Novel Gene Discovery in Primary Ciliary Dyskinesia

Mahmoud Raafat Fassad

Genetics and Genomic Medicine Programme

Great Ormond Street Institute of Child Health

University College London

A thesis submitted in conformity with the requirements for the degree

of

Doctor of Philosophy

University College London

Declaration

I, **Mahmoud Raafat Fassad**, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Primary Ciliary Dyskinesia (PCD) is one of the 'ciliopathies', genetic disorders affecting either cilia structure or function. PCD is a rare recessive disease caused by defective motile cilia. Affected individuals manifest with neonatal respiratory distress, chronic wet cough, upper respiratory tract problems, progressive lung disease resulting in bronchiectasis, laterality problems including heart defects and adult infertility. Early diagnosis and management are essential for better respiratory disease prognosis. PCD is a highly genetically heterogeneous disorder with causal mutations identified in 36 genes that account for the disease in about 70% of PCD cases, suggesting that additional genes remain to be discovered.

Targeted next generation sequencing was used for genetic screening of a cohort of patients with confirmed or suggestive PCD diagnosis. The use of multi-gene panel sequencing yielded a high diagnostic output (> 70%) with mutations identified in known PCD genes. Over half of these mutations were novel alleles, expanding the mutation spectrum in PCD genes. The inclusion of patients from various ethnic backgrounds revealed a striking impact of ethnicity on the composition of disease alleles uncovering a significant genetic stratification of PCD in different populations. Pathogenic mutations were also identified in several new candidate genes not previously linked to PCD.

Molecular and cell biology techniques were coupled with model organism studies to characterize the involvement of the new candidate genes in cilia motility and PCD. *Paramecium* was proven to be a good model for functional characterization of PCD potential candidate genes. The previously uncharacterized *C11orf70* was identified to play a highly conserved role in dynein assembly and intraflagellar transport (IFT)-related cilia cargo trafficking. Mutations identified in *DNAH9* resulted in a distinct motile cilia defect with mild respiratory symptoms, unusual in PCD. Mutations identified in two intraflagellar transport genes, *IFT74* and *WDR19*, linked together primary and motile ciliopathy phenotypes observed in the affected individuals.

Impact Statement

Motile cilia are continuously beating hair-like structures that clear airways, move gametes and determine left-right organ position. Defective cilia motility in primary ciliary dyskinesia (PCD) causes chronic respiratory problems and lung damage, infertility, hydrocephalus, malpositioned organs and heart defects. No single test currently confirms or excludes the diagnosis of PCD, a disease for which there is no cure. Working towards better PCD diagnosis and molecular understanding for improved disease management and development of personalized genetic therapies, this study used advanced sequencing techniques for genetic screening of PCD patients to aid novel gene discovery.

This study isolated several new genes not previously linked to PCD, providing fresh insights into: its cell biological basis by describing a new X-linked PCD subtype and a new player in dynein assembly (*PIH1D3, C11orf70*); genetic contribution to poorly recognised mild-spectrum subtypes of PCD (*DNAH9, CCDC103*); causes of respiratory incapacity within primary ciliopathies (*IFT74, WDR19*). By expanding the genetic landscape of PCD, we can improve genetic diagnosis. Genetic testing avoids the disadvantages and inconclusive results of other diagnostic testing methods currently used in healthcare, and consequently it will expedite the diagnostic pathway of PCD and facilitate better characterization of patients. Upon identification of the whole morbid genome of PCD, genetic testing can be considered a first-tier PCD diagnostic test.

Genetic results from this study have already contributed to two larger studies. Firstly, a published investigation of risk factors for laterality abnormalities and congenital heart defects in PCD, many being clinically actionable. This led to a recommendation for routine echocardiogram and abdominal ultrasound for all PCD patients. Secondly, a nationwide study, to be submitted shortly, to understand PCD genotype/phenotype correlations in the UK. These efforts will eventually help clinical management and inclusion in future therapeutic trials.

Patients from various ethnic backgrounds were recruited for this study, demonstrating successful international collaborations and the valuable outputs gained from multi-ethnic genetic studies in rare diseases. This will help in better genetic stratification of rare diseases, paving the way for future targeted therapy. This type of study can help countries with limited resources to pursue tailored testing strategies for genetic diagnosis of rare diseases. Moreover, this will help to provide appropriate genetic counselling to affected families.

Two successful training placements were undertaken in different laboratories abroad, building new multi-disciplinary collaborations across Europe in motile cilia research. Introducing *Paramecium* as a new model organism in PCD will help to identify new disease genes and validate the role of potential candidate genes in cilia motility. Bringing new proteomics skills to the UCL laboratory adds new prospects to the study of PCD disease mechanisms and how gene mutations cause disease phenotypes. Understanding the molecular basis of human diseases is an essential step to investigate future personalized therapeutic strategies.

Identifying motile cilia defects associated with primary ciliopathy phenotypes highlights the importance of investigating respiratory problems in ciliopathy syndromes. This can help in the management of the affected individuals, especially as respiratory failure is a common cause of death among these patients.

Acknowledgements

First and foremost, I feel indebted to Allah the most merciful and most compassionate who enabled me to accomplish this work

I would like to express my greatest gratitude to *Dr Hannah Mitchison* for her meticulous supervision and true concern to accomplish this work in the best possible image. My gratitude extends to *Prof Stephen Hart* and *Dr Eddie Chung* for their precious scientific guidance and continuous support.

This PhD would not have been possible without the immense help and valuable support from all my colleagues within Cilia Disorders Section. Special thanks go to my office and lab mate; *Mitali Patel* and my colleague and sister; *Dr Rasha Nour*. I would like to extend my thanks to all members of the UK PCD diagnostic and management service especially *Dr Amelia Shoemark* and my collaborators in Egypt, Palestine and Portugal. I am also grateful to *Dr Anne-Marie Tassin* and *Dr Karsten Boldt* for hosting me in their labs and providing unlimited help and constructive guidance.

Words are not enough to express my deepest thanks and endless gratitude to my parents, my brother; *Mohamed,* my sisters; *Shimaa* and *Asmaa*, my fiancée; *Salsabil* and all my friends for their endless support and continuous encouragement throughout my whole life. My appreciation and thanks to my cousin; *Ahmed*, his wife; *Jennifer* and their lovely daughter; *Rou'aia* for opening their house to me to stay with them over the past three years.

I would like to thank the British Council and the Egyptian government through the Ministry of Higher Education for giving me this opportunity to pursue PhD studies at University College London through Newton-Mosharafa Fund.

Last but not least, I would like to express my deepest thanks to all patients and their families for their kind cooperation.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

a.a	Amino Acid
ACMG	American College of Medical Genetics and Genomics
ALI	Air Liquid Interface
AONs	Anti-sense Oligonucleotides
ART	Assisted Reproductive Technologies
ATD	Asphyxiating Thoracic Dystrophy
ATS	American Thoracic Society
BB	Basal Body
BBS	Bardet-Biedl Syndrome
BCA	Bicinchoninic Acid
BEBM	Bronchial Epithelial Cell Basal Medium
BF	Basal Feet
BHB	Bio Herbe de Blé
BMI-1	B-cell specific Moloney murine leukemia virus integration site 1
bp	Base Pair
BSA	Bovine Serum Albumin
CADD	Combined Annotation Dependent Depletion
CBF	Cilia Beating Frequency
CBP	Cilia Beating Pattern
CC	Central Complex
CED	Cranio-Ectodermal Dysplasia
CF	Cystic Fibrosis
CGH	Comparative Genomic Hybridization
CHD	Congenital Heart Defect
Chr	Chromosome
CNV	Copy Number Variant
co-IP	Co-Immunoprecipitation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CT	Computed Tomography
ddH_2O	Double Deionized water

del	Deletion
delins	Deletion-Insertion
DIC	Differential interference contrast
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNAAF	Dynein Axonemal Assembly Factor
dpc	Days Post-Coitum
dsDNA	Double Stranded DNA
DynAPs	Dynein Assembly Particles
ERS	European Respiratory Society
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
FDR	False Discovery Rate
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GOS-ICH	Great Ormond Street Institute of Child Health
HEK	Human Embryonic Kidney
HC	Heavy Chain
HGNC	HUGO Gene Nomenclature Committee
hiPSC	Human induced pluripotent stem cell
hr	Hour
HSF	Human Splicing Finder
HSVM	High Speed Video Microscopy
IB	Immunoblotting
IC	Intermediate Chain
IDA	Inner Dynein Arm
IF	Immunofluorescence
IFT	Intraflagellar Transport
IGV	Integrative Genomics Viewer
IP	Immunoprecipitation
IPTG	Isopropyl β-D-thiogalactoside
kb	Kilo base

KD	Knockdown
LB	Lysogeny broth
LBAT	Lysogeny broth supplemented with Ampicillin and Tetracycline
LC	Light Chain
LFQ	Label Free Quantification
LRR	Leucine Rich Repeat
MAF	Minor Allele Frequency
min/s	Minute/s
MKS	Meckel-Gruber Syndrome
MT	Microtubule
MTEC	Mouse Tracheal Epithelial Cell Culture
NCBI	National Center for Biotechnology Information
NDK	Nucleoside Diphosphate Kinase
N-DRC	Nexin-Dynein Regulatory Complex
NGS	Next Generation Sequencing
nNO	Nasal Nitric Oxide
NRDS	Neonatal Respiratory Distress Syndrome
NTD	Neural Tube Defect
OD	Optical Density
ODA	Outer Dynein Arm
OMIM	Online Mendelian Inheritance in Man
P/S	Penicillin/Streptomycin
PBS	Phosphate Buffered Saline
PCD	Primary Ciliary Dyskinesia
PCR	Polymerase Chain Reaction
PDGFR	Platelet-Derived Growth Factor Receptor
PVDF	Polyvinylidene difluoride
qPCR	Quantitative Polymerase Chain Reaction
RBCs	Red Blood Cells
RefSeq	Reference Sequence
RGMC	Reduced Generation of Multiple Motile Cilia
RNAi	RNA interference

RS	Radial Spokes
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SD	Standard Deviation
sec	Second
SEM	Standard Error of the Mean
SHH	Sonic Hedgehog
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SRPS	Short Rib Polydactyly Syndrome
TAE	Tris Acetic Acid EDTA Buffer
TALENs	Transcription Activator-Like Effector Nucleases
TBS	Tris Buffered Saline
TBST	Tris Buffered Saline Tween 20
TEM	Transmission Electron Microscope
TGF	Transforming Growth Factor
Tm	Primer melting temperature
TPR	Tetratrico Peptide Repeat
TRX	Thioredoxin
UCL	University College London
USH	Usher
UV	Ultraviolet
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WHO	World Health Organization

Publications from this thesis

- **Fassad MR**, Shoemark A, Legendre M, Hirst RA, Koll F, le Borgne P, Louis B, Daudvohra F, Patel MP, Thomas L, Dixon M, Burgoyne T, Hayes J, Nicholson A, Cullup T, Jenkins L, Carr SB, Aurora P, Lemullois M, Aubusson-Fleury A, Papon JF, O'Callaghan C, Amselem S, Hogg C, Escudier E, Tassin AM, Mitchison HM "*Mutations in outer dynein arm heavy chain DNAH9 cause motile cilia defects and situs inversus*" **Am J Hum Genet.** 2018 Nov 20. https://doi.org/10.1016/j.ajhg.2018.10.016
- **Fassad MR**, Shoemark A, le Borgne P, Koll F, Patel M, Dixon M, Hayward J, Richardson C, Frost E, Jenkins L, Cullup T, Chung EMK, Lemullois M, Aubusson-Fleury A, Hogg C, Mitchell DR, Tassin AM, Mitchison HM "*C11orf70 mutations causing primary ciliary dyskinesia disrupt a conserved step in the intraflagellar transport-dependent assembly of multiple axonemal dyneins*" **Am J Hum Genet.** (2018)3;102(5):956-972
- Best S, Shoemark A, Rubbo B, Patel MP, **Fassad MR**, Dixon M, Rogers AV, Hirst RA, Rutman A, Ollosson S, Jackson CL, Goggin P, Thomas S, Pengelly R, Cullup T, Pissaridou E, Hayward J, Onoufriadis A, O'Callaghan C, Loebinger MR, Wilson R, Chung EM, Kenia P, Doughty VL, Carvalho JS, Lucas JS, Mitchison HM, Hogg C "*Risk Factors for Situs Defects and Congenital Heart Disease in Primary Ciliary Dyskinesia*" **Thorax** Published Online First: 30 August (2018). doi: 10.1136/thoraxjnl-2018-212104.
- Irving S, Dixon M, **Fassad MR**, Frost E, Hayward J, Kilpin K, Ollosson S, Onoufriadis A, Patel MP, Scully J, Carr SB, Mitchison HM, Loebinger MR, Hogg C, Shoemark A, Bush A "*Primary Ciliary Dyskinesia Due to Microtubular Defects is Associated with Worse Lung Clearance Index*" **Lung** (2018). 196(2), pp. 231-238

- Shoemark A, Moya E, Hirst RA, Patel MP, Robson EA, Hayward J, Scully J, **Fassad MR**, Lamb W, Schmidts M, Dixon M, Patel-King RS, Rogers AV, Rutman A, Jackson CL, Goggin P, Rubbo B, Ollosson S, Carr S, Walker W, Adler B, Loebinger MR, Wilson R, Bush A, Williams H, Boustred C, Jenkins L, Sheridan E, Chung EMK, Watson CM, Cullup T, Lucas JS, Kenia P, O'Callaghan C, King SM, Hogg C, Mitchison HM "*High prevalence of CCDC103 p.His154Pro mutation causing primary ciliary dyskinesia disrupts protein oligomerisation and is associated with normal diagnostic investigations*" **Thorax** (2018);73:157–166
- Olcese C, Patel MP, Shoemark A, Kiviluoto S, Legendre M, Williams HJ, Vaughan CK, Hayward J, Goldenberg A, Emes RD, Munye MM, Dyer L, Cahill T, Bevillard J, Gehrig C, Guipponi M, Chantot S, Duquesnoy P, Thomas L, Jeanson L, Copin B, Tamalet A, Thauvin-Robinet C, Papon JF, Garin A, Pin I, Vera G, Aurora P, **Fassad MR**, Jenkins L, Boustred C, Cullup T, Dixon M, Onoufriadis A, Bush A, Chung EM, Antonarakis SE, Loebinger MR, Wilson R, Armengot M, Escudier E, Hogg C; UK10K Rare Group, Amselem S, Sun Z, Bartoloni L, Blouin JL, Mitchison HM "*X-linked primary ciliary dyskinesia due to mutations in the cytoplasmic axonemal dynein assembly factor PIH1D3*". **Nat Commun** (2017) Feb 8;8:14279

Published abstracts

- **Fassad MR**, Shoemark A, Le Borgne P, Koll F, Patel M, Dixon M, Hayward J, Richardson C, Frost E, Jenkins L, Cullup T. Lemullois M, Aubusson-Fleury A, Hogg C, Mitchell DR, Tassin AM, Mitchison HM "*Genomic sequencing and reverse genetic approaches characterize C11orf70 gene: a novel player in the assembly of motor dyneins*" GENETICS RESEARCH 2018 Jul 12 (Vol. 100).
- Daudvohra F, **Fassad M**, Dixon M, Rogers A, Burgoyne T, Loebinger M, Hogg C, Mitchison H, Shoemark A "*S69 Genetic and structural characterisation of outer dynein arm variants causing primary ciliary dyskinesia*" Thorax (2017);72:A44
- Hogg C, Best S, Shoemark A, Rubbo B, Patel M, **Fassad M**, Carvalho J, Kenia P, Lucas J, Mitchison H "Genetic risk factors for laterality defects and congenital heart disease (CHD) in patients with primary ciliary dyskinesia (PCD)" European Respiratory Journal (2017);50: PA1852
- Shoemark A, Moya E, Hirst RA, Patel MP, Robson EA, Hayward J, Scully J, **Fassad MR**, Lamb W, Schmidts M, Dixon M, Patel-King RS, Rogers AV, Rutman A, Jackson CL, Goggin P, Rubbo B, Ollosson S, Carr S, Walker W, Adler B, Loebinger MR, Wilson R, Bush A, Williams H, Boustred C, Jenkins L, Sheridan E, Chung EMK, Watson CM, Cullup T, Lucas JS, Kenia P, O'Callaghan C, King SM, Hogg C, Mitchison HM "*A high prevalence CCDC103 p.His154Pro mutation causing primary ciliary dyskinesia is associated with normal diagnostic investigations*" European Respiratory Journal (2017);50: PA1851

Chapter 1 Introduction

1.1 Cilia structure

Cilia are complex hair-like organelles projecting from the surface of almost all cells in the human body.(1) The microtubular-based skeleton of the cilium is highly conserved across species to serve various physiological functions.(2) Various genetics and proteomics studies have identified over 1000 different proteins likely to reside in the cilium, highlighting the complexity of the cilia structure.(3)

The cilia structure can be divided into compartments including the basal body, transition zone, the axoneme, ciliary tip and the ciliary membrane.(1, 4) The basal body is formed when the centriole migrates to the apical cytoplasm. It is composed of nine triplet microtubules, distal appendages (transition fibres) connecting the basal body to the membrane at the cilia base and sub-distal appendages.(5, 6) The basal body forms the foundation for the ciliary axoneme, acting as a docking area for proteins involved in cilia assembly.(7) The transition zone acts as a gate for the ciliary compartment that controls which proteins can go in or out of the cilium. It has Y-shaped structures connecting the microtubules to the ciliary membrane.(8-10) (**Figure 1-1**)

Cilia are present either as a monocilium (one cilium per cell) or as multiple cilia (10s-100s cilia per cell).(1) Four different types of cilia can be distinguished in the human body based on their axonemal structural arrangement and motility. Motile cilia (multiple cilia) of the respiratory epithelium, brain ependyma and fallopian tube lining epithelium have a 9+2 axonemal arrangement: 9 peripheral doublets of microtubules surrounding a central pair of microtubules, along with microtubule-associated structures including inner (IDA) and outer (ODA) dynein arm motors, radial spokes (RS) and nexin-dynein regulatory complexes (N-DRC). The sperm flagellum has broadly the same structure as the motile cilium with subtle differences in the distribution of dynein arms along the axoneme.(11) Non-motile (primary) cilia (monocilium) are present in almost all cells and have nine pairs of microtubules at the periphery (9+0) with no central pairs nor dynein motors. Nodal motile cilia (monocilium) in the embryonic left-right organiser (the

embryonic node; a transient midline structure formed during gastrulation) have a similar structure to the primary cilia but with dynein subunits attached to the peripheral microtubules that provide their motility machinery.(12) The last type is the kinocilium of the inner ear hair cells, which has a 9+2 organization but without motility.(13, 14)

The emergence of cilia requires pre-assembly of many axonemal structures in compartments near the basal bodies.(15) In a recent pre-print, Huizar et al 2017(16) reported the presence of cytoplasmic foci where dynein assembly occurs. They showed that these are evolutionarily conserved compartments, proposing the term DynAPs (Dynein Assembly Particles) for these novel organelles where the assembly of multi-protein dyneins takes place.(16) Dynein arm assembly involves a chaperone-relay system for protein folding and maturation.(17) Once assembled, these multi-protein complexes are transported from the cytoplasm into the cilium through the transition zone and along the ciliary axoneme by an intra-ciliary protein transport system called intraflagellar transport (IFT).(18, 19) The IFT machinery is composed of multiple subcomplexes that move along the ciliary axoneme. Anterograde IFT (IFT-B complex) moves protein cargoes using the kinesin-2 molecular motor, from the transition zone to the ciliary tip where they are released. Retrograde IFT (IFT-A complex) moves cargoes for recycling back from the cilia tip to the base, using retrograde dynein-2 motors.(20, 21) (**Figure1-1**)

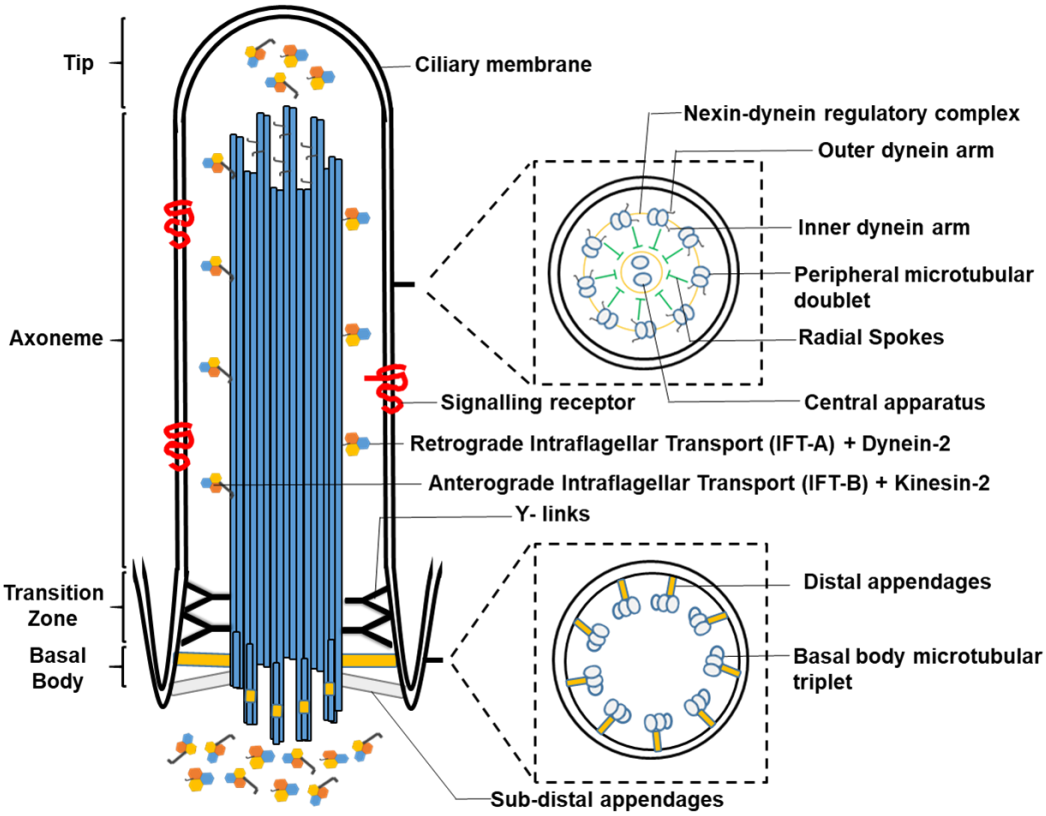


Figure 1-1 Structure of motile cilia

Cilia are cytoplasmic projections extending from the apical cell membrane. The cilium is composed of a basal body (a ring of triplet microtubules) that is attached to the base of the ciliary membrane through distal appendages. The transition zone, located distal to the basal body, contains microtubule doublets and is connected to the ciliary membrane by Y-links. The motile ciliary axoneme is formed of 9+2 microtubules connected through the nexin-dynein regulatory complex and radial spokes. Outer and inner dynein arm motors are attached to the peripheral microtubules. Intra-ciliary trafficking is governed by the intraflagellar transport system (IFT) in both directions (anterograde IFT-B carried by kinesin-2 motor towards the ciliary tip and retrograde IFT-A carried by dynein-2 towards the ciliary base.

1.2 Non-motile and motile cilia in health and disease

Although the basic structure of motile and immotile cilia is similar, they exert diverse distinct physiological functions during development and homeostasis.(22-24) Cilia related disorders are collectively known as ciliopathies. These conditions manifest in a wide spectrum of clinically and genetically heterogeneous diseases, however they still share an overlapping clinical phenotypic spectrum.(25, 26) Ciliopathies are defined as rare diseases as they typically affect less than 1 in 2,000 individuals in the general population.(27, 28) (**Figure 1-2**)

Non-motile or primary cilia serve as sensory organelles that regulate several intracellular signal transduction pathways.(29) The ciliary membrane contains cilia-specific receptors, ion channels and signalling molecules which receive extracellular signals (mechanical or chemical) and convert them into a signalling cascade intracellularly.(30) Signalling pathways such as sonic hedgehog (SHH), platelet-derived growth factor (PDGF), transforming growth factor β (TGF-β) and non-canonical Wnt pathways were shown to be mediated by primary cilia.(31-34) It was reported that IFT is not only essential for the transport of the axonemal components but also for the transport of channels in the primary ciliary membrane. Therefore, IFT plays a major role in the sensory function of the primary cilia.(20, 31)

Non-motile ciliopathies are clinically diverse diseases associated with a wide array of sensory, physiological and developmental abnormalities including polycystic kidney disease and other rare disorders like Usher (USH), Bardet-Biedl (BBS), Meckel-Grüber (MKS) and Jeune Asphyxiating Thoracic Dystrophy (JATD) syndromes. The phenotypic spectrum of primary ciliopathies usually involves multi-system anomalies (including the kidney, brain, eye and skeleton) that range clinically from embryonic lethality to isolated retinal degeneration.(35) There are several underlying mechanisms leading to primary ciliopathies, including abnormal cilia formation or maintenance, abolition of ciliary pathway components or defective trafficking of the signalling machinery.(4) Cilia motility defects have increasingly been

reported as a feature in primary ciliopathy syndromes, but these defects have not been widely studied.(36-39)

Some motile cilia perform a synchronized whip-like back and forth motion forming a wave-like pattern, known as the metachronal waveform of cilia motility. This is responsible for generating extracellular fluid flow or movement within fluids.(40) Motile monocilia in the embryonic left-right organizer exert a monodirectional motility (circular/rotatory beating pattern) which is essential for generating nodal flow (extraembryonic fluid flow), that leads to asymmetric expression of several genes. This determines left-right asymmetry of the embryo.(12, 41, 42) The motility of respiratory cilia is essential for proper mucociliary clearance through moving mucus and inhaled particles towards the nasopharynx to be removed by spitting or swallowing. In the same manner, ependymal cilia play a role in ependymal fluid movement and the beating of the ciliated epithelial lining of the fallopian tubes is important to direct fertilized ova towards the uterus.(43) Sperm flagellum beating is required for sperm movement through the female reproductive tract for fertilization of the ovum.(1) Defective cilia motility leads to primary ciliary dyskinesia (PCD), the only known motile ciliopathy in humans. About 50% of PCD-affected individuals have laterality problems.(44) There is a subtype of PCD known as Reduced generation of multiple motile cilia (RGMC), which has the same clinical phenotype but without laterality defects and is caused by defective multiciliogenesis.(43)

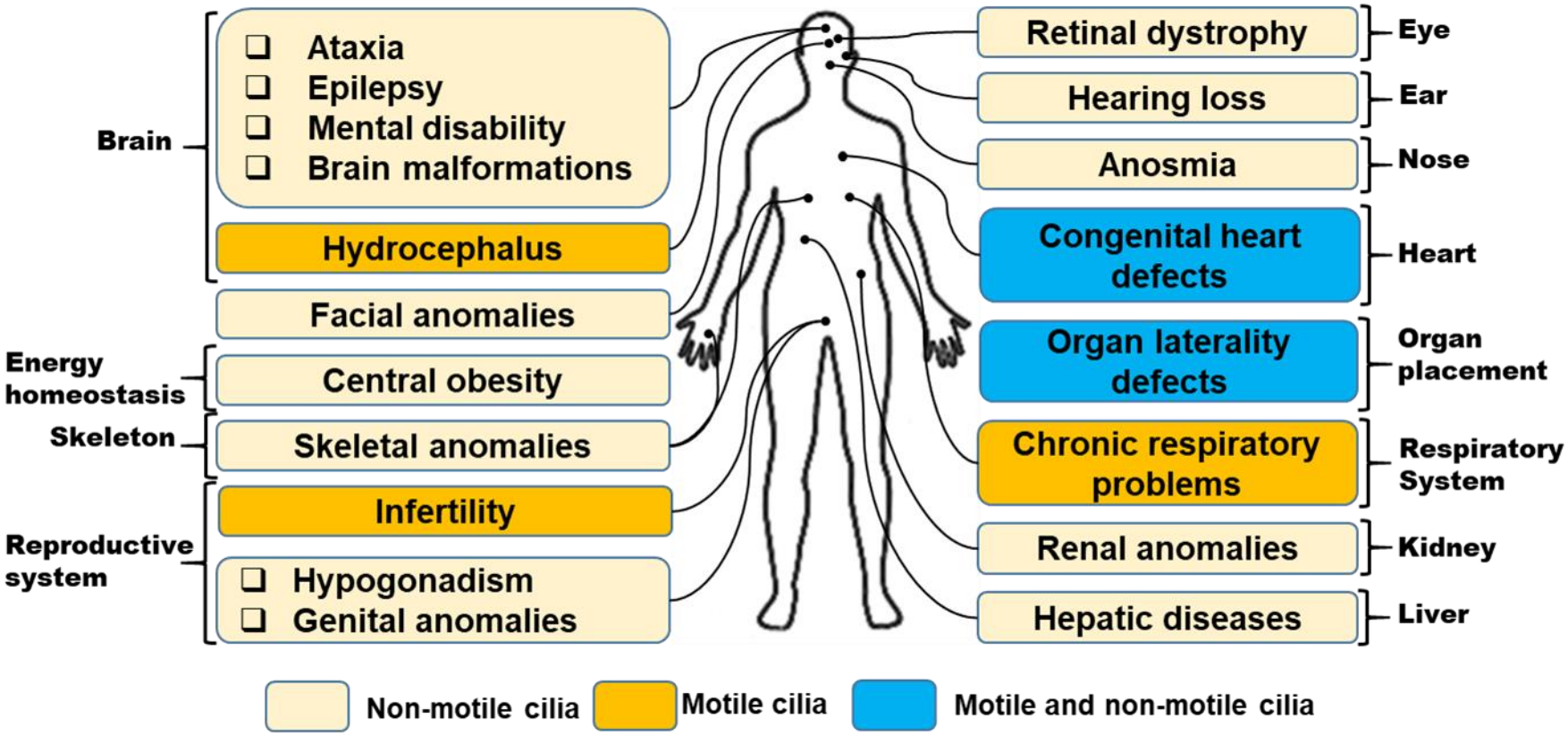


Figure 1-2 Multi-system disorders associated with non-motile and motile ciliopathies

Ciliopathies manifest with multi-system abnormalities and they share some overlapping phenotypic abnormalities. Ciliopathies can arise from non-motile or motile cilia defects or may be associated with defects in both types of cilia. (Figure adapted from Reiter JF and Leroux MR 2017) (4)

1.3 Primary Ciliary Dyskinesia

1.3.1 History of PCD

The first report of patients with bronchiectasis and situs inversus was in 1904 describing a condition that was firstly known as Zivert syndrome in the Soviet Union.(45) Later, it became widely known as Kartagener syndrome, named after a clinician from Zurich who clinically described it in 1933. He reported a familial form of bronchiectasis associated with malpositioning of the viscera, mainly dextrocardia and nasal polyps. In 1962, he identified 334 cases in the literature and described two more cases where bronchiectasis was associated with situs inversus.(46) In 1959, Gorham and Merselis concluded that Kartagener Syndrome follows an autosomal recessive mode of inheritance.(47)

In 1960 a clinician from Copenhagen, Arge, described 4 cases (3 males and 1 females) with complete transposition (mirror image) of the internal viscera (situs inversus totalis). They were offspring of a first-cousin marriage. All patients complained of recurrent respiratory problems and male infertility, however the affected female was fertile and had given birth to a child. Semen examination revealed immotile sperms. He concluded that infertility should be investigated in affected individuals with malpositioning of the internal organs.(48)

The underlying aetiology was not known until 1976 when Afzelius reported cilia immotility secondary to loss of dynein arms in patients with recurrent respiratory problems, situs inversus totalis and sperm immotility. At that time, he introduced the term “immotile cilia syndrome” to describe the congenital (inborn) immotility of the multiciliated epithelium of the respiratory system and the nodal monocilia of early embryos.(45, 49) In 1977, Afzelius and colleagues confirmed their previous findings of cilia immotility due to lack of dynein arms from the nasal epithelial cilia and sperm flagella. They hypothesized that the motility of monocilia in young embryos was essential to

determine the position of the heart and that in a case of cilia immotility, laterality is determined by chance.(50)

It was not until 1980 when the term “dyskinetic cilia syndrome” was firstly introduced. It was shown that immotile cilia syndrome was a misnomer as limited cilia motility could be observed in some affected individuals. Later in 1981, ‘primary ciliary dyskinesia’ was first proposed as the term to describe congenital ciliary dyskinesia and since then this term remains in use.(51, 52)

1.3.2 Clinical phenotypes

Primary ciliary dyskinesia (PCD) is a rare genetic disease, with a prevalence of about 1:10,000 individuals.(53, 54) PCD incidence is much higher within isolated or highly consanguineous populations like the Dutch Volendam population and the UK South Asian (primarily Pakistani) population, in whom disease prevalence is as high as 1:400 and 1:2,265 respectively.(44, 55)

Wherever cilia motility exerts an essential physiological function in a tissue or an organ, congenital defective cilia motility will lead to dysfunctional consequences presented as PCD. It is caused by motility defects of the monocilium at the embryonic left-right organizer, multicilia of specialized epithelial lining of the respiratory system and fallopian tubes, as well as the ependymal lining of brain ventricles and the highly modified sperm flagellum.(44) (**Figure 1-3**) **Table 1-1** summarizes the clinical symptoms of patients who need referral to a PCD diagnostic centre for further investigations. No treatments for PCD are available and disease management aims to control lung, sinus, and ear infections and remove trapped mucus from the airways. Early diagnosis significantly improves morbidity since lung damage can be delayed by specialist care.(56-58)

PCD can present early in life with unexplained neonatal respiratory distress.(59) Recurrent upper and lower respiratory tract symptoms represent the main clinical symptoms that arise from defective airway mucociliary clearance.(60) Respiratory symptoms include chronic nasal congestion, wet cough, rhino-sinusitis, serous otitis media, chronic middle ear effusion and

repeated lower respiratory infections. These can be complicated later in life with irreversible bronchiectasis and lung damage.(44) These symptoms are frequent within children and can overlap with many other paediatric diseases like cystic fibrosis, asthma and immunodeficiency, imposing a diagnostic challenge for clinicians.(61) This leads to delayed referral for proper diagnostic workup with delayed mean age at diagnosis of about 5 years.(54) PCD was found to account for about 5% of children with recurrent respiratory infections.(62) Bronchiectasis was found in about half of PCD patients (children and adults) and was reported in 32% of 84 cases with a mean age of PCD diagnosis at 6.4 years.(63, 64)

About 50% of PCD patients have abnormal laterality of their internal organs, most commonly situs inversus totalis. Previously, congenital heart problems were reported in 3.5 – 6% of PCD patients.(65-67) In a recently published study, we reported a frequency of 17.1% of congenital heart defects among PCD patients with or without laterality problems in the United Kingdom.(68)

Male infertility secondary to sperm immotility is common among adult PCD patients.(69) However, male patients with PCD can also have normal fertility as some genetic causes of escaped male infertility have already been identified.(55, 70) Dyskinesia of the cilia lining the fallopian tubes is believed to cause poor movements of the fertilized ovum towards the uterus, resulting in ectopic pregnancy.(69)

Beating of the ependymal cilia that line the brain ventricles generates the flow of the cerebrospinal fluid which moves signalling molecules along the central nervous system.(71, 72) Defective motility of the ependymal cilia can lead to hydrocephalus. Despite this being common in PCD mouse models, it is not a common complication in PCD patients. It was found that 10% of PCD patients with Reduced generation of multiple motile cilia (RGMC) had hydrocephalus. Interestingly, female infertility was high (22%) among the RGMC group too.(70, 73, 74)

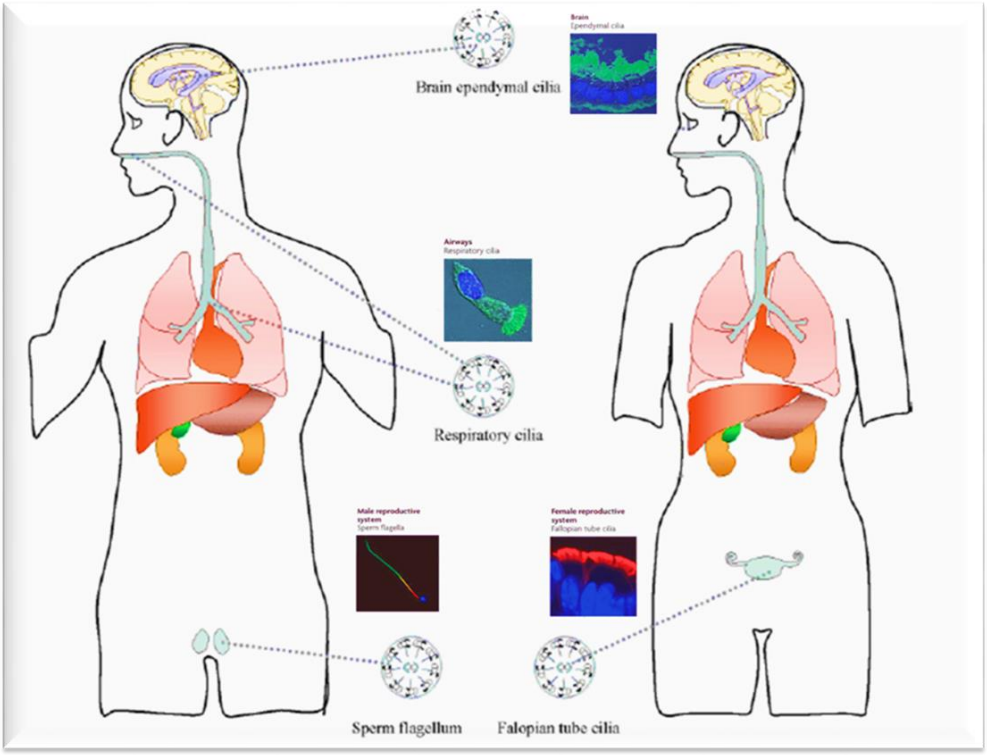


Figure 1-3 Motile cilia distribution in human body

Motile cilia are present in the epithelial lining the respiratory system and fallopian tubes, the ependymal lining of the brain ventricles and are structurally related to sperm flagella. Motile monocilia are present in the left-right organizer of the embryonic node. (Figure adapted from Ibañez-Tallon I et al 2003) (22)

Table 1-1 Clinical features that raise the suspicion of PCD and prompt further investigations

Patients with early onset recurrent respiratory problems and any of the following:
1. Situs anomalies and/or congenital heart defects
2. Unexplained neonatal upper and/or lower respiratory symptoms
3. Persistent wet cough, rhinitis or bronchiectasis after other common possible causes have been excluded
4. Positive family history of PCD
5. Adult males with dysmotile sperms
6. Chronic serous otitis media

The table adapted from Lucas JS et al 2016.(61)

1.3.3 Diagnostic odyssey of PCD

Definitive diagnosis of PCD imposes many clinical and diagnostic testing challenges. The clinical symptoms of PCD are heterogeneous and can overlap with many other diseases leading to a delay in referral for further diagnostic testing.(61) There is no single standard test to confirm or exclude the diagnosis of PCD and usually an array of testing is required.(75) The diagnostic workup of PCD is expensive and technically demanding requiring well-trained personnel and sophisticated equipment. These are not widely available in all centres or countries.(76, 77) There are no unified global guidelines of conducting or reporting results of any testing methods in PCD diagnostic workup.(61, 78, 79) Diagnostic tests that are currently used in PCD include measurement of the expelled nasal nitric oxide (nNO) level, high speed video microscopy (HSVM) analysis of the cilia beating pattern and frequency, transmission electron microscopy (TEM) analysis of the cilia ultrastructure, genetic analysis to identify the causative mutations and immunofluorescence staining (IF) studies to detect the presence or absence of ciliary proteins. **Table 1-2** summarizes the advantages and disadvantages of each diagnostic testing method.

In Europe, the European Respiratory Society (ERS) proposed a step-wise diagnostic algorithm where not all patients with a suggestive clinical and family history of PCD have to undergo all steps. Step 1 includes measurement of the nasal nitric oxide level and assessment of the cilia beating pattern and frequency by high speed video microscopy. Low nNO and/or abnormal HSVM analysis put PCD as a possible diagnosis. Step 2 includes analysis of the cilia ultrastructure by TEM. Identifying an ultrastructural defect of the ciliary axonemes by TEM confirms the diagnosis of PCD. Step 3 involves genetic testing in cases where TEM analysis showed normal or subtle defects of cilia ultrastructure. Genetic testing is also recommended in confirmed PCD cases with TEM analysis, to allow further characterization of the underlying defect since a single defect identified by TEM such as absent ODAs could be due to mutations in a number of different PCD genes.(44, 61, 75, 78)

In North America, the American Thoracic Society (ATS) recently proposed a different diagnostic algorithm where a panel of diagnostic tests are also recommended for PCD diagnosis. The ATS guidelines state that patients with clinical features of PCD should be referred for nNO measurement after excluding cystic fibrosis (CF). Two measurements showing a low level of nNO can diagnose PCD and subsequent genetic testing and/or TEM analysis is recommended for further characterization of the underlying defect as this can be helpful for the clinical prognosis and may suggest potential future therapeutic considerations. In cases where nNO measurement is not available or patients are too young to undergo testing, extended genetic testing is recommended. PCD diagnosis is confirmed by identifying bi-allelic pathogenic variants in a known PCD gene. In cases where the genetic findings are incomplete (identifying one pathogenic variant) or no genetic diagnosis was reached, TEM analysis should be done to confirm PCD by identifying a ciliary ultrastructural defect.(57, 79, 80) Some groups do not recommend using HSVM in clinical practice but using it only in research settings.(81, 82) The comparison between the ERS and the ATS guidelines is summarized in **Figure 1-4**.

Table 1-2 Advantages and disadvantage of PCD diagnostic tests currently used

Diagnostic test	Advantages	Disadvantages
Nasal nitric oxide (nNO)	1. There are guidelines and standardized protocols to conduct the test. 2. Good sensitivity and specificity.	1. Not suitable for children <6 years. 2. Small percentage of patients may have normal NO levels.
High speed video microscopy (HSVM)	1. Assesses functional defects of cilia motility. 2. There is abnormal cilia beating pattern and/or frequency in almost all PCD cases.	1. No standardized protocols and requires expertise and sophisticated equipment. 2. Subtle changes are difficult to be detected.
Transmission electron microscopy (TEM)	1. Assesses cilia ultrastructural defects. A hallmark defect confirms PCD diagnosis. 2. Widely used test and correlates to genetics	1. In 30% of cases there is no defect on TEM. 2. Requires well-trained personnel and highly sophisticated equipment.
Genetic testing	1. Overcomes diagnostic delay of other methods. 2. Better characterization of the patients.	1. In 20-30% of patients, no genetic cause has been identified. 2. Some variants are difficult to be interpreted.
Immunofluorescence staining (IF)	1. Relatively low cost, useful for research. 2. Several antibodies are commercially available.	1. No standardized protocols. 2. Cannot confirm diagnosis in many cases.

The table was adapted from Lucas JS et al 2016.(61)

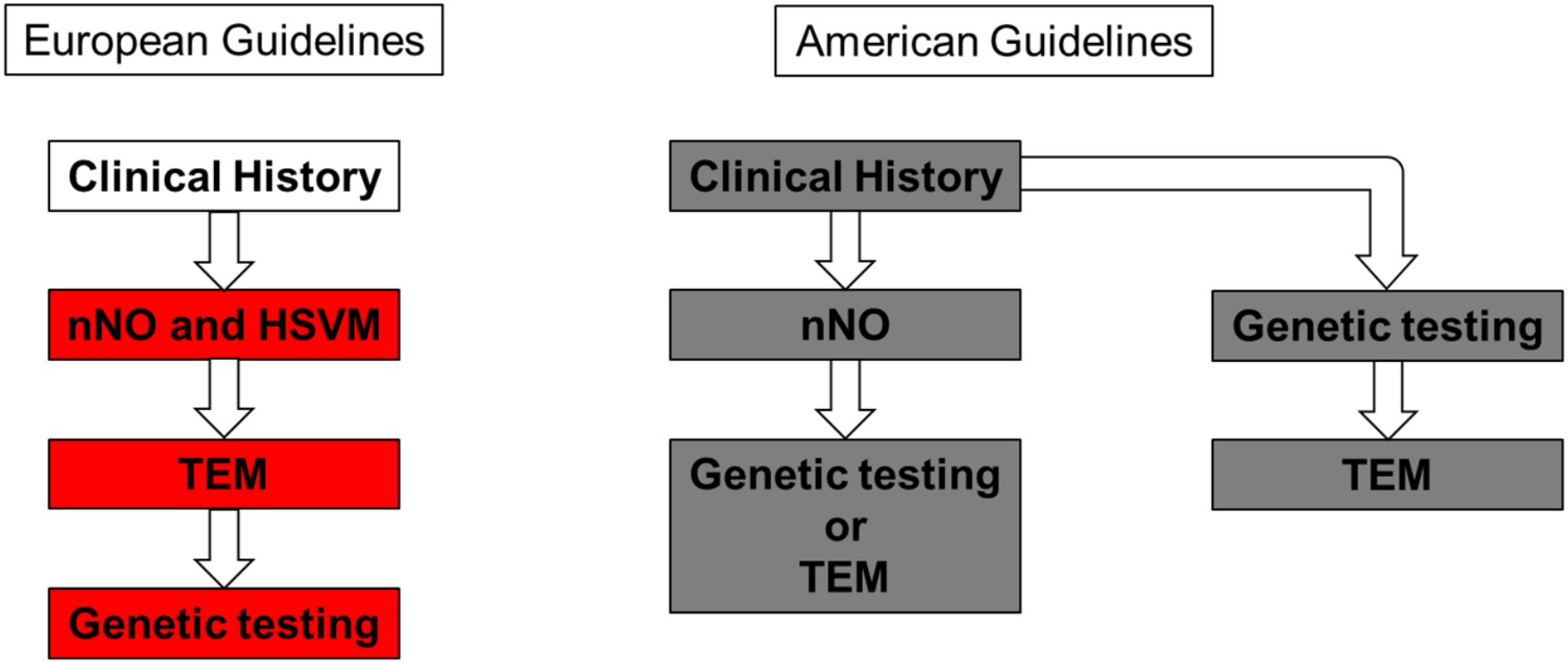


Figure 1-4 Summary of the diagnostic pathway for PCD in Europe and North America

The two most recent PCD diagnostic workflow guidelines follow different step-wise approaches in Europe and North America. Genetic testing is considered at step 3 in Europe (the ERS guidelines). In North America (the ATS guidelines), extended genetic panel testing is considered as a second step after two measurements of nasal NO and in cases where nNO measurement is not available or not possible, genetic testing is considered as the first-tier diagnostic method.

1.4 In vivo and in vitro studies of motile cilia

1.4.1 Model organisms to study motile ciliopathies

The structure of the motile ciliary axoneme is highly phylogenetically conserved across species. Many model organisms and animal models have been used to study the biology of motile cilia and the underlying disease mechanisms of PCD. They are valuable not only to identify potential candidate genes that may be defective in PCD but also for validation of the pathogenicity of human mutations.(83) The most popular model systems used to study the biology of cilia motility include: unicellular organisms like *Chlamydomonas reinhardtii*, invertebrates like *Drosophila melanogaster* and vertebrates like *Dario rerio*, *Xenopus* laevis and *Mus musculus*.(84) Phylogenetic analysis and comparative genomics have been used to identify motile cilia related genes.

Chlamydomonas reinhardtii, a unicellular bi-flagellate green alga, represents the historically most popular and the premier model organism used to study the structure and biology of motile cilia.(85) Isolation of several immotile *Chlamydomonas* mutant strains led to identifying many PCD-causing mutations through direct screening for pathogenic mutations in human orthologs of the *Chlamydomonas* potential candidate genes.(81, 86, 87) Other organisms like *Tetrahymena thermophila* and *Trypanosoma brucei* which are easy to be cultured in large numbers, have also provided a wealth of material for structural, biochemical and molecular studies, to allow for functional characterization of motile cilia related genes.(88-90)

Zebrafish *Dario rerio* is another popular animal model to study cilia motility. In zebrafish, motile cilia are found in several tissues including in Kupffer's vesicles (laterality organ of zebrafish embryo), the pronephros, otic vesicle and olfactory placode.(91-93) There are several advantages of using zebrafish as a model for PCD. Mating cycles yield many embryos for analysis. Embryos are transparent with short developmental time allowing the real time monitoring and visualization of organ development. The use of anti-sense

morpholinos also revolutionized functional characterization of many genes in a cheap simple way, despite the arising concerns about their off-target effects.(94-99)

The mouse *Mus musculus* is in many ways an ideal animal model to study the clinical-pathological consequences of PCD including developmental defects and physiological abnormalities e.g. laterality problems and mucociliary clearance defects.(60, 83) However, PCD mouse models typically develop hydrocephalus leading to a high mortality due to neurological complications. This has led to a difficulty in studying respiratory phenotypes in PCD mice.(60) This difficulty was overcome, for instance by introducing a conditional allele into transgenic mice causing postnatal deletion of *Dnaic1,* avoiding hydrocephalus and allowing studies of the progression of respiratory phenotypes.(100)

An alternate model for studying cilia structure and function is the *Paramecium tetraurelia*, which is a unicellular freely-living organism with a large number of motile cilia. With around 4000 cilia, it can provide a wealth of material for molecular and structural analysis to study the function of candidate genes in PCD, where small changes in cilia beat pattern can alter swimming behaviour.(101) Silencing of *Paramecium* genes by feeding a double-stranded RNA is a previously documented, efficient and easy method to knock down candidate gene expression.(102) Thus, *Paramecium* is an ideal unicellular model for functional characterization of candidate genes in PCD. This will be demonstrated in detail in **chapters 4&5**.

1.4.2 In vitro cell culture

In vitro culturing of multiciliated cells derived from human primary respiratory epithelial cells carries a lot of benefits for motile cilia research and PCD diagnostics. This provides an ex vivo model that can be used to study the molecular biology underlying multiciliogenesis. For diagnostic purposes, it can be used to differentiate between primary and secondary ciliary dyskinesia, as persistent ciliary dyskinesia after culture indicates that it is a primary defect rather than being secondary to something arising from mucus or infection in

the sample. Studying cilia beating dysfunction can also be evaluated using in vitro cell culture in attempts to study genotype/phenotype correlations.(103-105)

To establish the in vitro airway model, progenitor basal respiratory epithelial cells obtained from donors by nasal or bronchial brushing are cultured either at an air-liquid interface (ALI) or in a suspension, using specialized culture media. In ALI culture, air exposure is essential to initiate the differentiation of cells into a multiciliated pseudostratified epithelium resembling the airway epithelium. The cells from the brushing are placed onto collagen-coated, semipermeable membranes which sit on top of the culture medium, exposed to the air at the surface, with the cilia starting to be visible from around day 14 of ALI culture.(106) In suspension culture, the cells are cultured in the form of non-adherent polarized multicellular spheroids.(107) Many studies have used in vitro culture systems to confirm PCD diagnosis in difficult cases and to study factors regulating mucociliary differentiation and identify motile cilia related genes/proteins and associated molecular pathways.(108-110)

Fast increasing in popularity is the use of human induced pluripotent stem cells (hiPSC), which provide an unlimited source of cells that can be derived from patients or genetically engineered to carry mutations identified in the patients. They can be used as an in vitro model of PCD to investigate the molecular basis of the disease and also to study potential therapies by reprogramming them to differentiate to multiciliated cells.(111)

1.5 Novel gene discovery strategies in the era of next generation sequencing

Early gene discovery studies in PCD followed the approach of direct sequencing of human candidate genes with potential roles in cilia motility based on identification of their orthologs in model organisms, then subsequently linkage studies and homozygosity mapping were used.(86, 87, 112) In the past decade, the traditional genetic testing of a single candidate gene has been replaced by the massively parallel sequencing of multiple genes.(113)

Due to locus and allelic heterogeneity, next generation sequencing (NGS) technology has revolutionized the genetic diagnostics of PCD. Various NGS approaches have been used for novel gene discovery in PCD, including whole genome sequencing (WGS), whole exome sequencing (WES) and targeted multi-gene panel sequencing. This enabled fast and comprehensive genetic analysis and subsequently improved and expedited the diagnostic process.(56, 78-80, 114) (**Figure 1-5 & 1-6**)

With the introduction of NGS, the disease gene identification challenge shifts from the technical experiment to the data analysis phase. Tens of thousands of variants can be detected in a single NGS experiment. This necessitates applying careful, robust strategies for data interpretation and variant analysis.(115) Variant prioritization for Mendelian disorders assumes that the mutation should have a large impact and therefore should be very rare in the general population, with a direct effect on the protein encoded by the mutated gene.(116) The American College of Medical Genetics and Genomics (ACMG) has developed certain standards and guidelines for the interpretation of sequence variants identified by NGS. These guidelines described variant classification based on several levels of evidence e.g. population frequency, *in silico* predictions, functional evidence and segregation studies.(117)

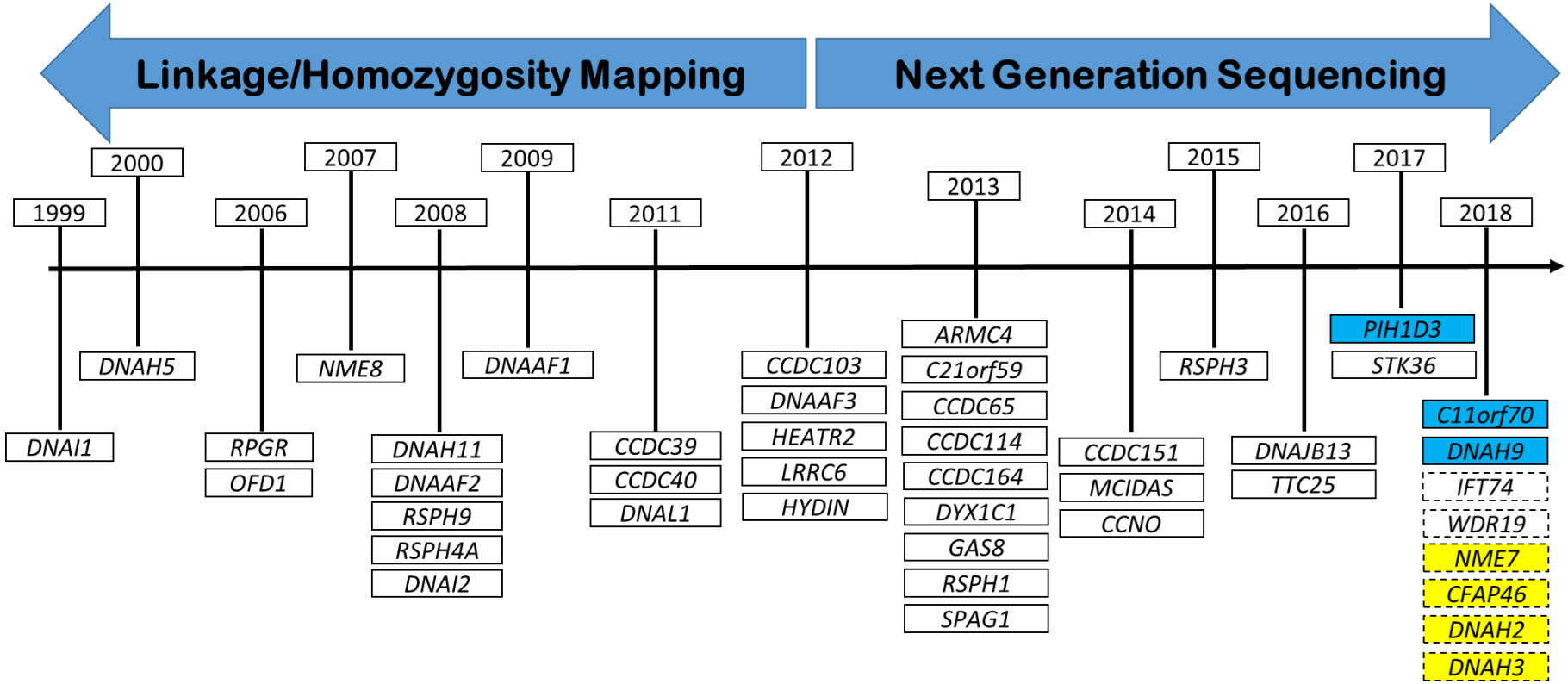


Figure 1-5 Timeline of gene discovery in primary ciliary dyskinesia

A timeline of gene discovery in PCD that started in 1999 with identifying mutations in *DNAI1* using a candidate gene approach. More than two thirds of PCD genes were discovered using next generation sequencing techniques, since 2012. Solid white boxed genes are published. The 2018 timeline genes are all referred to in the current study, blue boxes refer to newly published genes. Dotted white boxes refer to unpublished PCD genes but with functional confirmations. Yellow boxes indicate genes in which variants were detected without further functional characterization in this study.

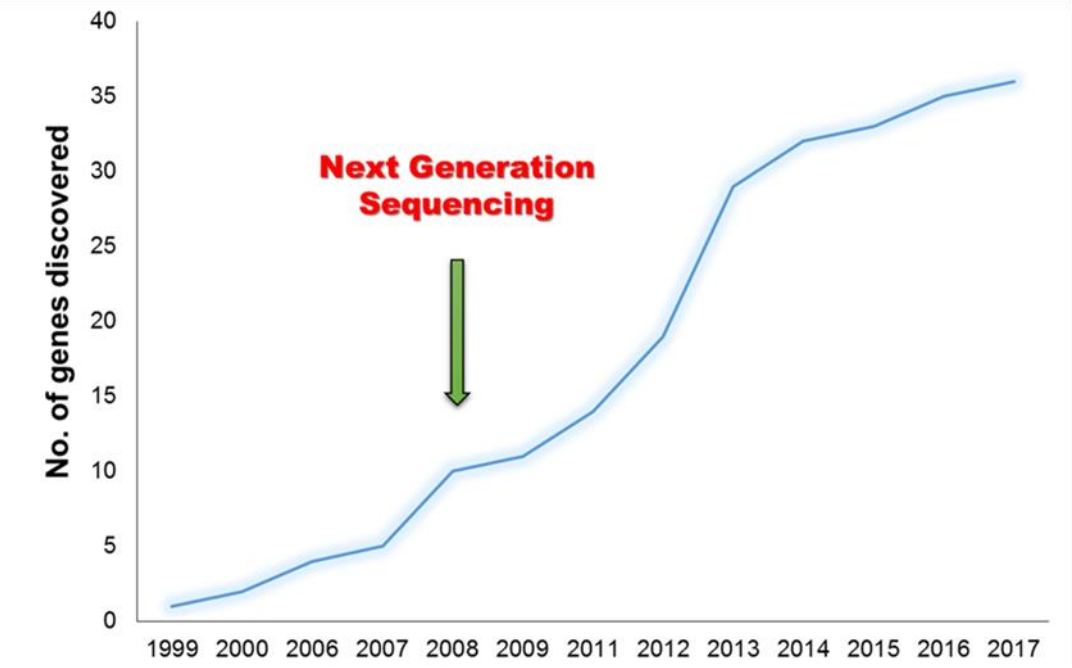


Figure 1-6 Next generation sequencing expedites PCD gene discovery

More than two thirds of known PCD genes were discovered using next generation sequencing techniques. This includes targeted sequencing, WES and WGS strategies.

1.5.1 Whole Genome Sequencing (WGS)

Whole genome sequencing (WGS) is considered the most comprehensive genetic testing. It is a powerful unbiased approach as it sequences the whole genome for detection of genetic variants. This approach imposes high cost and requires significant computational resources due to the massive data obtained. These considerations partially prohibit WGS from being widely used in disease-gene identification studies.(118, 119) WGS has the power to identify non-coding variants which are not readily studied in PCD. In a very recent pre-print, Ellingford et al (2018) presented the benefits of using WGS in PCD genetic research as they identified deep intronic variant in *DNAH11* in a PCD patient, reaching a definitive genetic diagnosis.(120)

With the advent of many promising whole genome sequencing initiatives like the Genomics England 100,000 Genome Project, it is expected that the sequencing costs will be reduced and new computational methods will be introduced for data analysis. Projects such as this, aims at sequencing of 100,000 whole genomes of NHS patients including patients with rare diseases, should greatly accelerate such progress in genomic medicine.(121) The use of whole genome sequencing in genetic testing carries the advantage of identifying more disease causing variants than other sequencing modalities. For example, WGS can identify regulatory and deep intronic variants which can also be uncovered through investigating RNA expression by RNA sequencing. Moreover, WGS can explicitly rule out other genetic causes e.g. novel structural variants which cannot be easily detected by other methods.(118, 122)

1.5.2 Whole Exome Sequencing (WES)

Whole exome sequencing (WES) is another unbiased powerful sequencing strategy that aims at sequencing of the whole complement of the coding exons (about 1% of the human genome). Typically, 20 000 to 50 000 variants can be detected in a single human exome. This depends on the exome enrichment kit used, the sequencing platform and the bioinformatics pipeline used for

analysis and variant calling.(115) In many cases, identification of the disease-causing variants by WES alone is not an easy process and additional strategies may be needed e.g. linkage studies, homozygosity mapping.

WES is more cost effective and less computationally demanding than WGS especially in case of searching for variants in protein-coding regions.(122, 123) It has been widely used for identification of many PCD genes and in many studies it was preceded or followed by homozygosity mapping.(80, 124)

1.5.3 Targeted multi-gene panel sequencing

The clinical utility of using NGS for variant detection is highly dependent on the ability to accurately isolate the genomic regions of interest. Targeted sequencing is more preferred than WGS and WES in the clinical settings as it is more affordable and yields higher coverage of the targeted regions.(123, 125) Moreover, targeted NGS ideally keeps errors due to PCR amplification or other sample manipulation to a minimum. There are several enrichment strategies currently being used to enrich for the target DNA sequences prior to sequencing e.g. emulsion PCR and hybridization-capture approaches.(126)

Targeted next generation sequencing is increasingly used in PCD diagnosis as the ATS guidelines recommend the use of extended gene panel screening for genetic testing of PCD patients.(127) This NGS approach was used for genetic stratification of PCD patients and novel gene discovery in the current study. It will be discussed in detail in the following chapters.

1.6 Genetics of PCD

PCD is generally inherited as an autosomal recessive disease with no apparent race or sex bias, though an X-linked form of PCD was recently described.(128, 129) Other very rare autosomal dominant and X-linked inheritance patterns have also been reported for ciliary dyskinesia-like phenotypes.(124, 130) PCD shows high allelic and locus heterogeneity as there are 36 genes, mutations in which lead to PCD.(131) To reach a specific genetic diagnosis of PCD, the guidelines state that bi-allelic mutations in an autosomal recessive state or hemizygous mutation in an X-linked gene should be identified.(58)

In the UK, genetic testing for PCD is currently placed at step 3, following the European diagnostic workflow guidelines (**Figure 1-4**). It is thereby used mainly to confirm the diagnosis and to characterize the underlying defect.(78) In North America, it is placed earlier in the proposed diagnostic algorithm as a second-tier diagnostic testing after nNO measurement. In cases where nNO measurement is not available or not suitable to the patients, extended genetic screening is considered to be a first line measure within the diagnostic workflow.(79)

Most of the genetic mutations known to cause PCD are of high pathogenic impact, being nonsense, frameshift or splice-site mutations while missense mutations are less reported.(58, 80) It is thought that the reported genetic mutations in known genes account for the disease in about 70% of PCD patients.(124, 131)

Mutations causing PCD have been identified so far in genes encoding proteins that are either structural ciliary components essential for motility, proteins essential for multiciliogenesis or genes involved in the assembly and transport of motile cilia structures.(124, 131) As the basic structure of the cilia is highly conserved across species, model organisms such as *Chlamydomonas reinhardtii* played a major role in the study of the genetics underlying human ciliopathies, including PCD.(132)

Many of ciliary genes in which mutations cause PCD, are expected to encode for protein components of the motility machinery of the cilia and play an essential role in cilia motility. To date, these include genes that encode for protein components of the outer dynein arm motors (ODAs), components of the ODAs docking complex, radial spoke (RS) proteins, protein components of the central complex (CC), components of the nexin-dynein regulatory complex (N-DRC) and proteins essential for the proper ring-array organization of the microtubules and assembly of the IDAs (cilia ruler proteins). There is another group of cytoplasmic proteins which are not structural components of the cilia but they play a major role in the dynein motor assembly and transport to the cilia. Furthermore, there is a group of genes that are essential for multiciliogenesis. Mutations in these genes lead to Reduced generation of multiple motile cilia (RGMC).(43) Other genes such as *RPGR* and *OFD1*, when mutated, lead to a syndromic form of defective motile cilia phenotype in which additional disease features such as retinitis pigmentosa and developmental delay are present.(36, 37, 39) (**Figure 1-7 &Table 1-3)**

In the following pages, I will list the genes known to be linked to PCD. I will summarize how each gene was discovered, the phenotype of the affected individuals and the results of model organism studies associated with each gene.

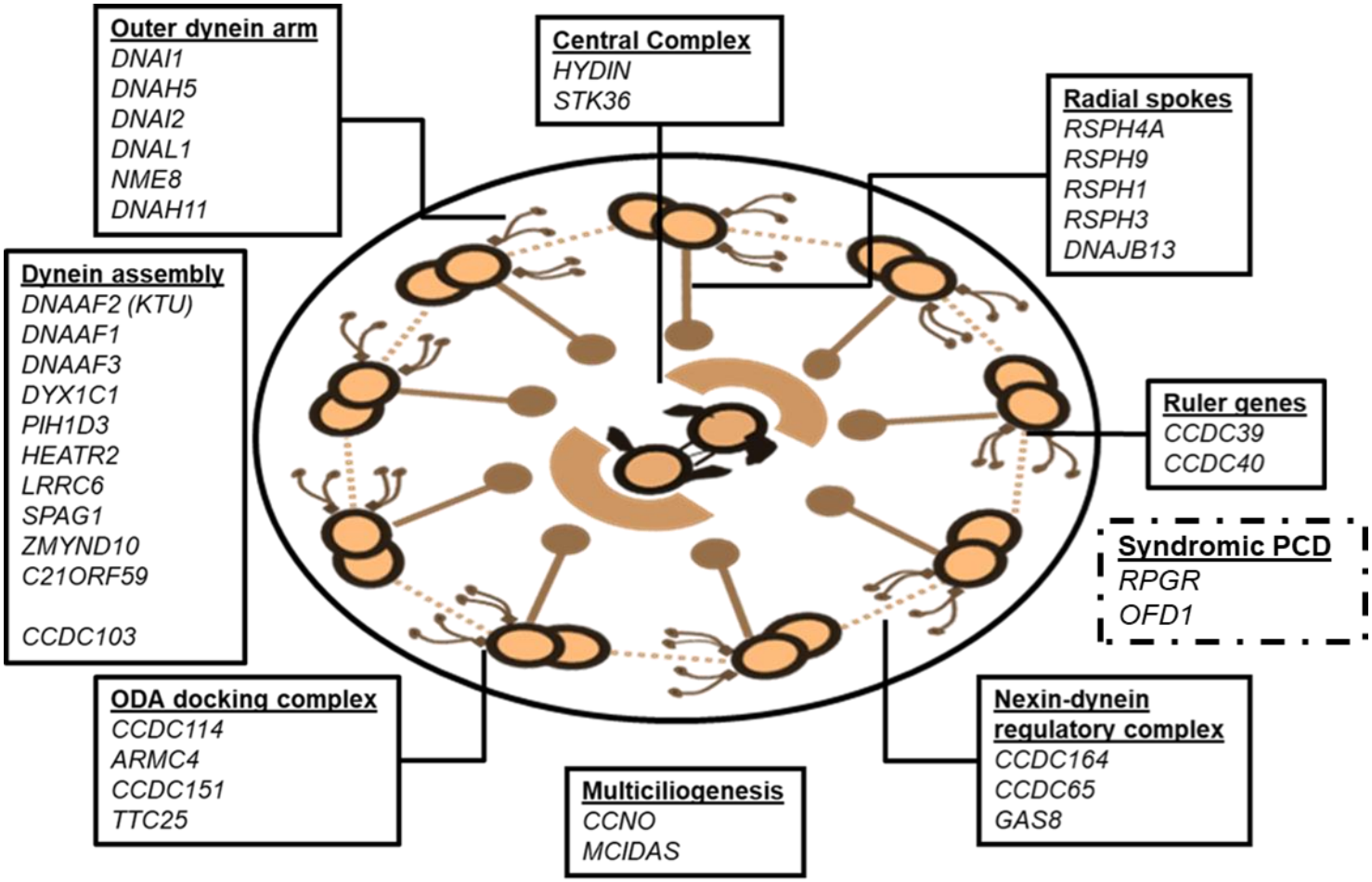


Figure 1-7 Groups of known genes when mutated lead to PCD

Table 1-3 List of known genes when mutated lead to PCD and associated defects

Gene	Locus	Role in axonemal integrity	TEM ultrastructural defect	Laterality	OMIM#
DNAI1	9p21-p13	ODA-IC	ODA loss	Yes	604366
DNAH5	5p15.2	ODA-HC	ODA loss	Yes	603335
DNAI2	17q25	ODA-IC	ODA loss	Yes	605483
DNAL1	14q24.3	ODA-LC	ODA loss	Yes	610062
TXNDC3	7p14-p13	ODA-IC	ODA loss	Yes	607421
DNAH11	7p21	ODA-HC	Normal	Yes	603339
CCDC114	19q13.32	ODA-docking complex	ODA loss	Yes	615038
ARMC4	10p12.1-p11.23	ODA-docking complex	ODA loss	Yes	615408
CCDC151	19p13.2	ODA-docking complex	ODA loss	Yes	615956
TTC25	17q21.2	ODA-docking complex	ODA loss	Yes	617095
RSPH4A	6q22.1	RS component	Mostly normal, CC defects	No	612647
RSPH9	6p21.1	RS component	Mostly normal, CC defects	No	612648
RSPH1	21q22.3	RS component	Mostly normal, CC defects	No	609314
RSPH3	6q25.3	RS component	Mostly normal, CC defects	No	615876
DNAJB13	11q13.4	RS component	Mostly normal, CC defects	No	610263
HYDIN	16q22.2	CC component	Normal, subtle CC defects	No	610812
STK36	2q35	Between RS head & CC	CC defects, lack of BF orientation	No	607652
CCDC164	2p23.3	N-DRC component	Normal, N-DRC loss	No	615288
CCDC65	12q13.12	N-DRC component	Normal, some MT disorganization	No	611088

GAS8	16q24.3	N-DRC component	Normal, some MT disorganization	No	605178
CCDC39	3q26.33	? N-DRC component	MT disorganization and IDA loss	Yes	613798
CCDC40	17q25.3	? N-DRC component	MT disorganization and IDA loss	Yes	613799
DNAAF2	14q21.3	Cytoplasmic Dynein assembly	IDA+ODA loss	Yes	612517
DNAAF1	16q24.1	Cytoplasmic Dynein assembly	IDA+ODA loss	Yes	613190
DNAAF3	19q13.42	Cytoplasmic Dynein assembly	IDA+ODA loss	Yes	614566
DYX1C1	15q21.3	Cytoplasmic Dynein assembly	IDA+ODA loss	Yes	608706
PIH1D3	Xq22.3	Cytoplasmic Dynein assembly	IDA+ODA loss	Yes	300933
HEATR2	7p22.3	Cytoplasmic Dynein assembly	IDA+ODA loss	Yes	614864
LRRC6	8q24	Cytoplasmic Dynein assembly	IDA+ODA loss	Yes	614930
SPAG1	8q22	Cytoplasmic Dynein assembly	IDA+ODA loss	Yes	603395
ZMYND10	3p21.3	Cytoplasmic Dynein assembly	IDA+ODA loss	Yes	607070
C21orf59	21q22.1	Cytoplasmic Dynein assembly	IDA+ODA loss	Yes	615494
CCDC103	17q21.31	? Dynein attachment	IDA+ODA loss	Yes	614677
CCNO	5q11.2	Multiciliogenesis	Oligocilia	No	607752
MCIDAS	5q11.2	Multiciliogenesis	Oligocilia	No	614086
RPGR	Xp11.4	??	variable	??	312610
OFD1	Xp22.2	Centrosomal protein	Short rare, disoriented cilia	Yes	300170

The table is adapted from Leigh M et al 2015.(133) TEM; Transmission electron microscopy, OMIM; Online Mendelian Inheritance in Man, ODA; Outer dynein arm, IC; Intermediate chain, HC; Heavy chain, LC; Light chain, RS; Radial spokes, CC; Central complex, N-DRC; Nexin-dynein regulatory complex, BF; Basal feet, MT; Microtubule, IDA; Inner dynein arms.

1.6.1 Outer dynein arm genes

Outer dynein arms (ODA), the force generators of axonemal motion, are very large macromolecular complexes attached to the peripheral microtubules doublets. They generate the oscillatory bending motion, or power of the stroke, of the axoneme by hydrolysing ATP. The ODA structure is highly evolutionarily conserved and it repeats every 24 nm along the length of the axoneme. They are composed of heavy, intermediate and light dynein chains. (134, 135) Mutations in genes encoding for ODA components lead to cilia ultrastructural defects with loss of ODAs and defective cilia motility.

DNAI1

In 1999, *DNAI1* was described to be the first human gene with identified mutations causing PCD, using candidate gene approach based on the high evolutionary conservation of the encoded protein and its role as a ciliary axonemal structural component across species. Mutations in *DNAI1* are associated with loss of ODA and cilia immotility. Locus heterogeneity of PCD was first demonstrated when mutations in *DNAI1* were excluded in patients with ODA loss ultrastructural phenotype.(87)

Further mutation screening studies on *DNAI1* expanded its mutation spectrum in PCD patients.(136, 137) In a later report, mutations in *DNAI1* were identified in 10% of a large cohort of PCD patients demonstrating a founder effect of one common mutation.(138) Screening of a large Polish PCD cohort indicated a worldwide prevalence of *DNAI1* mutations of about 7-10% in PCD families not preselected for ODA loss.(139)

DNAH5

DNAH5 was first considered as a potential candidate gene for PCD through homozygosity mapping. Screening of a large consanguineous Arabic family with PCD and ODA loss revealed a region of homozygosity by descent on chromosome 5p15-p14 where *DNAH5* gene is localized.(86) Mutations in its *Chlamydomonas* ortholog *oda2* were previously shown to cause slow

swimming and similar ultrastructural abnormalities to the defects seen in the patients.(140)

Linkage analysis followed by sequencing of all exons of *DNAH5* identified mutations carried by affected individuals from 8 families with PCD and ODA loss. In the same study, *in situ* hybridization analysis showed localization of DNAH5 in the left-right organizer of mouse embryos.(141) Moreover, mice deficient for *Dnahc5* developed respiratory problems and 50% of them had laterality problems.(142)

Screening of 109 white PCD families, identified mutations in *DNAH5* in 30 families. Absence of DNAH5 from the ciliary axoneme or its presence only in the proximal part of the patients' cilia was shown by immunofluorescence staining. *DNAH5* mutations were identified in 53% of a pre-selected cohort of patients with ODA loss.(143) In another study, mutations in *DNAH5* were identified in 15% of a clinically heterogeneous group of PCD patients.(144)

DNAI2

In 1991, Mitchell and Kang showed that the *Chlamydomonas* mutant strain *oda6* which carries mutations in the dynein intermediate chain gene *IC69,* has flagellar immotility associated with loss of ODA.(145) *DNAI2* was identified as the human ortholog of *IC69*. Later, *DNAI2* expression studies using northern blot analysis of various adult human tissues showed expression of *DNAI2* in the trachea and testes only. However, in this study, *DNAI2* mutation screening of 12 patients with ODA loss could not identify any mutations.(146)

In mice, *Dnaic2* mRNA expression was found in multiciliated tissue including the ovaries, testes and lung.(147) Mutant *dnai2* in Medaka fish conferred defective cilia motility in the Kupffer's vesicle causing laterality problems.(148, 149) Linkage analysis followed by direct sequencing of *DNAI2* in a large cohort of PCD patients identified loss of function mutations in 3 families with ODA loss. Immunofluorescence staining of *DNAI2* patients demonstrated loss of DNAH5 and DNAH9 form the patients' ciliary axoneme.(150)

DNAL1

In 2011, a missense mutation in *DNAL1* converting asparagine to serine at position 150 was reported in a PCD family with absent ODA. It was shown that the mutation affects the stability of the axonemal dynein light chain-1, DNAL1 and damages its interactions with dynein heavy chains and tubulin.(151) Immuno-gold labelling electron microscopy analysis of *Chlamydomonas* flagellar axoneme showed localization of LC1 near the microtubule-binding domain (MTBD) of Ɣ heavy chain dynein.(152)

Human *DNAL1* and mouse *Dnal1* were identified as highly conserved orthologs to *Chlamydomonas* axonemal dynein light chain-1 *(LC1)*. Direct interaction between DNAH5 and DNAL1 was demonstrated by co-immunoprecipitation. Expression analysis of *DNAL1* in various human tissues showed its expression only in testis but not in other tissues where motile cilia are not present. In adult mice, *Dnal1* was demonstrated to be expressed in respiratory epithelium and ependymal lining brain ventricles in addition to the node of mouse embryos.(153)

TXNDC3 (NME8)

Homozygosity mapping in a large Turkish family with Kartagener syndrome identified a candidate locus 7p14.1, which is known to contain the *TXNDC3* gene.(154) Later direct sequencing of 18 exons of the candidate PCD gene, *TXNDC3* gene in 47 patients (33 with ODA loss, 8 with normal ultrastructure and 6 males with infertility due to loss of sperm ODA) revealed nonsense mutation and common intronic variant in a compound heterozygous state in an affected girl with ODA lost in 66% of her examined respiratory cilia. It was shown that the intronic variant affects the ratio of the two *TXNDC3* isoforms causing reduced expression of TXNDC3d7 isoform.(155)

TXNDC3, also known as *NME8*, encodes a thioredoxin nucleoside diphosphate kinase, a member of a large family of enzymatic proteins that act as protein disulphide reductases. It contains thioredoxin (TRX) and nucleoside diphosphate kinase (NDK) domains. TXNDC3 is the human ortholog of the sea urchin IC1, a component of sperm outer dynein arms.(156, 157) *TXNDC3*

was also shown to be expressed in human sperm and respiratory epithelium.(155, 156)

DNAH11

Mutations in *DNAH11* were first described in a patient with a combined phenotype of cystic fibrosis and PCD with situs inversus totalis. Paternal uniparental isodisomy of chromosome 7 was described in this patient who was also carrying a homozygous F508del mutation in the *CFTR* gene. Direct sequencing of 82 *DNAH11* exons revealed a homozygous nonsense mutation in exon 52. Electron microscopy analysis of the patients' cilia revealed a normal axonemal ultrastructure.(158)

Linkage analysis followed by direct sequencing of *DNAH11* exons revealed compound heterozygous mutations in a large family with PCD. Immunofluorescence staining displayed normal localization of DNAH5 and DNAH9 in the patients' cilia associated with normal axonemal ultrastructure by electron microscopy analysis. High speed video microscopy confirmed abnormal beating pattern of the patients' respiratory cilia.(159) Mutant *Dnahc11* reproduced the human PCD phenotype in mice, which show laterality problems, respiratory symptoms, reduced sperm motility and normal cilia ultrastructure.(160)

Screening of 58 unrelated PCD patients with normal ciliary ultrastructure identified mutations in *DNAH11* in 22% of these patients who showed a hyperkinetic cilia beating pattern and reduced waveform amplitude.(161) Immunofluorescence staining revealed localization of DNAH11 in the proximal part of the respiratory cilia of healthy control.(162) Electron tomography showed a loss of >25% of the volume of outer dynein arms in the proximal part of cilia from patients with *DNAH11* mutations. This subtle ultrastructural defect could not be detected by TEM.(163)

1.6.2 ODA docking complex genes

In order to achieve this 24 nm periodicity along the axoneme, ODAs require a heterotrimeric docking complex (ODA-DC) to facilitate their anchoring into the microtubules.(164) Mutations have been identified in a number of genes encoding for protein subunits of ODA-DC, associated with defective cilia motility and ODA loss.

CCDC114

Whole exome sequencing (WES) in two distantly related affected individuals from a large family from Volendam village (North Holland), identified a mutation in a consensus splice donor site of exon 7 in *CCDC114*, leading to an intronic insertion of 79 base pairs (bp) and a frameshift in the CCDC114 protein. Direct sequencing of 44 individuals with either ODA or combined IDA+ODA loss led to the identification of another mutation in the consensus splice donor of exon 5 in one UK family.(55)

In a parallel study, whole exome sequencing identified mutations in *CCDC114* in one out of 6 families with ODA loss. Subsequent direct sequencing of *CCDC114* in probands of 58 unrelated families with ODA defects identified mutations in 5 more families.(165) The *Chlamydomonas* ortholog of CCDC114 is DCC2, a component of the outer dynein arm docking complex. In the mutant strains *oda1* where DCC2 is lost, ODA docking complexes are not assembled and ODAs cannot be attached to the microtubules leading to ODA loss.(164, 166)

ARMC4

Mutations in the *ARMC4* gene were first identified when a custom high-resolution array-CGH (comparative genomic hybridization) approach was used for Copy Number Variations (CNVs) screening in a cohort of ciliopathy individuals. A homozygous deletion of 8.4 kb spanning exons 4-7 of *ARMC4* was identified in two affected siblings. Direct sequencing of *ARMC4* in 135 PCD patients with ODA loss led to identifying mutations in *ARMC4* in patients from 9 more families.(167) It was previously shown that *Armc4* is expressed

in multiciliated cells with an expression profile similar to that of other ciliary genes in mouse.(168, 169)

In patients with *ARMC4* mutations, DNAH5 and DNAI2 are absent from the distal cilia but still present proximally. Interestingly, DNAH9 was not absent but mislocalized along the whole ciliary axoneme. *CCDC114* mutations were found to affect the localization of ARMC4, while CCDC114 localization was found not to be ARMC4-dependent.(167) The use of whole genome sequencing (WGS) for the first time in PCD identified loss of function mutations in *ARMC4* in a Pakistani family. It was also shown that the *ARMC4* expression is upregulated during multiciliogenesis.(170)

CCDC151

Autozygosity mapping and whole exome sequencing identified mutations in *CCDC151* in 3 unrelated families associated with loss of ODAs. Both CCDC114 and ARMC4 were lost in cilia from patients with *CCDC151* mutations. By co-immunoprecipitation, CCDC151 was found to interact with CCDC114 but not with DNAI1, DNAI2 and DNALI1.(171) *CCDC151* is highly conserved across species with motile cilia. It was shown to be associated with IFT-dependent dynein assembly in *Drosophila*. Knockdown of *Ccdc151* in zebrafish led to laterality defects and renal cyst formation.(172)

TTC25

Mutations in the *TTC25* gene were identified in two unrelated PCD families with ODA loss by homozygosity mapping and whole exome sequencing. CRISPR/Cas9 was used to introduce a deletion of *Ttc25* exon 2 and 3 in mice resulting in small-sized mice, hydrocephalus and laterality problems. Mutations in *TTC25* led to depletion of CCDC114, ARMC4 and CCDC151 from the patients' cilia.(173) Previous studies on *Xenopus* embryos displayed localization of TTC25 at cilia base and along the ciliary axoneme and its knockdown led to defective cilia assembly and short fewer cilia.(174)

1.6.3 Radial spokes genes

Radial spokes (RS) are T-shaped multi-protein projections, repeating every 96 nm periodicity, extending between the peripheral microtubule doublets and the central pair complex (CC) to regulate cilia motility.(175) It is believed to relay signals from the central pair of microtubules to the dynein arms. Mutations in RS genes in Chlamydomonas or PCD patients result in paralyzed or abnormally moving cilia, frequently associated will loss of the CC in the patients' cilia.

RSPH4A

SNP-based linkage analysis in 7 consanguineous PCD families with central complex (CC) abnormalities identified the 6q22.1 locus in 5 families in which there was a complete absence of the central pair of microtubules. Mutations identified in *RSPH4A* in this locus and haplotype analysis revealed a common ancestral founder *RSPH4A* mutation in UK-Pakistani families. RSPH4A protein was found to be 31% and 30% identical to two *Chlamydomonas* proteins RSP4 and RSP6 respectively.(176) Both RSP4 and RSP6 are known to be components of the head of radial spokes in *Chlamydomonas*.(177)

Direct screening for *RSPH4A* mutations in 184 Eastern European families with PCD estimated that *RSPH4A* mutations are responsible for 2-3% of PCD patients in the Eastern European population.(178) In another study, a splice site mutation was identified in 9 PCD patients from 7 families of Hispanic descent, indicating a founder effect in patients from Puerto Rico. None of these patients had laterality problems.(179)

RSPH9

Linkage analysis in two Bedouin families with PCD and intermittent loss of CC in cilia cross sections, identified the 6q21.1 locus which carries *RSPH9* as a potential candidate gene. Direct screening for *RSPH9* mutations identified inframe deletion p.Lys268del in both families.(176) In another study, the same mutation was identified in another family from Saudi Arabia indicating a possible founder effect of this mutation in the Arab Peninsula.(180, 181)

RSP9 is the *Chlamydomonas* ortholog of RSPH9 and it is a component of radial spoke head. *Chlamydomonas RSP9* mutants (*pf17*) display immotile flagella with loss of radial spoke head.(177, 182) Expression profile of *Rsph9* in mice revealed its enrichment in tissues with motile cilia and in the embryonic node. In zebrafish, *rsph9* was shown to be expressed in both motile cilia 9+2 and kinocilia 9+0 and its knockdown yielded defective cilia motility phenotype.(176, 183)

RSPH1

Human RSPH1 is the ortholog of the *Chlamydomonas* RSP1 protein, one of the protein components of the head of the ciliary radial spokes.(177) Homozygosity mapping and WES first identified nonsense mutation in the *RSPH1* gene in a female PCD patient with a predominant central complex (CC) defects, with no mutations in *RSPH4A* and *RSPH9*. Further direct screening for *RSPH1* mutations in 36 unrelated families with CC defect and no mutations in *RSPH9* and *RSPH4A*, found 10 families with bi-allelic mutations in *RSPH1*. None of the affected individuals had laterality problems and cilia motility displayed a mixed range of dyskinetic beating patterns (active, slow, immotile).(184)

In another study, WES identified splice site mutation in *RSPH1* in two affected siblings with normal ciliary ultrastructure. Further screening of *RSPH1* in 413 unrelated probands, identified mutations in nine additional families. Affected individuals showed milder features of PCD with better lung function and higher levels of nasal NO compared to patients with mutations in other PCD genes. Their cilia showed abnormal circular beating pattern with normal beating frequency.(185) In a parallel study using targeted gene panel sequencing in 70 patients with PCD, *RSPH1* mutations were identified in two families with isolated loss of the ciliary central pairs and no laterality defect.(186)

RSPH3

Homozygosity mapping and WES first identified nonsense *RSPH3* mutations in a PCD family with CC defects. In 26 unrelated families with CC defects and no mutations in *RSPH1, RSPH4A* or *RSPH9,* direct screening for *RSPH3*

mutations identified 4 more families with *RSPH3* mutations.(187) RSPH3 is the human ortholog of the *Chlamydomonas* RSP3 protein, which forms a dimer that creates the core of the radial spoke stalk.(188) It was shown that *RSPH3* mutations do not disrupt the localization of RSPH23 (radial spoke neck protein), RSPH1 and RSPH4A (radial spoke head proteins) but they lead to the depletion of RSPH11 (radial spoke stalk protein) from the ciliary axoneme.(187)

DNAJB13

WES and homozygosity mapping identified *DNAJB13* mutations in two unrelated families with PCD and male infertility. TEM analysis of cilia cross sections of the affected individuals showed loss of central microtubular pairs.(189) *DNAJB13* expression was found in mouse tissues with motile cilia including testes, lung, oviduct and brain. It was also shown to be present along the entire length of the mouse sperm flagellum and localized to the radial spokes of the ciliary axoneme.(190, 191)

1.6.4 Central complex genes

The central pair complex (CC) is composed of the two single central microtubules and the associated network of interconnected projections.(192) Nodal monocilia are lacking the central complex and mutations in genes encoding for CC protein components do not usually cause laterality problems.

HYDIN

In mouse, *Hydin* was shown to be expressed in the ependymal lining of the brain ventricles, the ciliated epithelium of the respiratory system and oviduct and in developing spermatocytes. A frameshift mutation in *Hydin* in *hy3* mice led to congenital hydrocephalus.(193) In *Hydin* mutant mice, the ciliary axoneme showed a normal ultrastructure apart from lacking a specific projection of the central complex. The cilia were unable to bend normally and there was a reduced cilia beating frequency.(194) In *Chlamydomonas*, HYDIN was shown to be a component of the central apparatus and its knockdown led to short flagella with loss of the C2b projection of the central complex.(195)

Homozygosity mapping and WES identified *HYDIN* mutations in PCD patients with no laterality problems. TEM analysis of cilia cross sections showed a loss of C2b projections; a similar finding to what was previously shown in *Chlamydomonas* and mice.(196)

STK36

WES identified a loss of function mutation in the *STK36* gene in a PCD patient without laterality problems. TEM analysis of the patient cilia showed loss of the central pair and lack of orientation of the basal feet of the cilia.(197) *Stk36 ($Fu^{-/-}$)* knockout mice show hydrocephalus and growth retardation associated with respiratory symptoms and infertility. No laterality problems were detected in these mice.(198) It was shown that STK36 is present motile cilia with probable localization between the radial spoke head and the central complex. It interacts with SPAG16 (a central pair component) suggesting a conserved role in central pair formation and it also interacts with KIF27 at the base of the cilia suggesting a potential function in cilia orientation.(197, 199)

1.6.5 Nexin-dynein regulatory complex genes

Nexin-dynein regulatory complexes (N-DRC) are large multi-molecular structures composed of at least 13 subunits linking between peripheral microtubule doublets.(200) The N-DRC is believed to be an important node for orchestrating dynein activity, where the nexin link is critical for transforming dynein-driven microtubule sliding into axonemal bending during beat. Mutations in different subunits have diverse phenotypic effects on cilia motility and the assembly of N-DRC.

CCDC164/DRC1

Mutations of *DRC1* in *Chlamydomonas* was shown to have more detrimental effects on motility and the assembly of both N-DRC and associated IDAs than other known *DRC* mutants.(201-203) Homozygosity mapping followed by direct screening for *CCDC164/DRC1* mutations identified mutations in 3 families with normal cilia cross sections but an almost total absence of N-DRC due to defective assembly of the nexin-dynein regulatory complex.(203)

CCDC65/DRC2

Direct sequencing of *CCDC65* exons in a cohort of 295 PCD patients identified 3 patients from two families with loss of function mutations. Normal cilia ultrastructure was noted in the patients with 5-15% of cilia cross sections showing microtubular disorganization.(98) In another study, homozygosity mapping in a PCD family revealed *CCDC65* as the only candidate gene that was highly conserved across all species with motile cilia. Direct sequencing of *CCDC65* exons identified a frameshift mutation in this family.(204) Morpholino knockdown of *ccdc65/FAP250* in zebrafish yielded a strong ciliopathy phenotype (pronephric cysts, laterality defects, body axis curvature and hydrocephalus).(98) FAP250, the *Chlamydomonas* ortholog of human CCDC65, is the DRC2 subunit of N-DRC.(205) The *Chlamydomonas* mutant strain *ida6* which has a mutation in *FAP250/CCDC65* shows motility defect, and defective assembly of IDAs and N-DRCs.(206, 207)

GAS8/DRC4

WES and targeted sequencing used to screen PCD patients, identified loss of function mutations in *GAS8*. TEM analysis revealed normal ciliary ultrastructure with noted mal-alignment of the peripheral doublets or axonemal disorganization in about 25.5% of the examined cilia cross sections, with or without IDA loss. DRC3/LRRC48 was shown to be lost in *GAS8* patients' cilia while CCDC39, DNALI1 and DNAH5 were still localized to the ciliary axonemes.(208, 209) It was also found that GAS8 is lost from the motile cilia in patients with mutations in *DRC1* and *DRC2*.(203, 204)

GAS8 expression profiling studies in human tissues showed that GAS8 is highly expressed in tissues with motile cilia: trachea, lungs and testes.(208) In mice, *Gas8* mRNA and protein were shown to be predominantly expressed in testes more than any other tissue. Gas8 protein was also detected along the sperm flagellum and in the cilia of the respiratory and fallopian tube epithelium.(210)

1.6.6 Molecular Ruler genes

CCDC39 and CCDC40 proteins were found to form a molecular ruler complex that supports the microtubular organization of the ciliary axoneme through promoting the accurate spacing of radial spokes along the axoneme with a regular 96-nm periodicity.(89)

CCDC39

Autozygosity mapping followed by sequencing in English sheepdogs with manifestations similar to human PCD identified truncating mutations in the *CCDC39* gene. TEM analysis of motile cilia from nasal and tracheal brushing biopsies from the affected dogs showed disorganization of the axonemal microtubule associated with loss of IDAs. Direct sequencing of *CCDC39* in PCD patients with similar ultrastructural defects to those noted in the dogs, identified mutations in *CCDC39* in 19 of 50 families screened.(211)

CCDC39 orthologs were only found in the species with motile cilia and its *Chlamydomonas* ortholog FAP59 was detected in the flagellar proteome. In mice, *Ccdc39* is highly expressed in tissues with motile cilia and also in the embryonic left-right organizer.(109, 168) Mouse ependymal cells carrying mutations in *Ccdc39* have shortened cilia with microtubular disorganization and loss of IDAs.(212) Mutations in *CCDC39* in human and *Chlamydomonas* lead to failure of the assembly of several IDA components and N-DRC leading to disorganization of microtubules.(211, 213)

CCDC40

Sequencing of *CCDC40* identified mutation in PCD patients with microtubular disorganization and loss of IDAs.(214) The ciliary ultrastructural defects caused by mutations in *CCDC39* and *CCDC40* are indistinguishable.(215, 216) Mutations in *CCDC40* resulted in lack of DNALI1, GAS8 and CCDC39 from the ciliary axonemes.(214)

In mice, mutations in *Ccdc40* lead to defective motile cilia phenotypes including laterality defects and hydrocephalus. Morpholino knockdown of *ccdc40* in zebrafish also resulted in laterality problems and a curved body axis

indicating a conserved role of *CCDC40* in cilia motility.(214) Studies in *Chlamydomonas* showed also that CCDC40 is required for the polyglutamylation of tubulin at the proximal part of the flagellum.(213)

1.6.7 Dynein assembly genes

Dynein arms are pre-assembled in the cytoplasm and then transported into the cilia.(15) This multi-step process requires cofactors and chaperones in addition to essential transport agents.(217) As the ciliary axoneme is highly conserved across species from the unicellular eukaryotes to mammals (218), delineation of the assembly mechanisms of the axonemal dyneins in the cytoplasm has been attempted using human genetics and PCD model organism studies.(217)

Mutations in 10 genes, implicated in dynein assembly, have all been identified to cause combined loss of inner dynein arms (IDAs) and outer dynein arms (ODAs). Interestingly, one protein previously considered for a potential role in dynein assembly, CCDC103, has recently been further examined and may not be a classic assembly factor.(219)

KTU/pf13/DNAAF2

KTU is the first cytoplasmic dynein pre-assembly factor identified in humans.(112) It was first analysed using a Medaka fish mutant, leading to impaired cilia motility associated with partial or complete loss of IDAs and ODAs.(92, 220) A *Chlamydomonas* mutant strain (*pf*13) also displays a deficiency of axonemal dyneins and immotile flagella.(221) PCD patients in whom *KTU* mutations were identified, presented with typical PCD manifestation and situs inversus totalis. High speed video microscopy showed complete immotility of the mutant respiratory cilia and sperms. TEM analysis revealed combined IDA + ODA loss from the ciliary axoneme.

It was also shown that HSP70 is one of the main proteins interacting with KTU.(112) This finding is consistent with the role of HSP70 in the assembly and transport of axonemal proteins.(222) In mouse tracheal epithelial cell culture (MTEC), it was shown that *KTU* expression is Foxj1-

dependent and that it plays an important role in multiciliogenesis downstream to Foxj1.(223)

LRRC50/DNAAF1

LRRC50 is the second identified human cytoplasmic dynein assembly factor. It harboured the first large genomic deletion found in PCD patients.(224) Zebrafish *lrrc50* was found to be expressed in all ciliated tissues and gene mutations lead to the development of cystic kidney disease with combined lack of dynein arms.(225) The *Chlamydomonas* ortholog, *ODA7* plays a role in the assembly of the axonemal dyneins and when mutated, leads to IDAs and ODAs loss.(226, 227) Mutations of *Dnaaf1* in mice cause hydrocephalus and postnatal lethality as well as laterality defects and respiratory symptoms. Interestingly, DNAH9 was found to be normally distributed in the airway motile cilia of the *Dnaaf1* mutant mice.(228)

Genotyping of 373 Chinese patients with neural tube defects (NTDs) and 222 matched controls with a multi-gene panel of 281 NTD candidate genes revealed 8 rare mutations in *DNAAF1*.(229) Moreover, it was shown that *DNAAF1* mutations alter the expression of genes related to neural tube closure in the patients' brain tissues compared to those of healthy controls.(229) However, It is still unknown how *DNAAF1* mutations can lead to the development of neural tube defects. In a forward genetics screening, heterozygous loss of function mutations in *dnaaf1* were found to predispose to testicular germ cell tumour (seminoma). In humans, germline heterozygous mutations in *DNAAF1* were identified in two families with history of seminomas suggesting potential novel tumour suppressor role of DNAAF1.(230) Recently, Verity L. Hartill et al (2018) have shown that mutations in *DNAAF1* can be associated with congenital heart defects in patients with no evidence of PCD.(231) These studies suggest that diverse roles can be played by DNAAF1. Further work is needed to elucidate these mechanisms and how they can interact during development.

DNAAF3

Mutations in *DNAAF3* were identified to cause complete absence of DNAH5, DNAH9, DNAI2 and DNALI1 from the ciliary axoneme with typical loss of IDAs and ODAs as demonstrated by TEM, suggesting that it plays a pivotal role in dynein assembly.(232) Mutations of the *Chlamydomonas* ortholog (*pf22*) cause immotility of the flagellum and disruption of IDA and ODA assembly.(221) Based on the *Chlamydomonas* model, Mitchison et al (2012) showed that abnormalities in the dynein heavy chains (HCs) resulted from defective *pf22,* are less severe than those caused by mutant *pf13* (*DNAAF2*) and *oda7* (*DNAAF1*) suggesting *DNAAF3* (*pf22*) plays a role in later stages of dynein assembly as it may be required for the dissociation of the co-chaperone complex after the accomplishment of the assembly process.(232)

In situ hybridization studies showed that the zebrafish *dnaaf3* is expressed in a pattern similar to that of other dynein assembly factors. Knockdown of *dnaaf3* in zebrafish leads to phenotypes consistent with defective cilia motility and PCD including abnormal axis curvature, renal cysts, hydrocephalus and laterality defects.(232)

DYX1C1/DNAAF4

Mutation screening of *DYX1C1* in PCD patients revealed several truncating mutations with a prevalent 3.5kb deletion causing loss of exon 7 in Caucasians through a likely founder effect. TEM analysis showed partial presence of axonemal ODAs with loss of IDAs. HSVM analysis showed some retained cilia motility with reduced frequency. DYX1C1 was shown to interact with DNAAF2 by co-immunoprecipitation study in HEK293 cells. The protein interactome of DYX1C1 identified in mouse trachea revealed interactions with many protein chaperones including HSP70 and HSP90.(233) Chen et al (2009) were able to show the upregulation of DYX1C1 in breast cancer and confirmed its interaction with HSP70 and HSP90.(234) Vaughan (2014) proposed a model where DYX1C1 interacts with DNAAF2 (KTU/pf13), forming an R2TP-like HSP90 co-chaperone complex for dynein assembly.(235)

Expression of *dyx1c1* was demonstrated in ciliated tissues of zebrafish embryos, while in adult zebrafish, the highest expression level was detected

in the testes. Knockdown of *dyx1c1* led to a phenotype of defective cilia motility with abnormal body curvature, laterality defects and kidney cysts. *Dyx1c1* morphant fish also has shortened cilia of variable degrees suggesting a possible role in regulating cilia length. TEM showed combined loss of IDAs and ODAs with persistent missing IDAs but with ODAs occasionally identified.(97) Tammimies et al (2016) showed that *DYX1C1* expression is regulated by *RFX* transcription factors through *X-box* promoter motifs and that knockdown of *RFX* transcription factors can alter *DYX1C1* expression. They also showed upregulation of expression of *DYX1C1* during ciliogenesis.(236)

PIH1D3/DNAAF6

PIH1D3 is the first X-linked gene identified to have a role in dynein assembly in humans, with different mutations causing PCD with combined IDA and ODA loss.(128, 129) Olcese et al (2016) reported an interaction between PIH1D3 and DNAI2 (ODA intermediate dynein chain) and proposed that PIH1D3 may play a role as a HSP90 co-chaperone in dynein assembly through an intermediate protein containing a TPR domain (possibly DNAAF4, in a similar fashion to DNAAF2).(128)

Previously, using *Pih1d3*$^{-/-}$ mutant male mice Fenglan Dong et al (2014) found that Pih1d3 can interact with Hsp70 and Hsp90 and concluded that it is essential for the cytoplasmic pre-assembly of the intermediate chains of ODAs in mouse sperm. On the other hand, this study could not show any protein interactions with Dnaaf1, Dnaaf2 and Dnaaf3.(237)

HEATR2/DNAAF5

Mutations in *HEATR2* were first identified by Horani et al (2012) as a cause of PCD in Amish communities.(238) HEATR2 is a highly conserved protein in organisms with motile cilia and flagella.(239) It is a cytoplasmic protein essential for ODA assembly as established by knockdown experiments in *Chlamydomonas* and human primary airway epithelial cells.(238)

It was found that *HEATR2* expression is RFX-dependent and it is upregulated during ciliogenesis. Knockdown experiments using the *Drosophila* ortholog in adult flies resulted in combined loss of ODAs and IDAs.

HEATR2 interacts with DNAI2.(240) These findings confirmed its role in ODA assembly while the mechanism underlying IDA loss is still unknown. A recent report showed that HEATR2 expression occurs early in ciliogenesis preceding known regulatory factors and that it co-localizes with SPAG1 and DNAAF2 to form an early dynein pre-assembly complex.(241)

LRRC6

Loss of function mutations of *LRRC6* were identified in PCD patients with combined loss of IDAs and ODAs.(242) LRRC6 protein contains N-terminal leucine rich repeats, an LRRcap domain (a motif occurring C-terminal to leucine-rich repeats) and a CS-like domain near the C-terminus.(243) It was also shown that DNAI1 and DNAH7 are lost from the cilia of PCD patients with *LRRC6* mutations, but DNAI1 was found to be trapped in the apical cytoplasm of ciliated cells implying that *LRRC6* plays a role in dynein assembly.(244)

In zebrafish, mutations in *seahorse/lrrc6l* affects cilia motility with laterally problems and pronephric cyst formation.(95) Moreover, a mutation of *tilB* (*LRRC*6 ortholog) in *Drosophila* was shown to cause loss of dynein arms and a male sterility phenotype.(245)

SPAG1

SPAG1 mutations were found in PCD patients with loss of IDAs and ODAs. It was shown that DNAH5 and DNALI1 are absent from the mutant ciliary axonemes implying a role of *SPAG1* in the pre-assembly of both dynein arms. Knockdown of the zebrafish ortholog of *SPAG1* leads to motile cilia related phenotypes with abnormal body curvature, hydrocephalus and abnormal heart looping.(246) It was previously shown that *SPAG1* expression is increased in pancreatic cancer and that it may promote motility of tumour cells.(247) Interactions of SPAG1 with both HEATR2 and DNAAF2 were recently shown, suggesting it has a role in the early steps of the cytoplasmic dynein pre-assembly process.(241)

ZMYND10/DNAAF7

Moore et al (2013) identified *ZMYND10* mutations in PCD patients with loss of IDAs and ODAs and discovered a putative hypomorphic allele p.Val16Gly where significant cilia motility is retained.(248) Immunofluorescence staining studies revealed an absence of DNAH5, DNAI2 and DNALI1 from the cilia in patients with *ZMYND10* mutations.(249)

Zmynd10 knockout mice show a typical PCD phenotype, indicating a possible role of ZMYND10 in regulating early steps of dynein pre-assembly through stabilization of DNAI1.(250) Moreover, the expression of *Zmynd10* in flies is dependent on *Rfx* and *fd3F* which are transcription factors regulating motile ciliary genes.(243, 251)

The interaction between ZMYND10 and LRRC6 was confirmed supporting their role in the dynein assembly process.(244, 252) A recent report identified a novel role for ZMYND10 in the dynein assembly process as a co-chaperone through conferring specificity of the FKBP8-HSP90 complex towards axonemal dynein proteins.(17)

C21orf59

TEM analysis of cilia from patients carrying mutations of *C21orf59* showed an absence of both IDAs and ODAs. Knockdown of *c21orf59/fbb18* in zebrafish and *planaria* resulted in motile ciliopathy phenotype with defects of ODA assembly. (98) Zebrafish *kur* mutants with mutated *c21orf59* have an evident defective cilia motility phenotype. It was also shown that there are interactions between c21orf59 and planar cell polarity (PCP) components implying its role in cilia polarity. A physical association was also shown between it and Lrrc6 indicating a role in dynein assembly process.(253)

CCDC103

Mutations in *CCDC103* were identified to cause PCD with variable loss of IDAs and ODAs and that's why its encoded protein was considered as a potential dynein assembly factor.(254) King et al (2015) have since shown that CCDC103 is tightly incorporated in the ciliary axoneme and that it binds directly with the microtubules. A model was proposed where CCDC103 forms

an array of high-affinity binding sites for the docking of outer dynein arms.(219) Recently, a hypomorphic p.His154Pro mutation in CCDC103 frequent within the UK-Pakistani PCD patients, was confirmed to affect CCDC103 protein oligomerization and to be associated with inconclusive clinical diagnostic results.(255)

1.6.8 Genes essential for multiciliogenesis

The assembly of motile cilia is regulated by several transcription factors including *FOXJ1* which is considered as the master multiciliogenesis gene.(256) The notch signalling pathway is also among the early regulators where reduced activity of this pathway promotes multiciliogenesis over mucous cell differentiation.(257) Mutations in PCD patients with oligocilia/motile cilia aplasia were identified in *MCIDAS* and *CCNO* genes.

MCIDAS/ MULTICILIN

MCIDAS is a transcriptional co-activator that was shown to promote multiciliogenesis in *Xenopus* skin and in ex vivo cultures of mouse airway epithelium.(73) WES and Sanger sequencing of *MCIDAS* exons in PCD patients with ciliary aplasia, identified loss of function mutations in *MCIDAS*. Mutations in *MCIDAS* were found to be associated with a severe reduction in basal bodies with occasional mislocalization of basal bodies in the cytoplasm. *MCIDAS* mutant cells failed to express *CCDC78* which is essential for deuterosomes formation.(74) In mice, it was shown that *Mcidas* acts upstream to *Foxj1* and that *Notch* signalling activation inhibits *Mcidas* expression.(258)

CCNO

WES identified homozygous frameshift mutation in *CCNO* in a large consanguineous family of Arab origin. Subsequently, screening for *CCNO* mutations in 3 consanguineous families with homozygous regions spanning *CCNO*, identified mutations in patients with complete absence or marked reduction of motile cilia. In vitro culture of the primary epithelial cells obtained from the affected individuals, identified a defect in primary multiciliogenesis. TEM analysis showed a marked reduction in centrioles and that the basal

bodies are mislocalized in the cytoplasm. Expression of *ccno* was shown to be upregulated by *mcidas* and repressed by *notch1* activation during multiciliogenesis in *Xenopus* embryos.(73)

In *Ccno* mutant mice, the multiciliated cells fail to generate deuterosomes and subsequently this leads to a significant reduction in the numbers of multiple motile cilia.(259) Mutations in *CCNO* were found to be associated with a higher incidence of hydrocephalus and female infertility in affected individuals.(70)

1.7 Aims and Objectives of the study

Primary ciliary dyskinesia (PCD) is a rare genetic disease with an autosomal or X-linked recessive inheritance pattern. The molecular basis of PCD is highly variable due to significant locus and allelic heterogeneity. PCD-causing mutations have been identified in 36 genes. These are estimated to account for the disease in about 70% of PCD patients, indicating additional genes are still to be discovered.

My project rationale was that genetic testing has the potential to overcome the diagnostic delay of other cilia-function diagnostic methods which are technically challenging, labour intensive and lead to inconclusive results in about 30% of cases. By identification of the whole genetic landscape of PCD, genetic testing can be developed into a gold standard, potentially even sole, test for PCD diagnosis. Genetics offers higher throughput improved testing for PCD, better disease management and counselling options for the affected individuals and their families and in future, will facilitate novel, personalized genetically targeted therapies.

In this study, I used targeted next generation sequencing followed by molecular, cell biology and model organism studies, to:

1) Evaluate the use of targeted multi-gene panel sequencing as a diagnostic tool for PCD with or without prior testing by other diagnostic methods.
2) Characterize the genetic landscape and the mutation spectrum of PCD among patients from various ethnic backgrounds.
3) Discover and characterize new genes linked to motor dynein assembly and transport in human.
4) Genetically evaluate affected individuals with a milder phenotype within the motile ciliopathy spectrum.
5) Define the genetic causes of combined non-motile and motile ciliopathy phenotypes.

Chapter 2 Subjects and Methods

2.1 Patient Enrolment

Participants were recruited through the UK national PCD diagnostic and management Services (London Royal Brompton Hospital, University Hospital Southampton, Leeds General Infirmary, Bradford Royal Infirmary, Birmingham Children's Hospital and Leicester Royal Infirmary). Diagnosis of PCD in the United Kingdom follows the European Respiratory Society (ERS) guidelines. Patients were also recruited by clinical partners in Portugal (Hospital de Santa Maria, Centro Hospitalar Lisboa Norte, Lisbon) where PCD diagnosis was also confirmed following the ERS guidelines.(78) Arab families with suggestive clinical manifestations of PCD were recruited through the clinical collaborators in Palestine (Makassed Hospital, East Jerusalem) and Egypt (Alexandria University Paediatrics Hospital, Alexandria). TEM analysis for Palestine patients was done at PCD unit, University Hospital Southampton. TEM analysis for some of the Egyptian patients was done at London Royal Brompton Hospital.

Written informed consent was obtained from all participants or their guardians. The study was ethically approved by the London Bloomsbury Research Ethics Committee under approval number 08/H0713/82 and from the respective ethics committees of the collaborating institutions. It was registered at UCL Great Ormond Street Institute of Child Health R&D office as project number 11MM03.

2.2 Materials

2.2.1 Equipment

Name	Supplier
Sorvall Legend RT Refrigerated Centrifuge	Thermo Scientific
Benchtop PCR plate centrifuge	Labnet International
Eppendorf Microcentrifuge	ThermoFisher Scientific
NanoDrop ND-1000 Spectrophotometer	Thermo Scientific
Qubit® 2.0 Fluorometer	ThermoFisher Scientific
MJ Research Tetrad Thermal Cycler	GMI
Agilent 4200 TapeStation	Agilent
NextSeq 550 system	Illumina
Electrophoresis systems	Bio-Rad
UVP Trans-illuminator	Cole-Parmer
LSM 710 inverted confocal microscope	Zeiss
CFX96 touch real-time PCR system	Bio-Rad
JEM-1400Flash TEM	JEOL
DM60 Leica microscope	Leica Microsystems GmbH
Innova® 44 Shaking Incubator	Eppendorf
Multiskan™ FC Microplate Photometer	ThermoFisher Scientific
Trans-Blot SD Semi-Dry Transfer Cell	Bio-Rad
TC20™ Automated Cell Counter	Bio-Rad

2.2.2 Kits and buffers

Name	Supplier
Oragene DNA collection kit	DNAgenotek
prepIT®•L2P reagent	DNAgenotek
Qubit dsDNA assay kit	ThermoFisher Scientific
SureSelectQXT Target Enrichment Kit	Agilent Technologies
Agencourt AMPure XP Kit	Beckman Coulter Genomics
Dynabeads MyOne Streptavidin T1	ThermoFisher Scientific
1X Low TE Buffer	ThermoFisher Scientific
NextSeq 500/550 High Output Kit v2 (150 cycles)	Illumina
BIOTAQ PCR kit	BIOLINE
10X TAE Buffer	Invitrogen
KAPA HiFi HotStart PCR Kit	KAPABiosystems
6x DNA Loading Dye	ThermoFisher Scientific
ExoSAP-IT PCR Product Cleanup Reagent	Affymetrix
Paraformaldehyde 16% Solution	VWR
Dulbecco's Phosphate Buffered Saline	ThermoFisher Scientific
Triton X-100	Sigma-Aldrich

Skim Milk Powder	Sigma-Aldrich
Antifade Mounting Medium with DAPI	Vector Laboratories
Dulbecco's Modified Eagle Medium	ThermoFisher Scientific
Fetal Bovine Serum (FBS)	ThermoFisher Scientific
Penicillin-Streptomycin (10,000 U/mL)	ThermoFisher Scientific
Trypsin-EDTA (0.05%), phenol red	ThermoFisher Scientific
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich
BEBM Basal Medium	Lonza
SingleQuot Kit Suppl. & Growth Factors	Lonza
PureCol, bovine collagen	Nutacon
Lipofectamine 2000/3000 Reagents	ThermoFisher Scientific
Opti-MEM Reduced Serum Medium	ThermoFisher Scientific
RNALater	Sigma-Aldrich
Trizol	Qiagen
Omniscript Reverse Transcription Kit	Qiagen
High-Capacity RNA-to-cDNA™ Kit	ThermoFisher Scientific
iQ™ SYBR® Green Supermix	Bio-Rad
SOC Medium	New England BioLabs
Lysogeny broth (LB) Agar	Sigma-Aldrich
LB Broth	Sigma-Aldrich
QIAprep Spin Miniprep Kit	Qiagen
Plasmid Maxiprep Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
Monarch® DNA Gel Extraction Kit	New England BioLabs
Glycerol	Sigma-Aldrich
cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail	Roche
Pierce™ BCA Protein Assay Kit	ThermoFisher Scientific
2x Laemmli Sample Buffer	Bio-Rad
PVDF Membranes for Western Blotting	ThermoFisher Scientific
Bovine serum albumin (BSA)	Sigma-Aldrich
Pierce ECL Western Blotting Substrate	ThermoFisher Scientific
Dynabeads Protein G for Immunoprecipitation	ThermoFisher Scientific
Anti-FLAG M2 Affinity Gel	Sigma-Aldrich
Bio Herbe de Blé	GSE-Vertrieb GmbH

2.2.3 Eukaryotic cells

Name	Supplier	Growth Medium
HEK293T	ATCC	DMEM with 10% FBS and 1% P/S
NHBE-BMI	Prof Stephen Hart	BEGM or ALI media

2.2.4 Bacterial cells

Strain	Supplier	Genotype
JM109 Competent Cells	Promega	endA1, recA1, gyrA96, thi, hsdR17 (rk–, mk+), relA1, supE44, Δ(lac-proAB),[F´traD36, proAB, laqIqZΔM15]
DH5α™ Competent Cells	ThermoFisher Scientific	F– Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1
XL10-Gold Ultracompetent Cells	Agilent Technologies	TetR Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F´ proAB lacIq ZΔM15 Tn10 (TetR) Amy CamR]
HT115-DE3 strain	*Paramecium* Lab, CNRS	[F-, mcrA, mcrB, IN(rrnD-rrnE)1, rnc14::Tn10(DE3 lysogen: lavUV5 promoter -T7 polymerase]

2.2.5 Plasmids

Vector	Transgene	Accession No.	Supplier
pcDNA3.1(+)-C-Myc	*IFT46*	NM_001168618.1	Genscript
pcDNA3.1(+)-C-eGFP	*C11orf70*	NM_032930.2	Genscript
pcDNA3.1+/C-(K)-DYK	*C11orf70*	NM_032930.2	Genscript
pcDNA3.1+/C-(K)-DYK	*IFT74*	NM_001099223.2	Genscript
L4440	*Paramecium C11orf70* RNAi	GSPATG00011350001	*Paramecium* Lab, CNRS
L4440	*Paramecium DNAH9* RNAi	GSPATG00011644001 GSTATG00038424001	*Paramecium* Lab, CNRS

2.2.6 Primary antibodies

Antigen	Antibody	Host species	Source	Application
Acetylated Tubulin	T7451	Mouse	Sigma-Aldrich	IF (1:500)
Acetylated Tubulin (Lys40)	YF488	Mouse	Proteintech	IF (1:500)
Conjugated Acetylated Tubulin	YF488-66200	Mouse	Proteintech	IF (1:500)
DNAH5	HPA037470	Rabbit	Sigma-Aldrich	IF (1:800)
DNALI1	HPA053129	Rabbit	Sigma-Aldrich	IF (1:200)
RSPH4A	HPA031196	Rabbit	Sigma-Aldrich	IF (1:400)
GAS8	HPA041311	Rabbit	Sigma-Aldrich	IF (1:500)
CCDC114	HPA042524	Rabbit	Sigma-Aldrich	IF (1:500)
DNAH9	HPA052641	Rabbit	Sigma-Aldrich	IF (1:500)
DNAI1	HPA021843	Rabbit	Cambridge Bioscience	IF (1:200)
FLAG	F1804	Mouse	Sigma-Aldrich	IB (1:1000)
IFT74	27334-1-AP	Rabbit	Proteintech	IF (1:500) IB (1:1500)
IFT81	11744-1-AP	Rabbit	Proteintech	IF (1:200) IB (1:500)
IFT88	13967-1-AP	Rabbit	Proteintech	IF (1:200) IB (1:300)
IFT46	HPA037909	Rabbit	Cambridge Bioscience	IB (1:100)
DYNC2LI1	15949-1-AP	Rabbit	Proteintech	IF (1:300)
WDR19	13647-1-AP	Rabbit	Proteintech	IF (1:200)
TEKT1	18968-1-AP	Rabbit	Proteintech	IF (1:200)
Polyglutamylated Tubulin	PolyE	Rabbit	A gift from J. Cohen	IF (1:500)
GFP	IgG	Rabbit	Interchim, France	IF (1:500)
GAPDH	G8795	Mouse	Sigma-Aldrich	IB (1:1000)

2.2.7 Secondary antibodies

Antibody	Host Species	Source	Application
Alexa 488, 594 – conjugated anti-rabbit IgG (H+L)	Goat	Invitrogen	IF (1:1000)
Alexa 488, 594 – conjugated Anti-mouse IgG1	Goat	Invitrogen	IF (1:1000)
HRP-conjugated Anti-mouse IgG	Goat	Bio-Rad	IB (1:2000)
HRP-conjugated Anti-rabbit IgG	Goat	Sigma-Aldrich	IB (1:2000)

2.2.8 Recipes

Solution	Composition
RBCs lysis buffer	15.405g Ammonium Chloride (0.144M NH_4CL),1.68g Sodium Bicarbonate ($NaHCO_3$), up to 2 litres distilled water
Nuclei lysis buffer	10 mL of 1 M Tris HCl (pH 8.1), 80 mL of 5 M NaCl, 4 mL of 0.5M EDTA (pH 7.5-8.0), 100g of Sodium Dodecyl Sulphate (SDS), up to 1 L distilled water
ALI medium	50:50 BEBM(v/v) + DMEM (v/v) (400mM glucose), 1x SingleQuots supplement kit, 100 U/mL penicillin, 100 mg/mL streptomycin, Retinoic acid (100nM)
Nonidet-P40 (NP-40) lysis buffer	150 mM sodium chloride, 0.5% NP-40, 50 mM Tris, pH 8.0
SDS-PAGE running buffer	25 mM Tris, 190 mM glycine, 0.1% SDS, pH 8.3
Towbin Transfer buffer	25 mM Tris, 190 mM glycine (20% methanol), pH 8.3
***Paramecium* Bio Herbe de Blé (BHB) Medium Stock**	50 g BHB powder + 750 ml H_2O, Boiled for 15 min, filtered on a thick layer of gauze and cotton and then adjust to 1 L with H_2O. Autoclaved at 120°C, for 20 minutes
Working *Paramecium* BHB Medium	Add 200 ml of stock solution to 1 L of Phosphate (17.5g $Na_2HPO_4.12H_2O$ in 1 L H_2O) and adjust to 20 L with H_2O. Autoclave at 120°C, for 20 minutes
***Paramecium* Dryl's Solution**	20 ml 0.1 M Sodium Citrate, 10 ml 0.1 M NaH_2PO_4, 10 ml 0.1 M Na_2HPO_4, 15 ml 0.1 M $CaCl_2$, up to 1 L H_2O

2.3 Methods

2.3.1 DNA Extraction

Genomic DNA was extracted from the whole blood samples collected in tubes containing EDTA anticoagulant using the following in-house protocol. All recipes are detailed in **Section 2.2.8**. Each blood sample was mixed with equal volume of RBCs lysis buffer and left at room temperature for 10 mins then centrifuged at 4000g for 10 mins and the supernatant was discarded (repeated twice). The pellet was re-suspended in nuclei lysis buffer, vortexed and incubated overnight in a water bath 60°C. 1 mL of 5M NaCl and 1 mL of chloroform were added and mixed well then centrifuged at 4000g for 30 mins with the brake off. The upper layer was transferred to a new tube and 1.2x volume of Isopropranolol was added slowly and gently mixed and then centrifuged for 5 mins to pellet the DNA. Afterwards, the DNA pellet was washed with 70% ethanol and centrifuged for 1 min. The supernatant was discarded and the pellet left to dry at room temperature, then the DNA was re-suspended in nuclease free water.

Saliva samples were collected using Oragene DNA collection kit. Genomic DNA was extracted using an ethanol precipitation protocol and prepIT®•L2P reagent (DNAgenotek) according to the manufacturer's protocol. In brief, the samples were incubated in water bath 50° C for an hour, then 1/25th volume of PT-L2P reagent was added with a vortex applied to mix, followed by incubation on ice for 10 mins. Centrifugation followed at 4000g for 10 mins, then the clear supernatant was transferred to a new tube. The DNA was precipitated by adding 1.2x volume of 95% ethanol with high speed centrifugation for 10 mins. 1mL 70% ethanol was used to wash the DNA pellet and the DNA was re-suspended in nuclease free water.

DNA concentration was quantified using a Thermo Scientific NanoDrop™ 1000 Spectrophotometer. 260/280 and 260/230 ratios were recorded to assess the purity of DNA.

DNA samples were stored in a -20 freezer until further usage.

2.3.2 Targeted Next Generation Sequencing

2.3.2.1 Panel design

A 'Motile ciliome' panel was designed to include genes that when mutated lead to primary ciliary dyskinesia along with other candidate genes selected with close relevance to cilia motility based on extensive literature review and candidate recommendations from other research groups working on cilia in zebrafish (Sudipto Roy, National University of Singapore, Singapore), mouse (Cecilia Lo, University of Pittsburgh School of Medicine, USA), Drosophila (Andrew Jarman, University of Edinburgh, UK). The motile ciliome panel underwent two design iterations. The first design included 651 genes and a revised second design included 321 genes halving the number of genes to remove repeated genes and to exclude biologically irrelevant genes. (**Figure 2-1 and Figure 2-2, Appendix table A-0-1 and A-0-2**)

Enrichment of the motile ciliome genes was performed using the SureSelectQXT Target Enrichment System (Agilent Technologies, Santa Clara, CA, USA). The Agilent SureDesign tool was used for panel design to capture all coding regions of the included genes and 25 bp at the exon-intron boundaries. The total numbers of targets, probes, capture size of each panel version and the panel design parameters are summarized in **Table 2-1.**

Table 2-1 NGS multi-gene panel design and mean coverage of targets

	651 gene panel	**321 gene panel**
Genome Build	(H. sapiens, hg19, GRCh37, February 2009)	
Targets	651	321
Region	Coding Exons	
Region Extension	25 bases from 3' end and 25 bases from 5' end	
Region Size	2.182Mbp	1.265 Mbp
Total Probes	50630	41785
Total Probe Size	3.536 Mbp	1.757 Mbp
Probe Tiling Density	2x	
Probe Strand	Sense	

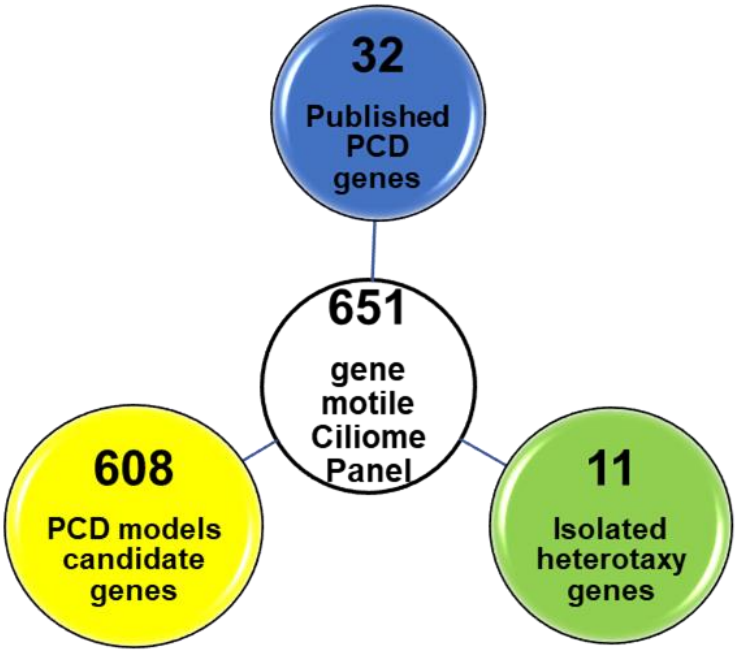


Figure 2-1 651 Multi-gene ciliome panel design

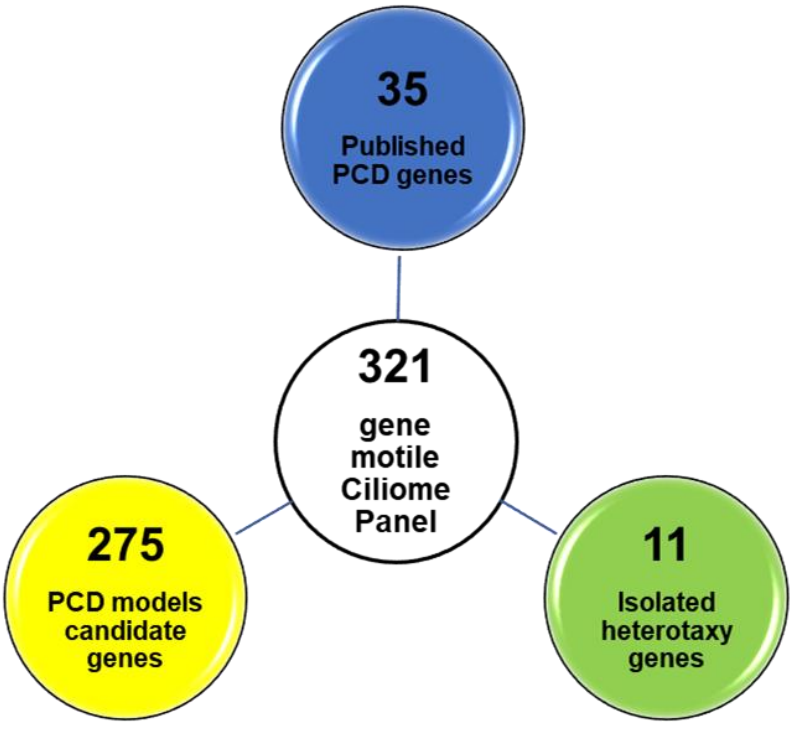


Figure 2-2 321 Multi-gene ciliome panel design

2.3.2.2 Library preparation

Library preparation was performed using the SureSelectQXT Target Enrichment Kit (Agilent Technologies) according to the manufacturer's protocol.

DNA samples were quantified using Qubit dsDNA assay kits (ThermoFisher) and then normalized to 25ng/µl in a 10 µl final volume in two dilution steps. At first, 2 µl of each DNA sample was used to quantify DNA concentration using the Qubit BR (Broad Range) Assay. Each sample was diluted with ddH_2O (double-deionized water) to achieve 50ng/µl. 2 µl of each diluted DNA sample was used to quantify again using the Qubit BR Assay and then each sample was diluted with ddH_2O to achieve 25ng/µl. 2 µl of each diluted DNA sample were used to quantify the sample again using the Qubit HS (High Sensitivity) Assay and then the concentrations were adjusted to ensure the final concentration is within +/-2ng of 25ng/µl in a final volume of 10 µl.

Approximately 50 ng of DNA was used for the enzymatic fragmentation and adaptor ligation in a single reaction. In a PCR plate on ice, 17 µl QXT buffer, 2 µl of each normalized DNA sample and 2 µl QXT enzyme mix were added in this exact order and mixed by pipetting 8-10 times, sealed with dome strip caps, vortexed for 20 sec followed by a brief spin and the plate was placed into the thermal cycler. For enzymatic fragmentation, the thermal cycling protocol included incubation at 45°C for 10 minutes then at 4°C for 1 min. After fragmentation, 32 µl 1X stop solution was added to each sample, the mixture was vortexed for 5 secs followed by a brief spin and incubation at room temperature for 1 min.

AMPure XP beads (Beckman Coulter Genomics) were used for the purification of the adaptor-tagged fragmented DNA samples. First, AMPure XP beads were taken out from the fridge and left at room temperature for 30 mins during the fragmentation reaction and vortexed vigorously to ensure even distribution of the beads within the solution. 52 µl of beads solution was added to each sample, vortexed for 5 sec and incubated at room temperature for 5 mins. Afterwards, the plate was put on a magnet for 3-5 mins and the

supernatant was removed and discarded. The beads pellets were then washed twice with 200 µl 70% ethanol. The samples were dried on a thermal cycler at 37°C for 1-3 min with the lid open. 11 µl of nuclease free water was added to the pellet, then was vortexed and spun down, incubated at room temperature for 2 mins then placed on a magnet for 2 mins. The supernatant was then transferred to a new PCR plate.

For amplification of adapter-tagged libraries, pre-capture PCR Reaction mix was prepared with reagents provided within the SureSelectQXT Target Enrichment kit. For each sample, a PCR reaction mix was added including: 25 µl nuclease free water, 10 µl Herculase II 5× Reaction Buffer, 0.5 µl 100 mM dNTP Mix (25 mM each dNTP), 2.5 µl DMSO, 1 µl SureSelectQXT Primer Mix and 1 µl Herculase II Fusion DNA Polymerase. The thermal cycling was performed as follows: 1 cycle for 2 mins at 68°C then 1 cycle for 2 mins at 98°C followed by 8 cycles (98°C for 30 sec, 57°C for sec and 72°C for 1 min) then 1 cycle at 72°C for 5 mins.

AMPure XP beads were used for purification of the amplified adaptor-tagged DNA fragments as mentioned above and 13 µl of the supernatant was transferred to a fresh PCR plate. 2200 TapeStation and D1000 ScreenTape (Agilent life technologies) were used to assess the quantity and the quality of the fragmented libraries. 1 µl of samples were mixed with 3 µl of D1000 reaction buffer. A peak size between 245-325 bp was achieved. (**Figure 2-3**)

For hybridization to capture libraries, 500-750 ng of the fragmented libraries were used. 5 µl of QXT Fast Blocker Mix was added to wells of DNA aliquots and pipetted up and down 8-10X then vortexed high speed for 5 secs which was followed by a brief spin. The plate was then transferred to the thermal cycler and the QXT Hybridization program was started, consisting of the following:

Segment	Number of Cycles	Temperatures	Time
1	1	95°C	5 minutes
2	1	65°C	10 minutes
3	1	65°C	1minute (PAUSE)
4	60	65°C	1 minute
		37°C	3 seconds
5	1	65°C	Hold

During the first two segments, a hybridization mix was prepared to include 2 µl 25% RNAase block solution, 2 µl of SureSelect Custom capture library, 6 µl QXT fast hybridization buffer and 3 µl of nuclease free water for each sample. When segment 3 started, the thermal cycler was paused and the hybridization mix was added to each sample, mixed by pipetting up and down 8-10 X, vortexed, spun down and returned to the thermal cycler to resume the program.

Dynabeads MyOne Streptavidin (Life Technologies) were used to capture the hybridized libraries. Prior to capture, streptavidin beads were vortexed vigorously and 50 µl of the beads (used for each hybridized sample) were washed three times with 200 µl binding buffer, re-suspended in 200 µl of binding buffer then mixed with 30 µl of hybridized sample using plate mixer at 1800 rpm for 30 mins. During mixing, wash buffer 2 was pre-warmed on the thermal cycler at 65°C (3x 200 µl for each sample). After 30 mins mixing, liquid was collected to bottom of wells by swift sharp flicking motion. The plate was put onto a magnet for 1 min before removing and discarding the supernatant. The beads were then re-suspended in 200 µl wash buffer 1 and mixed by pipetting 8-10 X, vortexed for 8 secs then the liquid was collected to the bottom of the wells by a sharp flicking motion again. The plate was put on a magnet for 1 min before removing and discarding the supernatant. The beads were then washed with wash buffer 2 for 3 times with incubation at 65°C for 10 mins at each wash. After removal of the supernatant after the final wash, the beads were re-suspended in 23 µl nuclease free water and placed on ice.

For multiplexing of samples in the sequencing run, the captured libraries were PCR amplified to add index tags followed by purification using AMPure XP beads. The indexing PCR reaction was prepared for each sample to include: 13.5 µl Nuclease-free water, 10 µl Herculase II 5× Reaction Buffer, 0.5 µl 100 mM dNTP Mix, 1 µl Herculase II Fusion DNA Polymerase with 1 µl of each P7 and P5 primers. The post-capture PCR program was as follows:

Segment	Number of Cycles	Temperatures	Time
1	1	98°C	2 minutes
2	12	98°C	30 seconds
		58°C	30 seconds
		72°C	1 minute
3	1	72°C	5 minutes

After purification using AMPure XP beads as mentioned above, the indexed libraries were assessed using 2200 TapeStation and high sensitivity D1000 ScreenTape (2 µl of sample mixed with 2 µl of high sensitivity D1000 buffer) and Qubit high sensitivity dsDNA assay. The peak of DNA fragment size was achieved between 325 and 450 bp. (**Figure 2-4**)

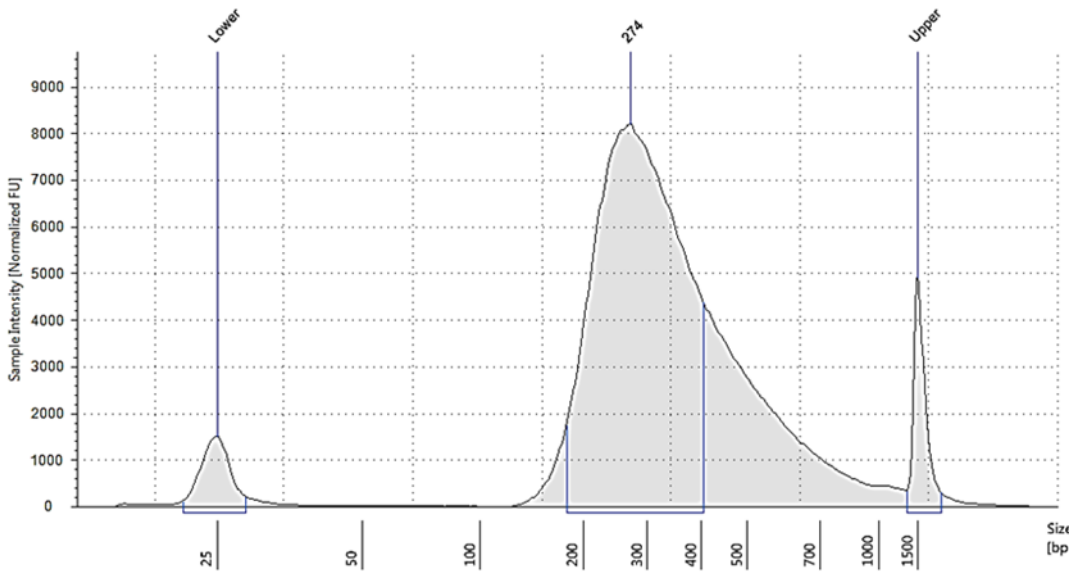


Figure 2-3 Representative electropherogram showing pre-capture analysis of amplified library DNA using the 2200 TapeStation with D1000 ScreenTape

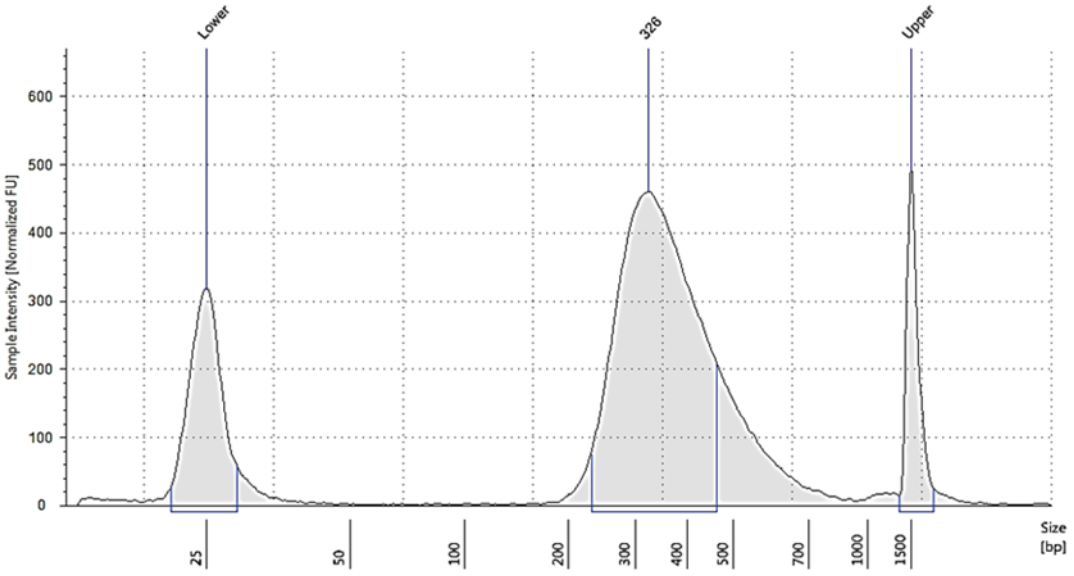


Figure 2-4 Representative electropherogram showing post-capture analysis of amplified indexed library DNA using the 2200 TapeStation with a High Sensitivity D1000 ScreenTape

2.3.2.3 Sequencing

48 indexed libraries were multiplexed per sequencing run using an Illumina HiSeq 2500 and HiSeq PE Rapid Cluster Kit v2 or NextSeq 550 platforms and NextSeq 500/550 High Output v2 kit (150 cycles) (Illumina, Inc., CA, USA) according to the manufacturer's protocol. Sequencing was done at the North-East Thames Regional Genetics Service Laboratories. The quality of the sequencing run was assessed during the run using Illumina sequencing analysis viewer software. (**Figure 2-5**)

2.3.2.4 Data analysis

Handling of the sequence data to call DNA variants was done by the bioinformaticians at the North-East Thames Regional Genetics Service. In brief, data in Fastq format from the NextSeq was subject to genome-wide alignment to the human genome reference (GRCh37/hg19) using BWA-mem (0.6.1-r104), variant calling using Freebayes and variant annotation using Alamut-batch (Interactive Biosoftware, Llc., Rouen, France) softwares through an in-house pipeline.(260) A search for larger insertion/deletion mutations and copy number variants (using ExomeDepth software) was separately performed.(261)

For manual variant filtration and prioritization, VCF files were converted to Excel format. Variants were filtered for significance to produce variant lists of interest in each patient with a recessive inheritance pattern (homozygous or compound heterozygous) and expected minor allele frequency (MAF) of <1%. The minor allele frequency was assessed in various control databases including ExAc databases, (262) 1,542 individuals in the Born-in-Bradford (BinB) cohort of UK South Asians (263, 264) and the almena database of genetic variants in Middle East and North African individuals (265). Only coding and splicing variants (±5bp from exon-intron junction) were retained. In the first analysis, synonymous coding variants were filtered out. In cases where no other variants were prioritized, synonymous variants in known PCD genes were re-visited and evaluated for possible splicing defects.

Variants were evaluated based on their predicted pathogenicity scores and impact at the protein level using *in silico* softwares including Polyphen-2, SIFT, Mutation Taster, Human Splicing Finder (HSF) and Combined Annotation Dependent Depletion (CADD) scores. The guidelines of the American College of Medical Genetics was used for pathogenicity scoring of the prioritized variants.(117) (**Figure 2-6**) In general, Class 3-5 variants according to this definition were prioritized for further confirmation and segregation studies.

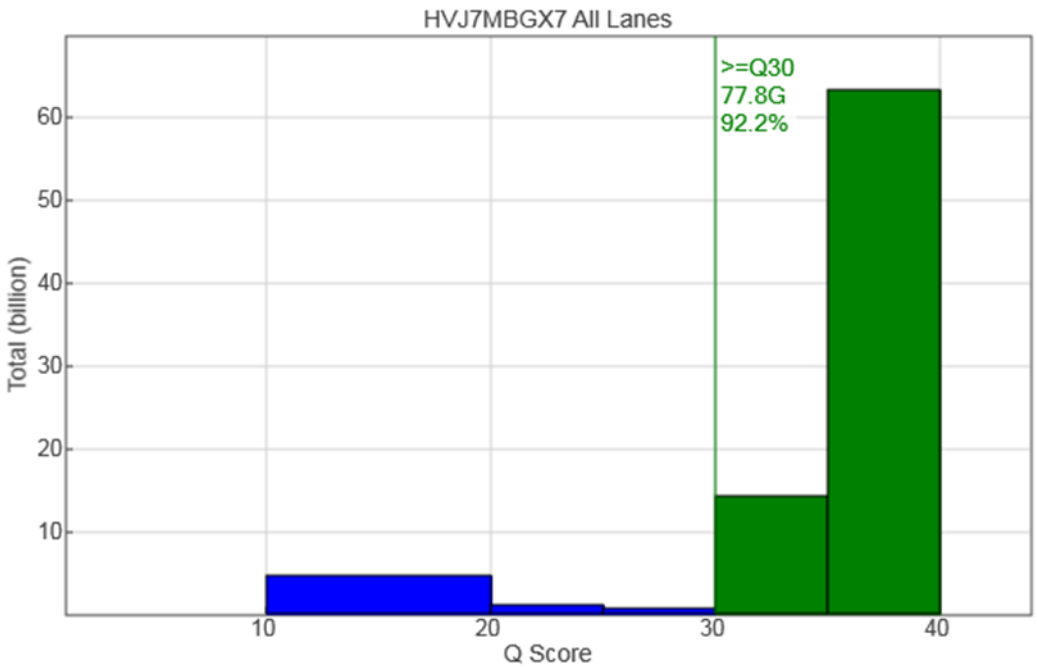


Figure 2-5 Representative snapshot of Q-score distribution pane in the sequencing analysis viewer software showing the number of reads by quality score

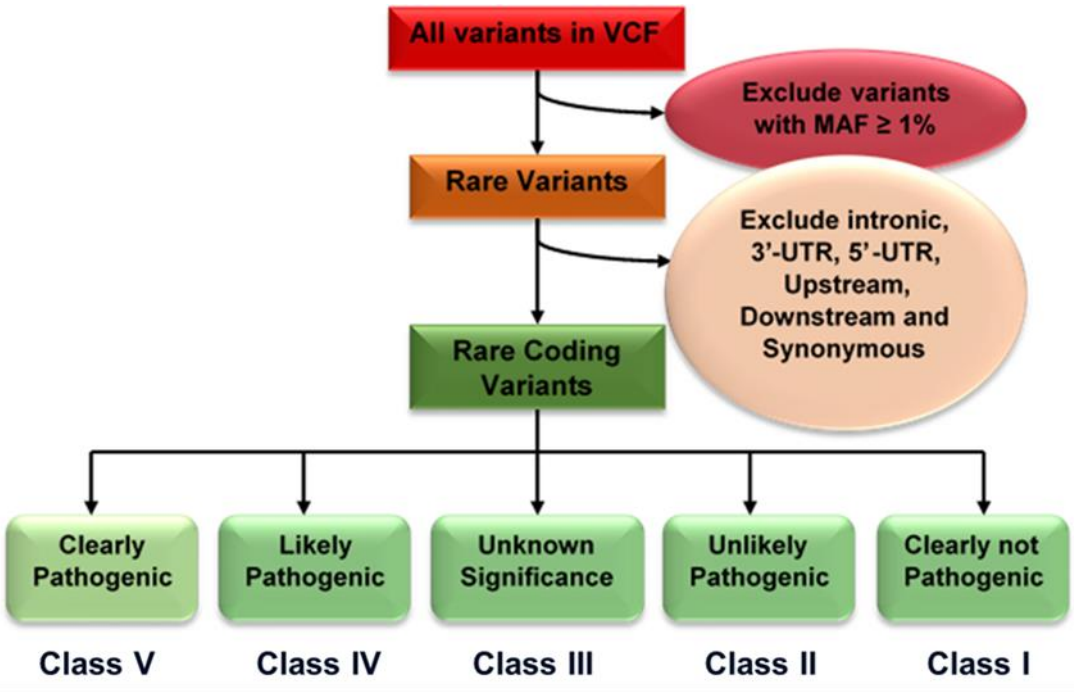


Figure 2-6 Variation filtration and prioritization workflow

2.3.3 Sanger confirmation and segregation studies

2.3.3.1 Primer design

The Primer-Blast tool within NCBI (www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to design specific primers flanking the exon of each prioritized variant. The primer pair specificity was checked against the human genome reference and their stringency was increased by increasing the number of mismatches between the primers and the unintended targets towards the 3' end of the primers.

2.3.3.2 Polymerase Chain Reaction (PCR)

Amplification of the DNA fragment harbouring the variant of interest was performed using the BIOTAQ PCR kit (BIOLINE) according to the manufacturer protocols. In brief, the PCR reaction volume was setup to 25 µl reaction through mixing 2.5 µl of 10x NH4 Reaction Buffer, 2 µl of 50mM MgCl2 Solution, 2.5 µl of 100mM dNTP Mix and 0.25 µl BIOTAQ enzyme with 200nM of each forward and reverse primers and 10-100ng of the genomic DNA and nuclease free water was added to reach a final volume of 25 µl. The thermal cycling conditions were set up to include an initial denaturation step at 95°C for 5 mins followed by 40 cycles (denaturation for 30 sec at 95°C, annealing for 30 sec with a temperature differs according to the primers Tm and extension at 72°C for 1 min per 1kb of the product size), followed by a final extension step at 72°C for 10 mins.

2.3.3.3 Agarose Gel Electrophoresis

1-2 % of agarose gels were prepared to separate PCR products and confirm there was no non-specific amplification. The appropriate amount of agarose (Hi-Pure Low EEO agarose, Biogene) was weighed and added to 1x Tris Acetic Acid EDTA (TAE) buffer (Invitrogen) enough for the gel tray. This was then heated in the microwave to melt the agarose and cooled to approximately 50°C. 10 mg/ml ethidium bromide was added to a final concentration of 0.2

µg/ml then the gel was poured and left to solidify for about 30 mins and then placed in a tank and covered with 1x TAE buffer. Samples mixed with a loading dye were loaded along with appropriate DNA ladder; either 1kb DNA Ladder or 100bp DNA Ladder (Promega). Electrophoresis was carried out at 140V until the dye was run to a sufficient distance for appropriate band size separation. The bands were then visualized by exposure to UV light.

2.3.3.4 Gel purification

A clean razor blade was used to cut the desired DNA bands from the agarose gel and the Monarch DNA Gel Extraction Kit (New England Biolabs) was used to purify DNA according to the manufacturer's protocol.

2.3.3.5 PCR cleanup

PCR products were cleaned up using ExoSAP-IT (Affymetrix) according to the manufacturer's protocol. 10 µl of PCR product was mixed with 2 µl of ExoSAP-IT and incubated at 37°C for 15 minutes followed by inactivation of ExoSAP-IT by heating to 80°C for 15 minutes.

2.3.3.6 Sanger sequencing

Sanger sequencing of DNA samples from the patients and available family members was outsourced at *Eurofins Genomics* using a Mix2Seq kit. The cleaned up PCR products were pre-mixed with 2 µl of the sequencing primer at a concentration of 10uM for a final volume of 17 µl.

2.3.3.7 Visualization and data analysis

Sequencing data was aligned to the human genome assembly GRCh37 and viewed using SnapGene viewer 4.1.9 (GSL Biotech LLC, Chicago, USA) or Sequencher software 5.2.4 (Gene Codes Corporation, Ann Arbor, MI, USA).

2.3.4 Immunofluorescence staining

2.3.4.1 Preparation of slides

Multiciliated epithelial cells were obtained by nasal brushing biopsies performed within the clinical diagnostic settings in the collaborating institutions by qualified personnel. The brushing samples of patients and normal controls were dislodged in serum free medium (M199) (Sigma-Aldrich Co. Ltd., Dorset, UK) and a couple of drops from the suspension were added on a glass slide using a Pasteur pipette. The slides were left to dry and then transferred to our laboratory to be stored at -80°C freezer until needed.

2.3.4.2 Immunostaining procedures

Prior to immunostaining, the slides from the patient of interest and a normal control were collected from the -80°C freezer and left to defrost. Paraformaldehyde 4% was added to the slides and left for 15 mins for fixation. 5% fat-free skimmed milk powder in a washing buffer (PBS + 0.1% Triton solution) was used as a blocking solution for 1 hour, followed by incubation with the diluted primary antibodies in a washing buffer for 2 hours at room temperature, then diluted Alexa Fluor secondary antibodies were added and incubated for 45 minutes. Each step was followed by 3 washes using PBS + 0.1% Triton solution. The residual solution after washing was removed by blotting the edges of each slide using a paper towel and anti-fade mounting medium with DAPI (Vector Laboratories Ltd. Peterborough, UK) was added then each slide was covered by a coverslip. The slides were left to dry in the 4°C fridge before imaging.

2.3.4.3 Confocal imaging and image processing

Confocal slice images of the patients and control ciliated cells and Z projections of the micro-injected *Paramecium* cells were obtained using Zeiss LSM 710 inverted confocal microscope and 63x Plan-Apochromat NA 1.4 WD 190 um Oil objective with DIC and laser lines for excitations at 405, 488 and

594 nm. Images were processed with the ZEN software suite. ImageJ software was used to create stacks and montages from 16-bit confocal images and to set scale bars.

2.3.5 Transmission Electron Microscopy

Transmission electron microscopy (TEM) analysis of the cilia cross sections was done whenever available within the diagnostic settings in the collaborating institutions. For TEM, nasal brush biopsies were fixed in 4% glutaraldehyde followed by processing and quantification according to a standard, well-established protocol.(266)

2.3.6 High Speed Video Microscopy

Cilia beat frequency and cilia beat pattern were assessed in ciliated epithelial cells obtained by nasal brushing then re-suspended in M199 medium by light microscopy at 37°C. Cilia beating was recorded using a 100X oil immersion lens by high speed video camera within the PCD diagnostic units of the collaborating institutions.(267)

2.3.7 Cell culture

2.3.7.1 Submerged culture

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Fetal Bovine Serum and 1% Penicillin-Streptomycin (10,000 U/mL). Primary normal human bronchial epithelial cells transduced by BMI-1 (B-cell-specific Moloney murine leukemia virus integration site 1) were grown in bronchial epithelial growth media (BEGM) on PureCol, bovine collagen coated cell culture flasks.

2.3.7.2 Air-Liquid Interface (ALI) culture

BMI-1 transduced NBHE cells (NBHE-BMI) in BEGM were seeded on 6.5 mm Corning Transwell polyester membrane cell culture inserts (0.4 μm pore size) at a seeding density of 0.5 X 10^6 cells/insert. BEGM was also added to the basolateral compartment and incubated for 48 hours when the apical and basolateral media were removed and ALI medium was added to the basolateral compartment only so the apical portions of the cells are exposed to air to stimulate multiciliogenesis for about 4 weeks until full ciliogenesis was reached with the basolateral ALI medium being changed every 2-3 days.(268, 269) The ALI medium recipe is detailed in **Section 2.2.8**.

2.3.8 RNA extraction

RNA was extracted from nasal brushing biopsies and from the ALI-culture transwell inserts using Trizol reagent (Qiagen Inc., Hilden, Germany). The dislodged nasal brushing biopsies or the harvested NBHE-BMI cells were centrifuged and medium removed. Samples were homogenized by adding 1 mL Trizol (per nasal brushing biopsy or 5 X 10^6 cells), pipetting up and down several times then were incubated for 5 mins at room temperature. 0.2 mL of chloroform per 1 mL of Trizol reagent was added with vigorous shaking and incubation for 2-3 mins, then centrifugation of the samples at 12,000 × g for 15 minutes at 4°C. The aqueous phase then was removed to a new tube and RNA was precipitated using 0.5 mL 100% isopropanol, incubated for 10 mins and centrifuged at 12,000 × g for 10 minutes at 4°C. The RNA pellet was washed in 1 mL 75% ethanol and left to air dry for 5-10 mins, and then re-suspended in RNAase free water. The RNA concentration was quantified using a NanoDrop™ 1000 Spectrophotometer (Thermo Scientific).

2.3.9 Real time quantitative PCR

2.3.9.1 Preparation of cDNA

0.5 – 2 µg of RNA was used for cDNA synthesis using either the Omniscript RT kit (Qiagen Inc., Hilden, Germany) or the High Capacity RNA-to-cDNA Kit (ThermoFisher Scientific) according to the manufacturers' protocol. All reactions were set up on ice to avoid premature cDNA synthesis and to minimize the risk of RNA degradation. cDNA samples were stored at -20°C until needed.

2.3.9.2 Real time qPCR

Primers used for qPCR were designed using the Primer-Blast tool with a chosen PCR product size of 80 -150 bp and the primers were designed to span exon-exon junctions. Real time PCR was done using the CFX96 Touch™ Real-Time PCR Detection System (BIORAD) and the iQ SYBR Green Supermix (BIORAD) according to the standard protocol. In brief, each PCR reaction was done in triplicate where 100 ng cDNA was mixed with 10 µl iQ SYBR Green Supermix and 150nM of each forward and reverse primers, with nuclease free water added to a final volume of 20 µl. The thermal cycling protocol was setup for a two-step PCR reaction starting with 3 min at 95°C for polymerase activation and DNA denaturation followed by 40 cycles of 15 sec denaturation at 95°C and 30 sec annealing/extension at 60°C. Melting curve analysis was done at 55-95°C (in 0.5°C increments) for 30 secs, to confirm the specificity of the primers used. Standard curve were generated separately for each gene of interest and the housekeeping genes using serial cDNA dilutions of pooled samples to assess the efficiency of PCR reactions.

2.3.9.3 Data analysis

Data was analyzed using Bio-Rad CFX Manager 3.1 software (Bio-Rad Labs Inc., CA, USA). Relative quantification of gene expression was performed using the delta-delta Ct ($2^{-\Delta\Delta Ct}$) method, where target gene expression was

normalized to an endogenous housekeeping gene (*GAPDH*) expression and to a calibrator sample (control) as previously described.(270) As each PCR reaction was done in triplicates, standard errors of the mean (SEM) represents variations between technical replicates.

2.3.10 Bacterial cell transformation

Transformation of the competent bacterial cells was performed according to the type of cells (listed in section **2.2.4**) and the recommended transformation protocol from the supplier. Briefly, in general, cells were thawed on ice, about 50 µl of cells were mixed with 1-100 ng plasmid DNA by flicking the bottom of the tube and the mixture then incubated on ice for about 30 mins. Heat shock of each tube was performed by placing the tubes into a 42°C water bath for 30-60 secs and then putting them back on ice for 2 mins. Transformed bacterial mixes were grown in 450 mL SOC outgrowth medium (New England Biolabs) for an hour in a shaking incubator at 37°C. The cells were then plated onto an LB agar plate supplemented with the appropriate antibiotic and incubated at 37°C overnight.

2.3.11 Plasmid purification

Plasmid DNA purification was performed using the QIAprep Spin Miniprep Kit or QIAGEN Plasmid Plus Kit (Qiagen Inc., Hilden, Germany) according to the amount of plasmid DNA yield needed. At the start, a single colony was inoculated overnight in 5ml LB broth with the appropriate antibiotic at 37°C in a shaking incubator. QIAprep Spin Miniprep Kit was used to purify plasmid DNA less than 20 µg, or the inoculated culture was used as an input to seed the growth of a 200 mL LB culture volume overnight for use with the Plasmid Plus Maxi-prep kit for a yield up to 1 mg, according to the manufacturer's protocol. Glycerol stocks were prepared for storage where 500 µl of transformed bacterial cells were mixed with 500 µl of 50% Glycerol and stored at -80°C in 2 ml Cryo-screw capped tubes.

2.3.12 Transfection

One day prior to transfection, HEK293T cells were seeded at a density of 0.5 x 10^6 cells/well in 6-well plates to reach 70-90% confluency at the time of transfection. Lipofectamine 3000 or 2000 (ThermoFisher) was diluted in Opti-MEM serum- free medium and used for application to the cells according to the manufacturer's protocol. When Lipofectamine 2000 was used, the DNA was diluted in Opti-MEM and then the diluted DNA was mixed with the diluted Lipofectamine 2000 reagent and left for 10 mins at room temperature then the DNA-Lipid complex was added to the cells. When Lipofectamine 3000 was used, the required amount of DNA was diluted in Opti-MEM and then P3000 reagent was added and mixed well, before mixing the diluted DNA with the diluted Lipofectamine 3000 reagent and incubating for 10-15 mins before transfection. 2.5 µg of DNA was used for transfection per well in a 6-well plate and the cells were incubated at 37°C and 5% CO2 for 48 hours. The amount of DNA and Lipofectamine used were scaled up/ down according to the surface area of the culture plate/dish/flask.

2.3.13 Western blotting

2.3.13.1 Sample preparation

Protein lysates were prepared from both cell culture and from nasal brushing biopsies. For preparation of lysates from cell culture, the culture dish was placed on ice and the cells were washed with ice-cold 1x PBS then NP-40 lysis buffer (0.5 % NP-40 in TBS) supplemented with cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail (Roche) Tablet/ 10 mL lysis buffer and Phosphatase inhibitor Cocktails 2 and 3 (Sigma) were added. Cell scrapers were used to scrape the cells off the plate and the suspension was incubated at 4°C on an end-over-end shaker for 20mins. The lysate was centrifuged at 10000 g for 10 mins at 4°C. The supernatant then was transferred to a new tube.

Nasal brushing biopsies were transferred in RNALater from the collaborating institutions. For each sample, RNALater was diluted with an equal amount of 1x PBS and centrifuged at 1000 g for 10 mins to pellet the cells and discard the supernatant. NP-40 lysis buffer supplemented with protease inhibitor cocktail was added and the pellet was homogenized by pipetting up and down, then the suspension was incubated at 4°C on an end-over-end shaker for 20-40 mins. The lysate was centrifuged at 1000 g for 10 mins at 4°C. The clear supernatant then was transferred to a new tube.

2.3.13.2 Protein quantification

Quantification of the protein yield was done by colorimetric assay using a Pierce BCA (Bicinchoninic Acid) Protein Assay Kit (ThermoFisher Scientific) in a microplate according to the manufacturer's protocol. In brief, 10-25 µl of each protein sample was added to 200 µl working reagent (50:1 BCA reagent A: B). Diluted Albumin standard was prepared according to the manufacturer's recommendation and mixed with 200 µl working solution. Samples and standards quantification were done in duplicates or triplicates. After incubation for 30 mins at 37°C and cooling down to the room temperature, absorbance was measured at 562 nm on a Thermo Multiskan Spectrum plate reader.

2.3.13.3 SDS-PAGE

Protein samples were mixed with equal volumes of 2x Laemmli Sample Buffer (Bio-Rad) and incubated at 95ºC for 5 mins then loaded into either 12% or 4-20% Mini-Protean TGX Precast Gels (Bio-Rad) submerged in a running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3) along with a colour prestained protein standard, broad range (11-245 kDa) (New England Biolabs).

2.3.13.4 Transfer of proteins

After running the gel for an adequate time, when the dye molecule reached the bottom of the gel, the gel was removed from the cast and equilibrated in

Towbin transfer buffer for 15 mins. Protein blotting was performed using a Trans-Blot® SD Semi-Dry Electrophoretic Transfer Cell (Bio-Rad), onto nitrocellulose membrane (previously wetted in the transfer buffer for 15 mins) for 15–30 minutes at 10–15 V. Thick filter papers were used to enclose a gel/membrane sandwich (paper/gel/membrane/paper) with any air bubbles first removed within the matrix to ensure a proper and complete transfer.

2.3.13.5 Membrane blocking and antibodies

After complete transfer, the membrane was blocked to prevent non-specific binding of antibodies to the membrane. Blocking of the membrane was done using 3%BSA + 2% Milk power in Tris Buffer Saline Tween20 (TBST) buffer and with incubation for 1 hour at room temperature with agitation.

After blocking, the membrane was incubated with diluted primary antibody in an adequate volume of the blocking buffer for 2 hours at room temperature or overnight at 4°C with agitation, to ensure that the membrane was homogenously covered and to prevent uneven binding of the antibodies.

The membrane was washed for 15 mins with TBST washing buffer, 5 mins per wash, to remove any residual primary antibodies. The membrane then incubated with diluted HRP-conjugated secondary antibodies for 1 hour at room temperature with agitation. This was followed by washing 3 times with TBST washing buffer, 5 mins per wash.

2.3.13.6 Development and imaging

For detection of horseradish peroxidase (HRP) activity from secondary antibodies, Pierce ECL Western Blotting Substrate (ThermoFisher Scientific) was used. The membrane was incubated with chemiluminescent substrate (a mix of reagents 1 and 2 at 1:1 ratio) for 1 min. The membrane then was covered with X-ray film and an automated x-ray developer was used to develop the film, or a ChemiDoc™ XRS+ System and Image Lab software (Bio-Rad) were used to develop digital images.

2.3.14 Affinity purification

This part of the project was done during a training placement at the Proteomics lab, Institute for Ophthalmic Research, University of Tübingen, Germany under the supervision of Dr Karsten Boldt.

2.3.14.1 Immunoprecipitation

An adequate volume (12.5 µl per one 14cm cell culture dish) of anti-FLAG M2 Affinity gel (Sigma-Aldrich) was used for immunoprecipitation of FLAG fusion proteins (FLAG tagged IFT74 constructs). The required gel volume was mixed with 600 µl 1x TBS and centrifuged at 5000 g for 30 secs at 4°C. The supernatant was discarded and the gel was mixed with 500 µl of protein lysis buffer and centrifuged again. After discarding the supernatant, the gel was washed twice with 500 µl wash buffer (150 mM sodium chloride, 0.1% NP-40, 50 mM Tris, pH 7.4) and mixed with a 2x gel volume of 1x TBS then transferred to plugged spin columns.

The washed anti-FLAG M2 Affinity resin was mixed with an appropriate amount of lysate and incubated at 4°C on an end-over-end shaker for 1-2 hours. The columns were then transferred to new collection tubes, unplugged and centrifuged at 100 g for 2-3 sec at 4°C. The flow-through was discarded and the resin was washed 3 times with 500 µl wash buffer. The spin columns were plugged again and transferred to a new tube. 200 µl of FLAG-elution buffer (200 µg/ml FLAG peptide in 1x TBS) was added and the columns were closed and incubated at 4°C on an end-over-end shaker for 10 mins. Then, the spin columns were unplugged again and centrifuged 1000g for 30 sec at 4°C.

2.3.15 Mass spectrometry

This part of the project was done during a training placement at the Proteomics lab, Institute for Ophthalmic Research, University of Tübingen Germany under the supervision of Dr Karsten Boldt.

2.3.15.1 Chloroform-methanol precipitation

800 µl methanol was added to 200 µl elute in a 2 ml Eppendorf tube, mixed and spin for 20 sec. Next, 200 µl chloroform was added, mixed and spin for 20 sec and then 600 µl water was added, vortexed for 5 sec and centrifuged for 1 min. The aqueous phase (upper layer) was then carefully removed and discarded. 600 µl methanol was added, mixed gently and centrifuged for 2 mins. The supernatant was removed and the pellet was left to air-dry.

2.3.15.2 Alkylation and tryptic digestion

The protein pellet was dissolved in 30 µl 50 mM ammonium bicarbonate (19.8 mg ammonium bicarbonate (Sigma, A6141-25G) dissolved in 5 ml water). Then, 4 µl RapiGest stock solution (RapiGest SF surfactant (Waters, 186001861) in 50 µl water added to 1 mg, homogenized, then stored at 4°C up to 4 weeks) was added at a final concentration 0.2%, and was vortexed strongly. 0.75 µL of 100 mM DTT (15.4 mg 1,4-dithiothreitol (Merck, 1.11474.0025) dissolved in 1 ml water) was added and vortexed then incubated at 60°C for 10 mins. The samples were cooled down to the room temperature and 1 µL of 300 mM iodoacetamide (55.5 mg 2-iodacetamide (Merck, 8.04744.0025) dissolved in 1 ml water) was added, vortexed and incubated at room temperature for 30 mins in complete darkness. 2 µL trypsin stock solution (1 µg/µl) was added and vortexed, then incubated over night at 37°C.

On the second day, the samples were centrifuged for 1 min at 9000 g at room temperature. The enzymatic digestion was stopped by applying TFA (trifluoracetic acid (Fluka, 40967) diluted in acetonitrile 80% (Sigma, 34967)

to a final concentration of 5%). The acidified solution was then added into a spin column immediately and incubated for 10 mins at room temperature, then centrifuged 16000 g at 22°C for 15 mins. The clear solution (between the bottom pellet and the upper oily phase) was transferred to a new tube. The peptide samples were then desalted and purified using StageTips and then separated by on a NanoRSLC3000 HPLC system (Dionex). The peptides were ionized by nano-spray ionization and detected using a LTQ Orbitrap Velos mass spectrometer (ThermoFisher Scientific). For identification of proteins, the raw spectra was searched against the human SwissProt databases using Mascot (Matrix Science, version 2.4.0). The results were validated by Scaffold (version 4.0.3, Proteome Software) employing the protein prophet algorithm.(271, 272)

2.3.15.3 Data analysis

MaxQuant software (a freely available quantitative proteomics software package, version 1.5.0.3) was used for identification and label free quantification, with a minimum ratio count of 2. The first search peptide tolerance was set to 20, the main search peptide tolerance to 4.5 ppm and the 're-quantify' option was selected. For peptide and protein identification the Bos Taurus subset of the Uniprot database (Release 2013_09) was used and contaminants were detected using the MaxQuant contaminant search. A minimum peptide number of 2 and a minimum length of 7 amino acids was tolerated. A false discovery rate (FDR) of protein and peptides was set at 1% with unique and razor peptides used for quantification. For statistical analysis including ratio, two-sample t-test (with permutation based FDR determination) and significance A (with Benjamini-Hochberg based FDR determination) calculation and visualization of the data, the freely available Perseus software package (version 1.6.1.3) was used.

2.3.16 *Paramecium* studies

This part of the project was done during a training placement in the *Paramecium* Lab at the Institute for Integrative Biology of the Cell (I2BC) – CNRS, France, under the supervision of Dr Anne-Marie Tassin. *Paramecium tetraurelia* is a unicellular organism that belongs to ciliate phylum. It is multiciliated organism (about 4000 cilia)and thus can allow molecular and biochemical analyses of potential candidate proteins in PCD.(101)

2.3.16.1 Phylogenetic analysis

To identify the orthologs of PCD candidate genes of interest in *Paramecium*, Dr France Koll (from the *Paramecium* Lab, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris Sud, Université Paris-Saclay) used the public databases http://cildb.cgm.cnrs-gif.fr/ and http://paramecium.cgm.cnrs-gif.fr/. As *Paramecium* has passed through 3 whole-genome duplications, we expected to find several orthologs to each human candidate gene.(273)

For the human *C11orf70* gene, the *Paramecium* ortholog was identified as *GSPATG00011350001.* For the human *DNAH9,* two sequences were identified as its orthologs in *Paramecium* genome corresponding to (*GSPATG00038423001, GSPATG00038424001, GSPATG00038425001*) and (*GSPATG00011643001, GSPATG00011644001, GSPATG00011645001*).

2.3.16.2 RNAi design

Each RNAi insert was cloned in the L4440 plasmid between its two convergent T7 promotors. HT115 bacteria (*E. coli* strain lacking RNAase III activity) were then transformed with the plasmid. The transformed bacteria containing the RNAi were prepared in the *Paramecium* feeding medium for intake by *Paramecium tetraurelia*.(102) For *Paramecium C11orf70* ortholog knockdown, 561 bp fragment corresponding to the 5' part of the gene was cloned into the L4440 plasmid. The sequence of *C11orf70*-RNAi insert was:

5’-
ATGTAAATTGAATCAGACAATTAAGTTACAAATCAGAGTTACTCTTTTTTT
AGGTAAGAGAATACGATTTTGGATGATAAAAAGTTTATGGAAATATTGTA
GAAATGGGGTCTATAACATAGCATCAAAGTGAGCACATTTTTATTTGATA
TCAAATTTGATCATCTAAATCCAAATCAATTTTTGTTGGATTTATTCAATT
CTAAGGACGTAAGAGGTTCATTGCATTATGTTTCATTTAAATAGAATGTG
TTGCTATCATAGATTAAATTTTAACCTTTGACATGCAAATCCATAAAATTG
GACTTATTCGATAAGTTAACAGAAGATAAAATTGTTGTCAAGGGTCACAT
CAAGTAGTGTTTCGAGGAATAATTTGAAAATATTCAGATAGCTGATGAAT
TACGTAAAGCACTTGTTTTGGAAGATTCGGAACAATATTGTGTATTCAAT
GAAGCTGACAGATAAGAATTATTATTCAAATTATTCCAAATTCTAGTTCT
GGGGGGATAGCTTTGTTAATATGAAGATGAAATACAAGCCTATTTAGAT
TGGACAAAG-3’.

For the *Paramecium DNAH9* ortholog knockdown, a 737 bp fragment was used which was 100% identical to part of the sequence of both the identified *DNAH9* orthologs. The sequence of the *DNAH9*-RNAi insert used was:

5’-
CGTTAGTACTAGAGATTTGTATGGTTACAATTTACCAACAAAGGAATGG
AAAGATGGTTTAGTTTCAAAGGTTTTGAGATCTCTAAGTGAAATTCAAGA
TGTTAATCCTAAATGGATATTACTTGATGGAGATTTAGATGCTAATTGGA
TTGAGTCTATGAATTCAGTGATGGATGACAATAAGATATTGACATTGGC
AAATAATGAAAGAATTCCATTGAAACCACACATGAGAATGCTTTTTGAAA
TCAGAGATCTTCGTTTTGCTACTCCTGCCACAGTGTCTAGAGCTGGTAT
CTTATATATTTCAGATGACAAGGGGTATTAATGGAGAGCCTACGTGAAA
TCATGGGTTAAGAATAACTTTAATGATGATAAATTTAAATAAGATTTACAA
AAACTATTTGATAGATATATTGAAGGTACTTTACTATTCTTAAAGAAACAT
TGTAAGACTTTGATTCCAGTCAATCCAATTTCTATGATTATCTCATTATG
CAAAGCCTTATTGCCATTATTATAAGGAGAAGTGAAGAATATGGAATATC
ATTTCGTGTATTGCTGTGTTTGGGCTATTGGAGGTGTTTTATCTGAAAAA
GACTCCATTGATTATCGTAAAGATTTCTCTAATTGGTGGAAAGGTGAAT
GGAAGACTTCAGTAAAATTCCCAAGTAAGGGTACAGTATTTGATTACTTT
GTTGAATAAAATTCAGAGAATGTTAAATTCGATGAATGGG-3’.

2.3.16.3 RNAi by feeding

The RNAi by feeding experiment was started by transfer of 4 ml of *Paramecium* culture stock to 50 ml bacterized BHB (Bio Herbe de Blé) feeding medium supplemented with β-Sitosterol and incubated at 28°C for 2 days. One day prior to starting the feeding, a colony of HT115 E. coli strain carrying an L4440 plasmid with the relevant insert targeting the *Paramecium* gene of interest, was picked from an LB plate containing ampicillin (100 µg/ml) and tetracycline (12.5 µg/ml) and inoculated into 10 ml of LB medium

supplemented with the same antibiotics (LBAT) placed overnight at 37°C in the shaking incubator.

On the day of feeding, the bacterial culture was diluted (1/100) in 10 ml fresh LBAT medium and incubated at 37°C in a shaking incubator for 2-3 hours up to an OD (600 nm) of 0.4 to 0.6. 25 µl of IPTG was added to reach 0.4 mM final concentration and incubated at 37°C in a shaking incubator for 3 hours, then the bacterial culture was centrifuged for 15 mins at 4000 rpm (4°C), the supernatant was discarded and the pellet was re-suspended in 5-10 ml BHB medium with ampicillin (without tetracycline, as *Paramecium* is intolerant to it), the concentration then being adjusted to get an OD (600 nm) of 0.25-0.3, at which point IPTG (2.5 µl/ml) and β-Sitosterol (2 µl/ml) were added.

The silencing feeding medium remains biologically active for about 24 hours and that is why feeding medium was prepared the same way everyday over the feeding period (1 – 3 days), to ensure continuous gene silencing. A single *Paramecium* cell was introduced into 200-300 µl of freshly prepared silencing feeding medium in a depression slide, and incubated overnight at 27°C. For the gene silencing to be done over few days, a single cell was isolated the next day in the same amount of a fresh silencing medium.

For mass culture of *Paramecium*, the cultures were inoculated at a density of 40-100 cells/ml and incubated at 27°C without agitation with silencing medium replaced everyday freshly prepared, over the period of feeding. The cells were then harvested by removing the grass debris by filtering the culture through a tight plug of sterilized gauze in a funnel followed by centrifugation at 170-200g for 2 mins with the *Paramecium* pellet then washed with Dryl's solution 2-3 times.(274)

All RNAi knockdown experiments were done ≥ 3 times to ensure reproducibility of the results. Silencing of the *ND7* gene was used as a negative control for the silencing experiments as it affects trichocyst exocytosis, without affecting cilia motility.(275)

2.3.16.4 Paramecium cilia function testing

To assess the level of gene silencing at the RNA level, from a mass culture of *Paramecium* prepared as described above, total RNA was extracted using the RNAeasy Micro kit (Qiagen). 1 µg RNA was used to prepare cDNA using Superscript IIII (Invitrogen) and random hexanucleotides (Invitrogen). qPCR was performed using specific primers as listed in (**Table 2-2**). Phenotype assessment was done after RNAi silencing.

Table 2-2 Primers used for qPCR to assess the level of knockdown in Paramecium

Gene	Forward primer	Reverse primer
GAPDH	GAGAGCCGGAAGAGCTGCTA	TGGTGGAACTCTGAAGGCCATA
C11orf70	TGCTCGAAAATATGCAGATAAGGA	TGATGGGTTCACAACGACAT
DNAH9	TCCAGCACCTAGATCAGGTC	CTGGTTTCTTCTCGCTGTCT
	TTCCAGCACCTAGGTCGGGCCA	CTGGCTTTTTCTCACTATCC

Paramecium swimming velocity

To assess *Paramecium* swimming pattern and velocity, 3 *Paramecium* cells were transferred into a drop of conditioned BHB medium and tracked under dark-field microscope using MetaVue software for 10 sec with an image taken every 0.3 sec. For image analysis and measurement of the swimming velocity, freely available ImageJ software was used.

Cilia beating frequency

To assess *Paramecium* cilia beating frequency after RNAi knockdown experiments, high speed video microscopic analysis of silenced *Paramecium* cilia beating was compared to that of the ND7 knockdown control using a 63X oil immersion lens on an upright DM60 Leica microscope at 500 frames per seconds at 37°C. This analysis was done with Dr Amelia Shoemark at Royal Brompton Hospital PCD diagnostic Unit or with Ms Dani Lee at Prof. Christopher O'Callaghan's lab at UCL Great Ormond Street Institute of Child Health. The CiliaFA plugin in ImageJ was used to quantify cilia beat frequency in *Paramecium*.(104)

Cilia ultrastructure

Sample fixation and processing was done by Dr Michel Lemullois (at the Paris *Paramecium* lab). Analysis of cilia cross sections was done, blinded to knockdown status, with Dr Amelia Shoemark at the Royal Brompton Hospital using a digital camera (AMT 16X, Deben UK). The numbers of missing IDA and ODA arms was recorded per each cilia cross section both in the silenced and control strains.

Cilia gross morphology

Immunofluorescence studies of *Paramecium* were done with Dr Anne Aubusson-Fleury (from the *Paramecium* Lab). Paramecium cells were fixed in 2% paraformaldehyde in PHEM buffer (Pipes 60mM, Hepes 25 mM, EGTA 10mM, MgCl2 2mM, adjusted to pH 6.9 with NaOH) for 15 minutes. The cells were then permeabilised with 1% Triton X-100 in PHEM for 15 mins and washed with PBS/BSA 3% three times. 1:500 polyglutamylated tubulin antibodies were used for immunostaining.

2.3.16.5 Paramecium microinjections

Paramecium injection was done by Dr Anne-Marie Tassin (at the *Paramecium* Lab). First the *Paramecium* orthologs of *C11orf70* and *IFT46* were cloned into pPXV-GFP and pZZ-GFP02 plasmids respectively. Linearized plasmids (5 µg/µl) were microinjected into the macronucleus of *ND7-1* mutant cells unable to discharge their trichocysts along with plasmid DNA directing the expression of wild type *ND7* under an inverted Nikon phase-contrast microscope, using a Narishige micromanipulation device and an Eppendorf air pressure microinjector. The *Paramecium* cells were left to divide 3-4 times and then assessed for the expression of the fluorescent proteins and the ability to discharge their trichocysts.

2.3.16.6 Paramecium deciliation

The *Paramecium* cells were incubated with vortexing in 10mM Tris pH 7.4, 1mM CaCl2, 5% ethanol for 30 sec.

2.3.17 Online databases used in this study

CILDB http://cildb.cgm.cnrs-gif.fr/

***Paramecium*DB** http://paramecium.cgm.cnrs-gif.fr/

RNAi off-target tool

http://paramecium.cgm.cnrs-gif.fr/cgi/tool/alignment/off_target.cgi

dbSNP build 141 https://www.ncbi.nlm.nih.gov/projects/SNP/

ExAC http://exac.broadinstitute.org/

Exome Variant Server http://evs.gs.washington.edu/EVS/

1000Genomes http://1000genomes.org/

NCBI Primer-BLAST tool https://www.ncbi.nlm.nih.gov/tools/primer-blast/

Human Splicing Finder http://www.umd.be/HSF3/

SIFT http://sift.jcvi.org/

Polyphen-2 http://genetics.bwh.harvard.edu/pph2/

Mutation Taster http://www.mutationtaster.org/

CADD http://cadd.gs.washington.edu/

MAPP
http://mendel.stanford.edu/sidowlab/downloads/MAPP/index.html

PhastCons http://compgen.cshl.edu/phast/

PhyloP http://compgen.bscb.cornell.edu/phast/help-pages/phyloP.txt

BLAST/BLOSUM62 https://blast.ncbi.nlm.nih.gov/Blast.cgi

DGV http://dgv.tcag.ca/dgv/app/home

OMIM http://www.omim.org/

RefSeq http://www.ncbi.nlm.nih.gov/RefSeq

Ensembl Genome Browser http://www.ensembl.org/index.html

NCBI	https://www.ncbi.nlm.nih.gov/
GTEx	https://gtexportal.org/home/
Uniprot	https://www.uniprot.org/
IntAct	https://www.ebi.ac.uk/intact/

Chapter 3 Deciphering the morbid genome of PCD using targeted NGS

3.1 Introduction

PCD is generally inherited as an autosomal recessive disease. However, very rare autosomal dominant and X-linked inheritance patterns have been reported for ciliary dyskinesia-like phenotypes.(124, 128-130) PCD-causing mutations have been identified in either structural ciliary genes, genes implicated in multiciliogenesis or genes playing a role in other aspects of ciliary motility, for example dynein assembly and transport factors.(56, 124) To reach a specific genetic diagnosis of PCD, bi-allelic mutations in an autosomal recessive state or a hemizygous mutation in an X-linked gene should be identified.(58) Most of the mutations known to cause PCD are of high impact in terms of predicted pathogenicity being nonsense, frameshift or splice-site mutations whilst missense mutations are less reported.(58, 80)

The diagnostic workup of PCD is complex and it may require multiple testing modalities to confirm diagnosis.(44, 57, 78) This requires obtaining specific samples from the patients to study cilia motility and ultrastructure through nasal brushing biopsies, or more invasive procedures like bronchial biopsies. Recurrent sampling may be required in cases of insufficient samples or inconclusive results.(58) This can be inconvenient and traumatic to the patients. These tests are expensive, require highly trained personnel and sophisticated equipment that are only available in certain specialized centres.(76) Genetic testing can overcome these hurdles, as DNA samples can be easily obtained from blood or saliva and then transported to any laboratory.(124) With identification of the whole genetic landscape of PCD possible in the near future, genetics can increasingly be considered as a first-line diagnostic test.

Various genetic screening approaches have been used to investigate the genetic background of PCD. These include the early single candidate gene approach by direct screening of candidates of interest that were highlighted from information on their evolutionary conservation and their role in cilia motility, arising from protein alignments and functional or mutational studies in model organisms. This candidate gene approach led to the discovery of

DNAI1 which is the first human gene linked to PCD, as published in 1999.(87) Homozygosity mapping and linkage studies used for gene discovery in PCD have also been successful.(86, 141) However, due to high genetic heterogeneity, the use of next generation sequencing technology from around 2010 has revolutionized the genetic diagnostics and research in PCD. This approach has enabled fast and comprehensive genetic analysis and subsequently improved and expedited the diagnostic process, since it provides parallel sequencing of targeted gene panels or at a more global scale, of the whole exome or genome.(56, 78-80, 114)

Targeted sequencing is considered to provide inexpensive genetic testing with a high read depth of the targeted regions and less of data burden.(276) It facilitates data analysis and enables the interpretation of the identified sequence variants with minimal chances of incidental findings.(277) On the other hand, there are limitations of using this approach for genetic diagnosis including the challenge to identify certain variants such as copy number variants (CNVs), especially in regions with lower coverage. A certain gene of interest implicated in the disease may not be included in the panel and to achieve high efficiency, gene panels require extensive literature search, curation and continuous updates of the gene list in the panel.(276)

In this chapter, I will present the diagnostic yield of using custom panel sequencing in a large PCD cohort from various ethnic backgrounds. This has revealed a striking impact of ethnicity on the genetic stratification of PCD. It found that not all PCD mutations are private or family-specific, and there are some mutations that occur more frequently in certain ethnicities while other more universal mutations were detected in different ethnicities. Genetics has not only helped in the diagnosis of PCD, but helped in better characterization of the PCD phenotypic paradigm as well.

3.2 Results

3.2.1 Characteristics and ciliary ultrastructural phenotypes of the PCD patient cohort enrolled for genetic testing

A total of 180 samples from 175 unrelated families were recruited with ethnicities including European, South-Asian, Arab and a smaller group of families from other nationalities. The South Asian families were originally from Pakistan, India, Bangladesh, Nepal and Sri Lanka. The Arab families included families from Egypt, Palestine, Kuwait and Iraq. The other families were from Turkey, Chile, Afro-Caribbean, Somalia, West Africa, South Africa, Hong Kong and mixed ethnicities. Consanguinity data was self-reported. Parental consanguinity was found in 27% of all the families. 25 out of 29 Arab families (86%) and 12 out of 35 South-Asian families (34%) were consanguineous. In 14 families, PCD diagnosis was found to be extremely unlikely based on reviewing their standard diagnostic testing results. In these individuals, no variants of interest were prioritized in the screened genes. These families were excluded from further analysis. The ethnic breakdown and the consanguinity state of the 161 remaining families consistent for PCD affection status is shown in **Table 3-1.**

Table 3-1 Consanguinity among different ethnicities

Ethnicity	No. of families	Consanguinity		
		Yes	No	Unknown
European	74	1	66	7
South-Asian	35	12	14	9
Arab	29	25	3	1
Other	15	7	4	4
Unknown	8	2	2	4

Amongst these 161 families, TEM ultrastructure analysis of the respiratory cilia had been performed during the clinical diagnostic workup in 134 families (83%). Of these 134, as summarized in **Table 3-2**, PCD was confirmed in 97 families (60%) by identifying a ciliary ultrastructural defect and was deemed highly suggestive in 37 families (23%) who had a typical PCD clinical history but with inconclusive diagnostic results i.e. no ultrastructural defect was identified or the defects were not prevalent enough among the examined cilia cross sections to confirm the diagnosis. In the final 27 families (17%), although their clinical history was highly suggestive of PCD, diagnostic TEM analysis had not been performed or results were not available.

The most commonly identified ultrastructural defect was ODA loss, found in 45% of families either alone in 23% (31 families) or combined with IDA loss in 22% (30 families). Other reported ultrastructural defects comprised microtubular disorganization with or without IDA loss (11%), central microtubular complex defects (6%), predominant IDA loss without loss of ODA (4%) and a lack of cilia cross sections (5%).

Table 3-2 Transmission Electron Microscopy findings and mutations in known PCD genes

TEM (Ultrastructural phenotype)	No. of families (%)	Mutated known PCD gene	(n)
Inconclusive TEM analysis (No apparent defect, or few observed defects not enough to make a diagnosis)	37 (28%)	***DNAH11***	**15**
		HYDIN	**6**
		DNAH5	**3**
		CCDC103	**2**
		RSPH1	**1**
		OFD1	**1**
		DNAI2	**1**
		ZMYND10	**1**
Outer dynein arm loss	31 (23%)	***DNAH5***	**17**
		DNAI1	**3**
		ARMC4	**2**
		DNAI2	**1**
		CCDC151	**1**
		SPAG1	**1**
		PIH1D3	**1**
Combined Inner and Outer dynein arm loss	30 (22%)	***LRRC6***	**5**
		DNAAF3	**5**
		CCDC103	**3**
		DNAH5	**3**
		ZMYND10	**2**
		DYX1C1	**1**
		DNAAF1	**1**
		HEATR2	**1**
		DNAI1	**1**
Microtubular disorganization ± Inner dynein arm loss	15 (11%)	***CCDC40***	**6**
		CCDC39	**5**
		CCDC65	**1**
		RSPH9	**1**
		RSPH1	**1**
Central pair defect	8 (6%)	***RSPH4A***	**4**
		RSPH1	**2**
		RSPH9	**1**
Inner dynein arm loss	6 (4%)	***CCDC103***	**2**
		CCDC164	**1**
		CCDC40	**1**
Lack of cilia cross sections	7 (5%)	***CCNO***	**2**
		MCIDAS	**2**
		DYX1C1	**1**
		RPGR	**1**

3.2.2 Targeted NGS yields high diagnostic output in PCD patients

Targeted multi-gene panels were designed to include all known PCD genes at the time of panel design and other potential candidates. Variants in known PCD genes were identified in 128 of the 161 PCD families (79%) (**Figure 3-1**). In 116 of these families (72%), variants identified were either bi-allelic in autosomal genes or hemizygous in an X-linked known PCD gene (i.e. complete genetic diagnosis was achieved). The full list of identified variants is shown in **Table 3-3.** The RefSeq accession numbers used for the nomenclature of the mutations are listed in **Appendix Table A-0-3**. Detailed data about allele frequency of each identified variant in control databases, segregations results, TEM defect, consanguinity and ethnicities of the families are shown in **Appendix Tables A-0-4 & A-0-5.**

In 12 families (7%) a single variant allele (single heterozygous) was found in a known autosomal recessive gene but with no second allele identified. Seven of these single heterozygous mutations were protein-truncating variants and 5 of them had been reported before in PCD patients in previous studies. Having one mutant allele despite identified in a known PCD gene was considered an incomplete genetic diagnosis; however, all 12 families were included in further analysis, as their genetic mutations occurred in genes which when mutated would cause ultrastructural defects that were consistent with the patients' actual reported TEM results, supporting them as likely causal alleles (**Table 3-4 and Appendix Table A-0-5**). All these variants were confirmed by Sanger sequencing and segregated within the families whenever DNA samples from family members were available.

In 13 families (8%), all the known PCD genes were excluded and bi-allelic variants in candidate genes for PCD were identified. All these variants were confirmed by Sanger sequencing and segregated within the families, whenever DNA samples from family members were available. Functional characterization of these genes and confirming the role of their loss of function in causing PCD will be demonstrated later in the following chapters. **Table 3-5** shows a list of variants prioritized in 10 potential PCD candidate genes, 4 of

which are further analysed in **Chapters 4-6**. No functional studies are presented for the other 6 putative novel PCD genes and their variants, due to time constraints and a lack of patient samples.

In the remaining 20 families (12%), no putative significant sequence variants were identified. PCD diagnosis was confirmed by ultrastructural defects in more than half of these patients lacking genetic results (**Figure 3-2**).

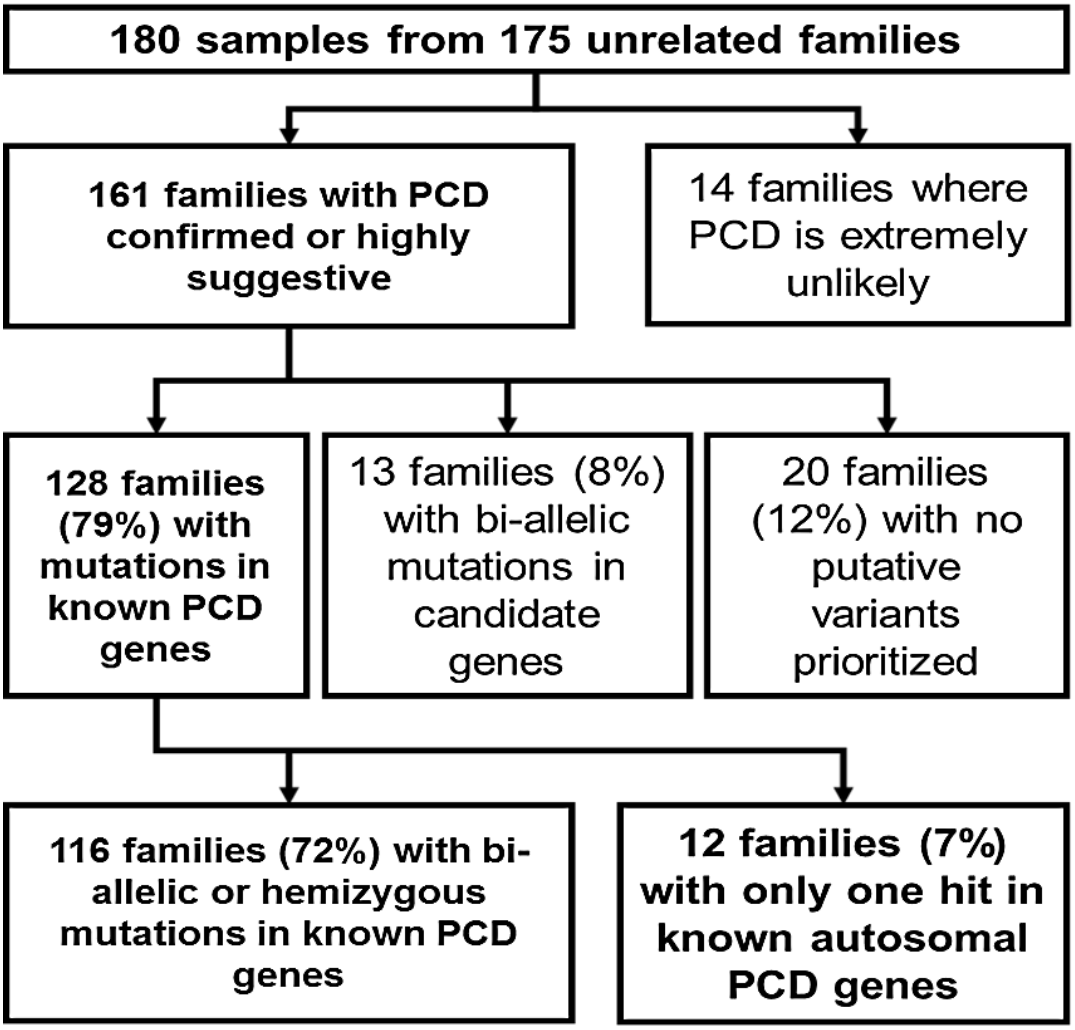


Figure 3-1 Targeted next generation sequencing yields a high diagnostic output in PCD patients

Flow chart of the families enrolled in the study and their genetic diagnostic output: 180 samples (from 175 unrelated families) were submitted to next generation sequencing using 2 versions of motile Ciliome panels. PCD diagnosis was confirmed or highly suggestive in 161 families, complete genetic diagnosis with mutations identified in known PCD genes was confirmed in 72% of families and in 7% of families, only one mutant allele was identified in known PCD gene.

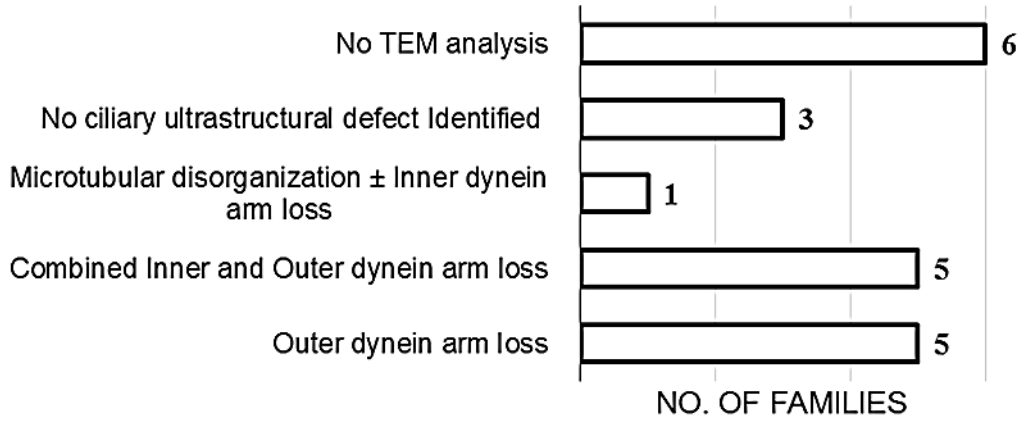


Figure 3-2 Transmission Electron Microscopy findings in families with no genetic results

In 20 families where no putative sequence variants were prioritized, there was no TEM analysis done in 6 families. For the other 14 families with TEM results; an ultrastructural defect was identified in 11 families (5 families with ODA loss, 5 families with combined IDA+ODA loss and 1 family with microtubular disorganization and IDA loss) while in 3 families, inconclusive TEM results were reported.

Table 3-3 List of bi-allelic and hemizygous variants in known PCD genes

Study ID	Gene	Genomic Coordinates (GRCh37)	c. Nomenclature	p. Nomenclature	Impact	Reported before
G001	***DNAAF3***	Chr19:55676764	c.296del	p.Glu99Glyfs*88	Frameshift	-
		Chr19:55673064 - 55673065	c.609_610delinsTGGGA	p.Ala204delinsGlyThr	Inframe delins	-
G002	***DNAH5***	Chr5:13701398	c.13486C>T	p.Arg4496*	Nonsense	(143)
		Chr5:13701426	c.13458_13459insT	p.Asn4487fs*1	Frameshift	(125, 143)
G003	***RSPH9***	Chr6:43638659 - 43638661	c.801_803delGAA	p.Lys268del	Inframe del	(176)
G004	***RSPH4A***	Chr6:116938246	c.460C>T	p.Gln154*	Nonsense	(176)
G005	***CCDC40***	Chr17:78073323	c.3181-3C>G	p.?	Splice site	-
G006	***LRRC6***	Chr8:133645009	c.630del	p.Trp210Cysfs*12	Frameshift	(252)
G007	***CCDC40***	Chr17:78032753	c.1414del	p.Arg472Glyfs*3	Frameshift	-
		Chr17:78071119	c.3097A>T	p.Lys1033*	Nonsense	(114)
G008	***CCDC40***	Chr17:78063562	c.2712-1G>T	p.?	Splice acceptor	(216)
		Chr17:78013765	c.248del	p.Ala83Valfs*84	Frameshift	(214)
G009	***DNAAF3***	Chr19:55677290 - 55677292	c.162_164delinsG	p.Val55Glyfs*28	Frameshift	-
G010	***DNAI1***	Chr9: 34459053	c.48+2dup	p.Ser17Valfs*12	Splice donor	(138)
G011	***ZMYND10***	Chr3:50381271	c.212T>C	p.Leu71Pro	Missense	-
G012	***CCDC114***	Chr19:48805987	c.1093C>T	p.Gln365*	Nonsense	-
G013	***DNAH5***	Chr5:13708395	c.13175T>G	p.Phe4392Cys	Missense	-
		Chr5:13870977	c.3733C>T	p.Arg1245Cys	Missense	-
G014	***DNAI1***	Chr9:34514434	c.1612G>A	p.Ala538Thr	Missense	(138)
G015	***RSPH1***	Chr21:43906573	c.275-2A>C	p.Gly92Alafs*10	Splice acceptor	(184)
G016	***DNAI1***	Chr9: 34459053	c.48+2dup	p.Ser17Valfs*12	Splice donor	(138)
G017	***HYDIN***	Chr16:70894739	c.11843C>G	p.Pro3948Arg	Missense	-
		Chr16:70913319	c.10438G>A	p.Val3480Met	Missense	-
G018	***CCDC40***	Chr17:78055687-78055691	c.1819_1823delinsT	p.Leu607Trpfs*33	Frameshift	-
G019	***CCDC103***	Chr17:42979917	c.461A>C	p.His154Pro	Missense	(254, 255)

G020	***DNAH5***	Chr5:13753599	c.10615C>T	p.Arg3539Cys	Missense	(144)
		Chr5:13753399	c.10815del	p.Pro3606Hisfs*22	Frameshift	(278)
G021	***HYDIN***	Chr16:70989412	c.6182A>G	p.Asn2061Ser	Missense	-
		Chr16:71101211	c.2057C>T	p.Ala686Val	Missense	-
G022	***CCNO***	Chr5:54528218	c.538dup	p.Val180Glyfs*55	Frameshift	-
G023	***DNAH5***	Chr5:13817711-13817715	c.6930_6934delinsG	p.Asn2310Lysfs*15	Frameshift	-
		Chr5:13701426	c.13458_13459insT	p.Asn4487fs*1	Frameshift	(143)
G024	***RPGR***	ChrX:38170000	c.646G>T	p.Glu216*	Nonsense	-
G025	***DNAH11***	Chr7:21723534	c.5593C>T	p.Arg1865*	Nonsense	-
G026	***DNAH5***	Chr5:13829802	c.6261T>G	p.Tyr2087*	Nonsense	-
G027	***DYX1C1***	Deletion of exon 7			CNV	(233)
G028	***DNAH5***	Chr5:13839639	c.5710-2A>G	p.Cys1904-Lys1909del	Splice acceptor	(144)
G029	***DNAH5***	Chr5:13737356	c.11455+5G>A	p.?	Splice Region	-
		Chr5:13829802	c.6261T>G	p.Tyr2087*	Nonsense	-
G030	***RSPH9***	Chr6:43638659 - 43638661	c.801_803delGAA	p.Lys268del	Inframe del	(176)
G031	***CCDC65***	Chr12:49312106	c.658G>T	p.Glu220*	Nonsense	-
G032	***DNAH11***	Chr7:21675540	c.4552C>T	p.Gln1518*	Nonsense	-
		Chr7:21726874	c.5778+1G>A	p.Val1821Thrfs*7	Splice donor	(161)
G033	***LRRC6***	Chr8:133645009	c.630del	p.Trp210Cysfs*12	Frameshift	(252)
G034	***PIH1D3***	ChrX:106462133	c.266G>A	p.Trp89*	Nonsense	(128)
G035	***OFD1***	ChrX:13785391-13785393	c.2745_2747delinsC	p.Tyr916Serfs*7	Frameshift	-
G036	***DNAH11***	Chr7:21789974	c.8932C>T	p.Gln2978*	Nonsense	-
		Chr7:21599378-21599383	c.853_857delinsG	p.Arg285Glufs*22	Frameshift	-
G037	***CCDC103***	Chr17:42979917	c.461A>C	p.His154Pro	Missense	(254, 255)
G038	***RSPH4A***	Chr6:116953417	c.1962_1966delinsC	p.Asp655Ilefs*83	Frameshift	-
G039	***DNAAF3***	Chr19:55673053	c.621dup	p.Val208Cysfs*12	Frameshift	(232)
G040	***DNAH5***	Chr5:13753399	c.10815del	p.Pro3606Hisfs*22	Frameshift	(278)
		Chr5:13701425-13701426	c.13458_13459insT	p.Asn4487fs*1	Frameshift	(143)
G041	***DNAH11***	Chr7:21934608	c.13040T>C	p.Leu4347Pro	Missense	-
		Deletion of exons 68-75			CNV	-
G042	***DYX1C1***	Deletion of exon 7			CNV	(233)

G043	***DNAH5***	Chr5:13753613	c.10601T>C	p.Phe3534Ser	Missense	-
		Chr5:13701425-13701426	c.13458_13459insT	p.Asn4487fs*1	Frameshift	(143)
G044	***CCDC40***	Chr17:78032754-78032755	c.1415del	p.Arg472fs3*	Frameshift	(216)
G045	***DNAH5***	Chr5:13886106	c.2710G>T	p.Glu904*	Nonsense	-
G046	***LRRC6***	Chr8:133645009	c.630del	p.Trp210Cysfs*12	Frameshift	(252)
G047	***DNAAF3***	Chr19:55677221	c.228+5G>C	p.?	Splice site	-
G048	***DNAAF3***	Chr19:55672025-55672026	c.1030_1031delinsG	p.Pro344Glyfs*64	Frameshift	-
		Chr19:55670783	c.1273G>T	p.Gly425*	Nonsense	-
G049	***CCDC40***	Chr17:78013765	c.248del	p.Ala83Valfs*84	Frameshift	(214)
G050	***DNAH5***	Chr5:13928248	c.232C>T	p.Arg78*	Nonsense	(143)
		Chr5:13753399	c.10815del	p.Pro3606Hisfs*22	Frameshift	(278)
G051	***DNAH5***	Chr5:13830872-13830873	c.5890_5894dup	p.Leu1966Serfs*9	Frameshift	-
		Chr5:13820505	c.6791G>A	p.Ser2264Asn	Missense	(143)
G052	***DNAH5***	Chr5:13708285	c.13285C>T	p.Arg4429*	Nonsense	-
		Chr5:13788830	c.8642C>G	p.Ala2881Gly	Missense	(278)
G053	***SPAG1***	Chr8:101226129	c.1519dup	p.Ile507Asnfs*5	Frameshift	(279)
G054	***DNAH5***	Chr5:13820533	c.6763C>T	p.Arg2255*	Nonsense	-
		Chr5:13770983	c.9480T>A	p.Cys3160*	Nonsense	-
G055	***RSPH4A***	Chr6:116949221	c.1351C>T	p.Gln451*	Nonsense	-
		Chr6:116937902	c.116C>A	p.Ser39*	Nonsense	(179)
G056	***CCDC39***	Chr3:180372650-180372652	c.830_831del	p.Asn276Lysfs*4	Frameshift	(216)
G057	***CCDC103***	Chr17:42979917	c.461A>C	p.His154Pro	Missense	(254, 255)
G058	***MCIDAS***	Chr5:54518828-54518829	c.332_333delinsG	p.Ala111Glyfs*22	Frameshift	-
G059	***LRRC6***	Chr8:133645009	c.630del	p.Trp210Cysfs*12	Frameshift	(252)
G060	***ARMC4***	Chr10:28257856	c.1233_1234delinsT	p.Leu411Phefs*48	Frameshift	-
		Chr10:28229509	c.1969C>T	p.Gln657*	Nonsense	(167)
G061	***DNAH5***	Chr5:13845040	c.5177T>C	p.Leu1726Pro	Missense	(280)
		Chr5:13902162	c.1730G>C	p.Arg577Thr	Missense	(143)
G062	***RSPH1***	Chr21:43906573	c.275-2A>C	p.Gly92Alafs*10	Splice acceptor	(184)
G063	***DNAH5***	Chr5:13792147	c.8404C>T	p.Gln2802*	Nonsense	(143)
		Chr5:13830135	c.6249G>A	p.Met2083Ile	Missense	-

G064	***RSPH4A***	Chr6:116949326	c.1456G>C	p.Ala486Pro	Missense	-
G065	***DNAH5***	Chr5:13753399	c.10815del	p.Pro3606Hisfs*22	Frameshift	(278)
G066	***DNAH5***	Chr5:13841167	c.5557A>T	p.Lys1853*	Nonsense	-
G067	***ARMC4***	Chr10:28250600	c.1283C>G	p.Ser428*	Nonsense	-
G068	***CCDC103***	Chr17:42979917	c.461A>C	p.His154Pro	Missense	(254, 255)
G069	***CCDC40***	Chr17:78022417	c.712G>T	p.Glu238*	Nonsense	-
		Chr17:78023861	c.940-2A>G	p.?	splice acceptor	(216)
G070	***CCDC40***	Chr17:78013764-78013765	c.248del	p.Ala83Valfs*84	Frameshift	(214)
		Chr17:78061404	c.2450-2A>G	p.?	splice acceptor	-
G071	***LRRC6***	Chr8:133645009	c.630delG	p.Trp210Cysfs*12	Frameshift	(252)
G072	***DNAH11***	Chr7:21727067	c.5846G>A	p.Arg1949Gln	Missense	-
		Chr7:21940699	c.13380_13383dup	p.Ala4462Leufs*22	Frameshift	-
G073	***DNAH11***	Chr7:21824139-21824142	c.9581_9582del	p.Leu3194Glnfs*10	Frameshift	-
		Chr7:21658796	c.4333C>T	p.Arg1445*	Nonsense	-
G074	***DNAH11***	Chr7:21775289	c.7472G>C	p.Arg2491Pro	Missense	-
		Chr7:21747335	c.6565C>T	p.Arg2189*	Nonsense	-
G075	***DNAAF1***	Deletion of exon 1,2&3			CNV	-
G076	***RSPH1***	Chr21:43906573	c.275-2A>C	p.Gly92Alafs*10	Splice acceptor	(184)
		Chr21:43906565	c.281G>A	p.Trp94*	Nonsense	(186)
G077	***DNAI2***	Chr17:72306303	c.1494+2dup	p.?	Splice donor	-
G078	***CCDC103***	Chr17:42979839-42979840	c.383dup	p.Pro129Serfs*25	Frameshift	(254)
G079	***DNAH5***	Chr5:13753399	c.10815del	p.Pro3606Hisfs*22	Frameshift	(278)
		Chr5:13758990	c.10384C>T	p.Gln3462*	Nonsense	(278)
G080	***DRC1***	Deletion of exons 11,12&13			CNV	-
G081	***CCDC39***	Chr3:180359783-180359784	c.1871_1872del	p.Ile624Lysfs*3	Frameshift	-
G082	***CCDC103***	Chr17:42978470	c.104G>C	p.Arg35Pro	Missense	(254)
G083	***LRRC6***	Chr8:133645203	c.436G>C	p.Asp146His	Missense	(242)
G084	***MCIDAS***	Chr5:54516635	c.718-1G>A	p.?	Splice acceptor	-
G085	***CCDC151***	Chr19:11537077	c.850C>T	p.Gln284*	Nonsense	-
G086	***CCDC40***	Chr17:78011940	c.48A>G	p.?	Synonymous	-
G087	***CCDC39***	Chr3:180361907	c.1665+1G>T	p.?	Splice donor	-

G088	***CCDC40***	Chr17:78013904	c.387C>G	p.Tyr129*	Nonsense	-
G089	***DNAH5***	Chr5:13792231	c.8320T>C	p.Trp2774Arg	Missense	-
G090	***LRRC6***	Chr8:133627283	c.974+1G>A	p.?	Splice donor	-
G091	***CCDC40***	Deletion of exon 11&12			CNV	-
G092	***DNAH5***	Chr5:13737558	c.11258del	p.Asn3753Thrfs*5	Frameshift	-
		Chr5:13885116-13885117	c.2964_2965del	p.Thr990Asnfs*2	Frameshift	-
G093	***DNAH5***	Chr5:13753399	c.10815T>G	p.Asp3605Glu	Missense	-
		Chr5:13845060	c.5157C>T	p.?	Synonymous	-
G094	***ZMYND10***	Chr3:50380737	c.510+1del	p.?	Splice donor	-
G095	***DNAH5***	Chr5:13820514	c.6782T>G	p.Leu2261Arg	Missense	-
		Chr5:13914037	c.1351C>T	p.Gln451*	Nonsense	-
G096	***CCDC40***	Chr17:78063675	c.2824_2825insCTGT	p.Arg942Thrfs*57	Frameshift	-
		Chr17:78069149	c.2920C>T	p.Gln974*	Nonsense	-
G097	***RSPH1***	Chr21:43906573	c.275-2A>C	p.Gly92Alafs*10	Splice acceptor	(184)
G098	***CCNO***	Chr5:54529084	c.263_267dup	p.Val90Serfs*6	Frameshift	(73)
G099	***ZMYND10***	Chr3:50382928	c.83G>A	p.Gly28Asp	Missense	-
G100	***DNAH11***	Chr7:21932181	c.12646G>T	p.Glu4216*	Nonsense	-
G101	***DNAH11***	Chr7:21641013	c.3426-1G>A	p.?	Splice acceptor	-
		Chr7:21631042-21631056	c.2514_2528delinsC	p.Gln838Hisfs*26	Frameshift	-
G102	***CCDC39***	Chr3:180359783-180359784	c.1871_1872del	p.Ile624Lysfs*3	Frameshift	-
G103	***CCDC39***	Chr3:180359783-180359784	c.1871_1872del	p.Ile624Lysfs*3	Frameshift	-
G104	***DNAH11***	Chr7:21598487	c.563T>C	p.Met188Thr	Missense	-
G105	***DNAI2***	Chr17:72297203	c.883C>T	p.Arg295*	Nonsense	-
G106	***CCDC39***	Chr3:180359783-180359784	c.1871_1872del	p.Ile624Lysfs*3	Frameshift	-
G107	***DNAH11***	Chr7:21659634	c.4438C>T	p.Arg1480*	Nonsense	-
		Chr7:21940815-21940821	c.13494_13500del	p.Ser4498Argfs*15	Frameshift	-
G108	***DNAH5***	Chr5:13762883	c.10229C>T	p.Thr3410Met	Missense	-
		Chr5:13842004	c.5281C>T	p.Arg1761*	Nonsense	-
G109	***RSPH9***	Chr6:43638611	c.760delG	p.Arg254Alafs*76	Frameshift	-
G110	***CCDC39***	Chr3:180337130	c.2182C>T	p.Gln728*	Nonsense	-
G111	***CCDC40***	Chr17:78063675	c.2824_2825insCTGT	p.Arg942Thrfs*57	Frameshift	-

G112	***DNAH11***	Chr7:21940815-21940821	c.13494_13500del	p.Ser4498Argfs*15	Frameshift	-
G113	***CCDC39***	Chr3:180381653	c.210+2T>C	p.?	Splice donor	-
G114	***ZMYND10***	Chr3:50380758	c.490C>T	p.Gln164*	Nonsense	-
G115	***DNAH11***	Chr7:21920318	c.12196-2A>G	p.?	Splice acceptor	-
		Chr7:21742391	c.6244C>T	p.Arg2082*	Nonsense	(281)
G116	***CCDC103***	Chr17:42979839-42979840	c.383dup	p.Pro129Serfs*25	Frameshift	(254)

When only one mutation is reported in the family, this indicates it is a homozygous mutation or hemizygous in an X-linked gene (*PIH1D3, RPGR,* and *OFD1*). When there are two mutations reported in the family, this indicates a compound heterozygous inheritance.

Table 3-4 List of single mutant alleles (single heterozygous) in 5 known PCD genes

Study ID	Gene	Genomic Coordinates (GRCh37)	c. Nomenclature	p. Nomenclature	Impact	Reported before
G117	***DNAH5***	Chr5:13701426	c.13458_13459insT	p.Asn4487fs*1	Frameshift	(125, 143)
G118	***DNAH11***	Chr7:21726760	c.5665G>T	p.Gly1889Trp	Missense	-
G119	***DNAH11***	Chr7:21745115	c.6506C>A	p.Ser2169*	Nonsense	(160)
G120	***HYDIN***	Chr16:70916766	c.10012G>T	p.Glu3338*	Nonsense	-
G121	***HEATR2***	Chr7:794249	c.1048C>T	p.Arg350Trp	Missense	-
G122	***DNAI1***	Chr9: 34459053	c.48+2dup	p.Ser17Valfs*12	Frameshift	(138)
G123	***DNAH11***	Chr7:21747434	c.6664C>T	p.Arg2222*	Nonsense	-
G124	***HYDIN***	Chr16:70917997	c.9805T>G	p.Tyr3269Asp	Missense	-
G125	***HYDIN***	Chr16:71171147	c.950G>A	p.Arg317Gln	Missense	-
G126	***DNAH5***	Chr5:13753399	c.10815del	p.Pro3606Hisfs*22	Frameshift	(278)
G127	***HYDIN***	Chr16:70972589	c.6923A>T	p.Asp2308Val	Missense	-
G128	***DNAH5***	Chr5:13753399	c.10815del	p.Pro3606Hisfs*22	Frameshift	(278)

All the variants listed in this table are single heterozygous with no second mutation identified in these genes in each family.

Table 3-5 List of prioritized variants in 10 candidate PCD genes

Study ID	Gene	Genomic Coordinates (GRCh37)	c. Nomenclature	p. Nomenclature	Impact	ExAc-MAF	Functional Characterization
G129	***DNAH3***	Chr16:20981269	c.8303A>G	p.Glu2768Gly	Missense	Not in ExAc	-
		Chr16:21115787	c.2366+5C>T	p.?	Splice region	0.000445	-
G130	***WDR19***	Chr4:39278645	c.3722C>T	p.Pro1241Leu	Missense	0.0007830	Done
G131	***DNAH9***	Chr17:11725235	c.8708-2A>G	[p.Glu2904Aspfs*53]	Splice acceptor	3.234e-5	Done
		Chr17:11775054	c.10193G>T	p.Arg3398Leu	Missense	Not in ExAc	Done
G132	***CFAP46***	Chr10:134628114	c.7252G>A	p.Val2418Ile	Missense	0.0002007	-
		Chr10:134692953	c.3982G>A	p.Ala1328Thr	Missense	0.003138	-
G133	***C11orf70***	Ch11:101953902	c.776A>G	p.His259Arg	Missense	8.242e-06	Done
G134	***CFAP46***	Chr10:134663843	c.5857G>A	p.Ala1953Thr	Missense	0.0007667	-
		Chr10:134726261	2:c.2397T>G	p.Ile799Met	Missense	0.01365	-
G135	***DNAH14***	Chr1:225195248	c.1107+2T>C	P.?	Splice donor	0.001004	-
G136	***DNAH2***	Chr17:7674655	c.4370T>G	p.Phe1457Cys	Missense	0.0006342	-
		Chr17:7630579	c.368G>C	p.Gly123Ala	Missense	Not in ExAc	-
G137	***C11orf70***	Chr11:101918589	c.154C>T	p.Gln52*	Nonsense	1.653e-05	Done
		Chr11:101937308	c.361C>T	p.Arg121*	Nonsense	8.254e-06	Done
G138	***NME7***	Chr1:169292350	c.278+5G>T	p.?	Splice region	Not in ExAc	-
G139	***TTC30B***	Chr2:178417158	c.334C>T	p.His112Tyr	Missense	Not in ExAc	-
G140	***IFT74***	3 kb genomic deletion encompassing exon 2			CNV	-	Done
G141	***DNAH9***	Chr17:11837266	c.12367G>A	p.Asp4123Asn	Missense	Not in ExAc	Done

When a single mutation is reported in the family, this indicates it is a homozygous mutation. Two mutations indicate compound heterozygosity.

3.2.3 Genetic landscape of PCD

Mutations reported in this study were identified among different categories of PCD genes. *DNAH5* was confirmed as the overall most prevalent single gene with mutations identified in 27 families (21%). This was followed by mutations in *DNAH11* as the second most frequently mutated gene in this cohort where mutations were identified in 16 families (13%). The most prevalent mutation group (affecting together 49 families (38%) of all families with mutations in known PCD genes) were found to affect components of the ODAs (*DNAH5, DNAH11, DNAI1,* and *DNAI2*). The second most common mutation group affects the cytoplasmic dynein assembly factors (*CCDC103, LRRC6, DNAAF3, ZMYND10, DYX1C1, DNAAF1, PIH1D3, SPAG1,* and *HEATR2*) in about 29 families (23%) with mutations in known PCD genes. Mutations in *CCDC39* and *CCDC40* together were identified in 21 families (16%). (**Figure 3-3**)

There was a striking stratification of the genetic basis of disease in different ethnicities, clearly detected when the cohort was divided into three broad ethnic groups: European, Arabic and South-Asians (primarily Pakistani). The most frequently mutated gene was found to be different in different population, for example demonstrated by *DNAH5* mutations which were found in 22 out of 60 families (37%) of European origin, while mutations in the same gene were found in only one family of South Asian origin and in two Arabic families. *LRRC6* and *CCDC103* were the most frequently mutated genes in South-Asian families (9 out of 25 families (36%)). *CCDC39* and *CCDC40* were the two major mutated genes identified in Arabic families (10 out of 24 families (42%)). (**Figure 3-4**)

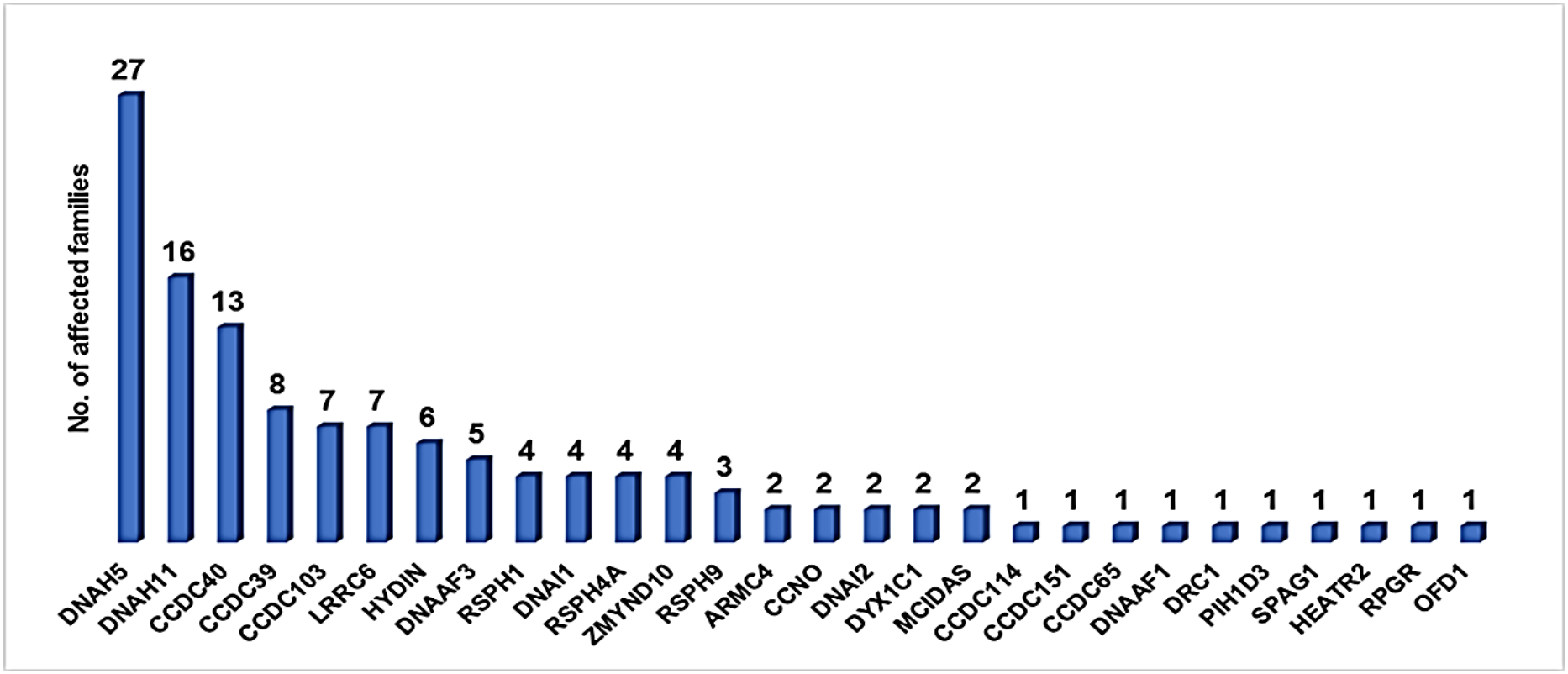


Figure 3-3 Genetic stratification of families with mutations in known PCD genes

Mutations in *DNAH5* are the most frequently identified mutations followed by mutations in *DNAH11* then mutations in *CCDC40* and *CCDC39.*

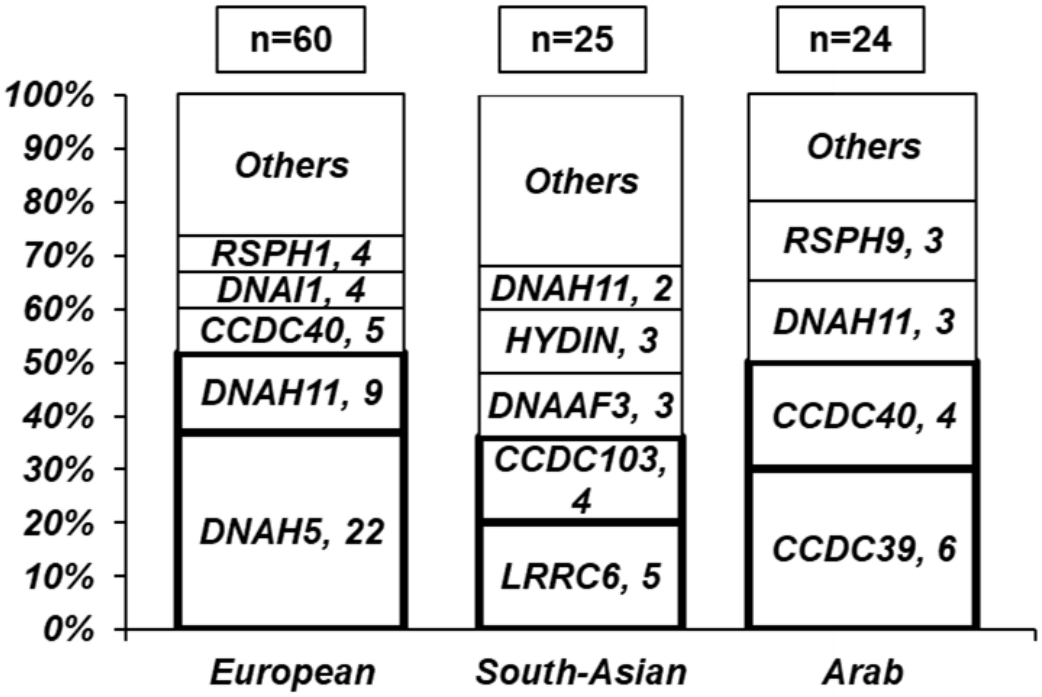


Figure 3-4 Known PCD genes with mutations identified in different ethnicities

A chart summarising the number of families with mutations in different genes in three different populations, illustrating that different ethnicities have different genetic stratification of PCD patients. Amongst 109 European, Arabic and South Asian families, the genes most commonly detected to carry mutations are *DNAH5* in European families (37%), *LRRC6* and *CCDC103* in South-Asian families (36%) and *CCDC39* and *CCDC40* in Arabic families (42%).

3.2.4 Expanding the mutation spectrum in known PCD genes

Homozygous mutations in the previously published autosomal PCD genes were found in 58% of unrelated patients (74 families), from all ethnic groups. 31% of patients (39 families) were found to carry mutations in autosomal genes in a compound heterozygous state and in 3 families, hemizygous mutations were identified in known X linked genes (*PIH1D3, RPGR,* and *OFD1*). In 12 families (9%), only one mutant allele was identified (**Figure 3-5**). In Arabic (23/24 (96%)) and South-Asian (17/25 (68%)) families the majority of mutations in known PCD genes were found to be homozygous. Interestingly, homozygous mutations in known PCD genes were identified in 1/3 of Europeans despite consanguinity having been self-reported in only one family of European origin (**Figure 3-6**).

167 mutations in known PCD genes that were detected in the total 128 PCD families, were classified based upon their predicted impact on the respective encoded proteins. This included a majority (77%) of predicted protein truncating mutations that were consisted of frameshift (32%) and stop-gain or nonsense (26%) mutations, mutations affecting splicing (15%) and copy number variants (CNV, 4%). Inframe deletions or deletion/insertion mutations accounted for 2% and missense mutations were about 21% of all identified mutations (**Figure 3-7**). **Appendix Table A-0-6** summarize the in silico predications of the impact of missense mutations in Polyphen, SIFT, Mutation Taster softwares and their CADD scores.

All the identified variants were classified based on the guidelines for the interpretation of sequence variants developed by the American College of Medical Genetics and Genomics (ACMG).(117) Pathogenic variants (class 5) represent 40% of the variants prioritized in known PCD genes (either in a bi-allelic or hemizygous state) and likely pathogenic variants (class 4) comprised 50% of variants. Class 3 variants of unknown clinical significance represented 10% of the identified variants (**Figure 3-8**).

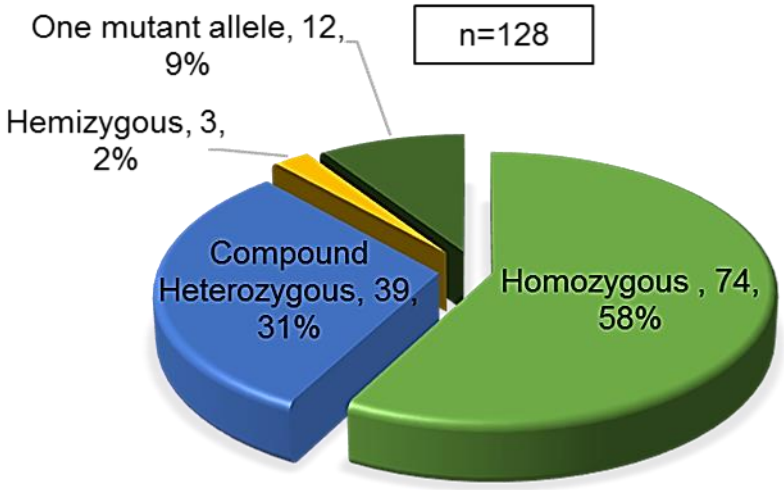
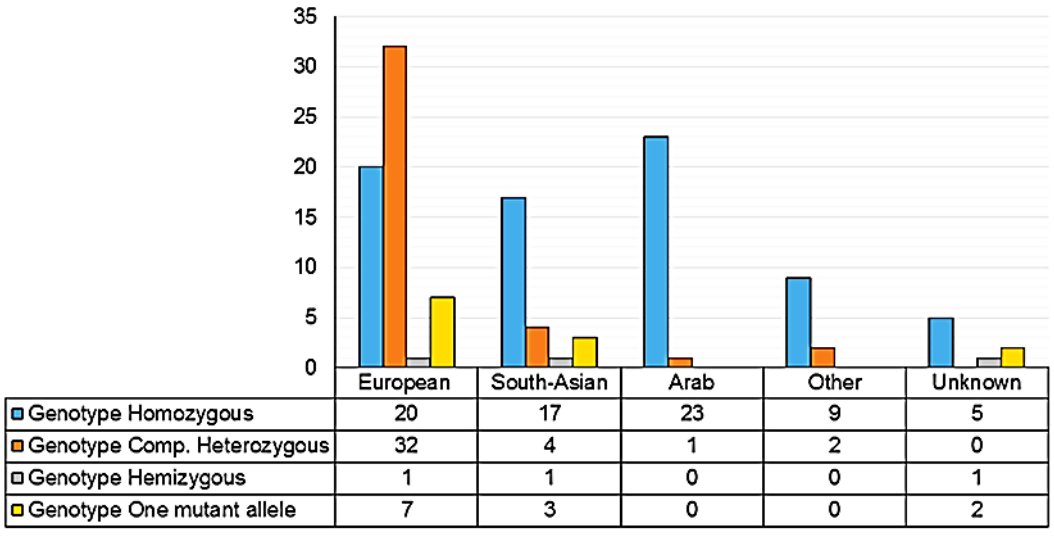


Figure 3-5 Number of unrelated families with mutations in known PCD genes, based on the mutation zygosity state

Out of 128 unrelated families, mutations in known PCD genes were classified according to their zygosity status showing that 58% of mutations identified in this study were present in patients in a homozygous state.



	European	South-Asian	Arab	Other	Unknown
Genotype Homozygous	20	17	23	9	5
Genotype Comp. Heterozygous	32	4	1	2	0
Genotype Hemizygous	1	1	0	0	1
Genotype One mutant allele	7	3	0	0	2

Figure 3-6 Genotype status of mutations in known PCD genes among different ethnicities

Out of 128 unrelated families, in Arab and South-Asian families the majority of mutations in known PCD genes were found to be homozygous. One third of mutations in European populations were also homozygous (20 out of 60 families).

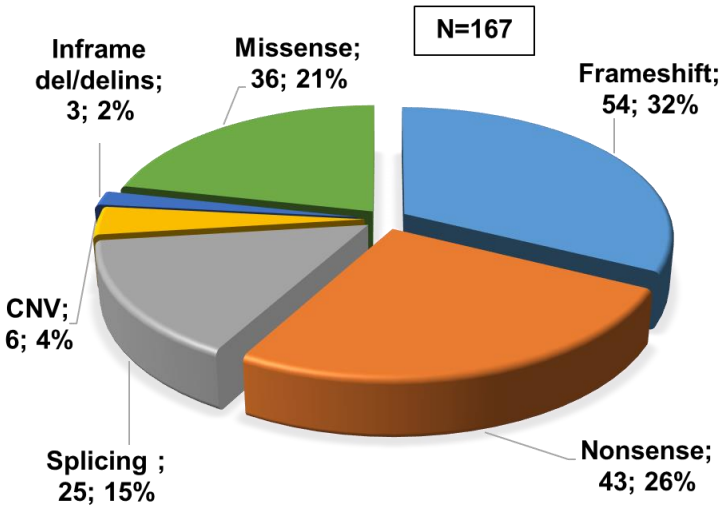


Figure 3-7 Types of mutations found in known PCD genes

Out of 167 mutations (bi-allelic, hemizygous and single heterozygous) identified in 128 families and classified according to their impact on the respective encoded proteins: frameshift and nonsense mutations were the most prevalent (58%) with (15%) splicing defects and (21%) missense changes. Collectively, 77% of the mutations are predicted to have a protein truncating effect (frameshift, nonsense, splicing defects and CNVs).

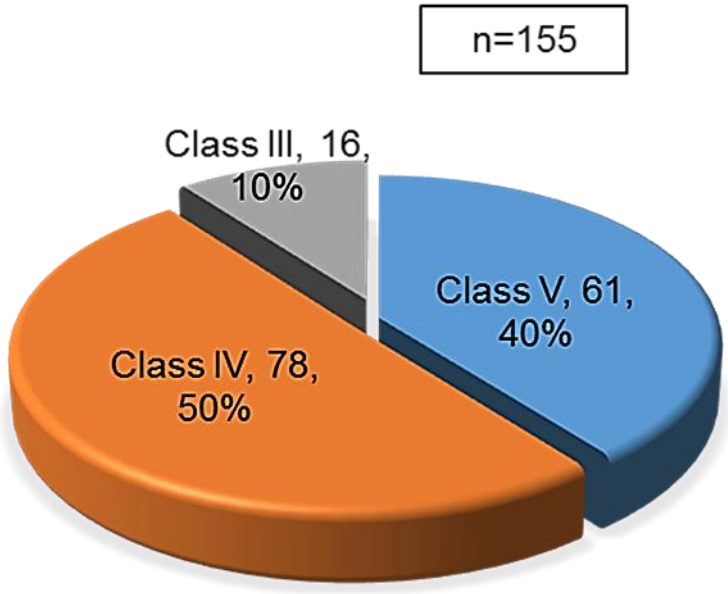


Figure 3-8 Pathogenicity classification of mutations in known PCD genes

Out of 155 mutations identified in 116 families in a bi-allelic state in the autosomal genes or a hemizygous state in the X-linked genes (complete genetic diagnosis) and classified based on the ACMG guidelines, (40%) were class 5 (clearly pathogenic), and (50%) of these mutations were class 4 (likely pathogenic) and only 10% were class III (variants of unknown significance).

3.2.5 Frequency and spectrum of mutations in different ethnicities

A number of common recurrent mutations were found in this study including a homozygous inframe deletion in *RSPH9* (**c.801_803del, p.Lys268del**) detected in two Arabic families. This mutation was previously described in 3 Bedouin families from across the Arab Peninsula.(180, 181) Moreover, another possible Arabic founder mutation, a homozygous frameshift *CCDC39* mutation (**c.1871_1872del, p.Ile624Lysfs*3**) was identified in 4 Arabic families from Palestine. This mutation has not been reported before and it is not present in the ExAc, almena or BinB control databases.

In European families, 3 frequent mutations were identified in *DNAH5*. Two were reported before associated with possible founder effects among Europeans (**c.10815del, p.Pro3606Hisfs*22**) and (**c.13458_13459insT, p.Asn4487fs*1**) (125, 143, 278) and the other mutation is a nonsense mutation (**c.6261T>G, p.Tyr2087***). The latter mutation was identified in two unrelated families, one in a homozygous state and the other in a compound heterozygous state along with another splice region mutation. These two families are originally from Portugal but no consanguinity was reported. This mutation has not been reported before and is not present in ExAc control database.

Three previously reported recurrent mutations were also identified in *DNAI1, RSPH1* and *DYX1C1* in European families.(138, 139, 184, 233) These included a splice-donor mutation in *DNAI1* (**c.48+2dup, p.Ser17Valfs*12**) carried in a homozygous state in affected individuals from two unrelated families and in another family as the only mutation (single heterozygous) allele. A splice acceptor mutation was identified in *RSPH1* (**c.275-2A>C, p.Gly92Alafs*10**) in 4 families (3 homozygous and 1 compound heterozygous) and also a homozygous deletion of 3.5 kb region including exon 7 of *DYX1C1* was found in two unrelated families.

The most frequent mutant allele identified in this study was the previously described frameshift mutation (**c.630del, p.Trp210Cysfs*12**) in the

LRRC6 gene.(252) It has been found in homozygous state in 5 families of South-Asian origin. In the ExAc database, it is present in a heterozygote state in controls from South Asia with an allele frequency of 0.001394 and in European populations with an allele frequency of 2.997e-05. It is also present in the BinB database in carrier state (10/1542).

In *CCDC103*, a homozygous frameshift mutation (**c.383dup, p.Pro129Serfs*25**) has been identified in two unrelated families of South-Asian background. This mutation was previously reported in PCD patients from the same geographical region. Although this variant is not reported in ExAc, it is found once before in heterozygous state in 1542 individuals in the BinB database. Another previously reported *CCDC103* missense mutation (**c.461A>C, p.His154Pro**) has also been detected in families from South-Asian, European and other ethnicities.(254, 255) This variant was detected in a heterozygous state in controls from several populations in the ExAc database (South Asian (0.003331), Latino (0.001296), European non-Finnish (0.001185), European Finnish (0.0001515), and African populations (0.0002893)).

In *CCDC40*, two mutations were present in more than one family. The first one was a previously described frameshift mutation (**c.248del, p.Ala83Valfs*84**) which has been identified in 3 families; two European and one from South Africa.(214) In the ExAc database, this mutation is found in heterozygous state in controls from European (0.0008243), African (0.0002039) and Latino populations (8.64e-05). The other mutation was a novel frameshift mutation (**c.2824_2825insCTGT, p.Arg942Thrfs*57**) identified in 2 families: in a homozygous state in an Egyptian family and in a compound heterozygous state with another nonsense mutation in a Portuguese family. It is not present in any control databases.

A frameshift mutation (**c.13494_13500del, p.Ser4498Argfs*15**) in *DNAH11* was present in a homozygous state in two cousins from Egypt and also in a compound heterozygous state along with another nonsense mutation in a family of European origin. This novel mutation is not reported in any control databases. These recurrent mutations are summarized in **Table 3-6**.

Table 3-6 Frequent mutations in each ethnic group enrolled in the study

Mutation	No. of alleles	Ethnicities	ExAc-MAF	almena-MAF	BinB-MAF	Reported before
CCDC39 (**c.1871_1872del, p.Ile624Lysfs*3**)	8	Arab	Not in ExAc	Not in almena	Not in BinB	No
RSPH9 (**c.801_803del, p.Lys268del**)	4	Arab	0.00005765	3 / 0.00151	Not in BinB	(180, 181)
DNAH5 (**c.10815del, p.Pro3606Hisfs*22**)	8	European	0.0001483	Not in almena	Not in BinB	(278)
DNAH5 (**c.13458_13459insT,p.Asn4487fs*1**)	5	European	0.00005783	Not in almena	Not in BinB	(125, 143)
DNAH5 (**c.6261T>G, p.Tyr2087***)	3	European	Not in ExAc	Not in almena	Not in BinB	No
DNAI1 (**c.48+2dup, p.Ser17Valfs*12**)	5	European	0.0004624	Not in almena	Not in BinB	(138, 139)
RSPH1 (**c.275-2A>C, p.Gly92Alafs*10**)	7	European	0.0003625	Not in almena	Not in BinB	(184)
DYX1C1 (**3.5kb del spanning exon 7**)	4	European	NA	NA	NA	(233)
LRRC6 (**c.630delG, p.Trp210Cysfs*12**)	10	South-Asian	0.000206	Not in almena	10/0.0065	(252)
CCDC103 (**c.383dup, p.Pro129Serfs*25**)	4	South-Asian	Not in ExAc	Not in almena	1/0.0006	(254)
CCDC103 (**c.461A>C,p.His154Pro**)	8	Multiple ethnicities	0.001261	2 / 0.00101	6/0.0039	(254, 255)
CCDC40 (**c.248del, p.Ala83Valfs*84**)	4	Multiple ethnicities	0.0004794	Not in almena	Not in BinB	(214)
CCDC40 (**c.2824_2825insCTGT, p.Arg942Thrfs*57**)	3	Multiple ethnicities	Not in ExAc	Not in almena	Not in BinB	No
DNAH11 (**c.13494_13500del, p.Ser4498Argfs*15**)	3	Multiple ethnicities	Not in ExAc	Not in almena	Not in BinB	No

3.2.6 Synonymous mutations predicted to affect splicing can cause PCD

Two synonymous mutations were identified. They are predicted to affect splicing. In an Arabic family (G086), a homozygous synonymous change (**c.48A>G, p.Gly16Gly**) was found in *CCDC40*. The mutation was confirmed by Sanger sequencing and correctly segregated within the family. The TEM analysis revealed classic *CCDC40*-associated microtubular disorganization with IDA loss in the affected children from this family. This mutation is not found in any control databases including ExAc, dbSNP, 1000Genomes, almena and BinB databases. Using Human Splicing Finder 3.0 analysis, it was predicted to activate an exonic cryptic donor site (**Figures 3-9& 3-10**).

In a European family (G093) where TEM showed ODA loss from the cilia of the probands, compound heterozygous mutations were identified in *DNAH5* with one allele being another synonymous mutation (**c.5157C>T, p.Phe1719Phe**) predicted to affect splicing through creation of an exonic splicing silencer (ESS) and alteration of an exonic splicing enhancer (ESE) site as predicted by Human Splicing Finder 3.0 (**Figure 3-11**). This variant was found only once in a heterozygous state in ExAc with a total allele frequency of 8.241e-06 and it was not found in the almena or BinB databases. By Sanger sequencing, this variant was confirmed in the patient but it was not present in the mother (the father's DNA was not available). In this case, the other *DNAH5* allele was a missense mutation (**c.10815T>G, p.Asp3605Glu**) which was absent from ExAc, dbSNP, 1000Genomes, almena and BinB databases. This variant was also confirmed by Sanger sequencing and the mother was found to be a carrier. This mutation is predicted to be damaging using SIFT and Mutation Taster *in silico* predication tools with a pathogenicity score of 1.000 (sensitivity: 0.00; specificity: 1.00) using Polyphen-2. It has a CADD score of 24.1 (CADD score ≥ 20 indicates the variant is among 1% most deleterious substitutions in the human genome). (**Figure 3-12**)

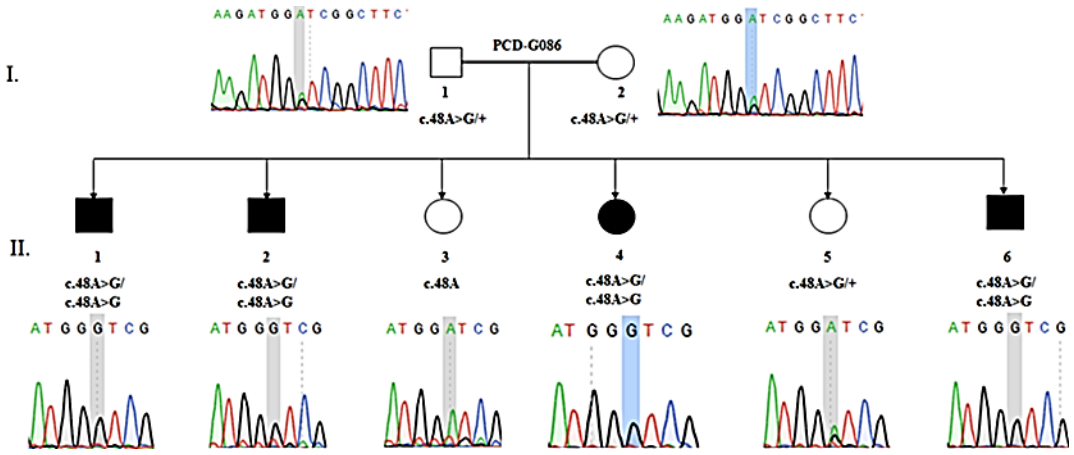


Figure 3-9 Familial segregation of a CCDC40 synonymous mutation in an Arabic family (G086) with cilia microtubular disorganization and IDA loss

The mutation is homozygous in 4 affected siblings while the two parents and one of the normal siblings were found to be heterozygous carriers.

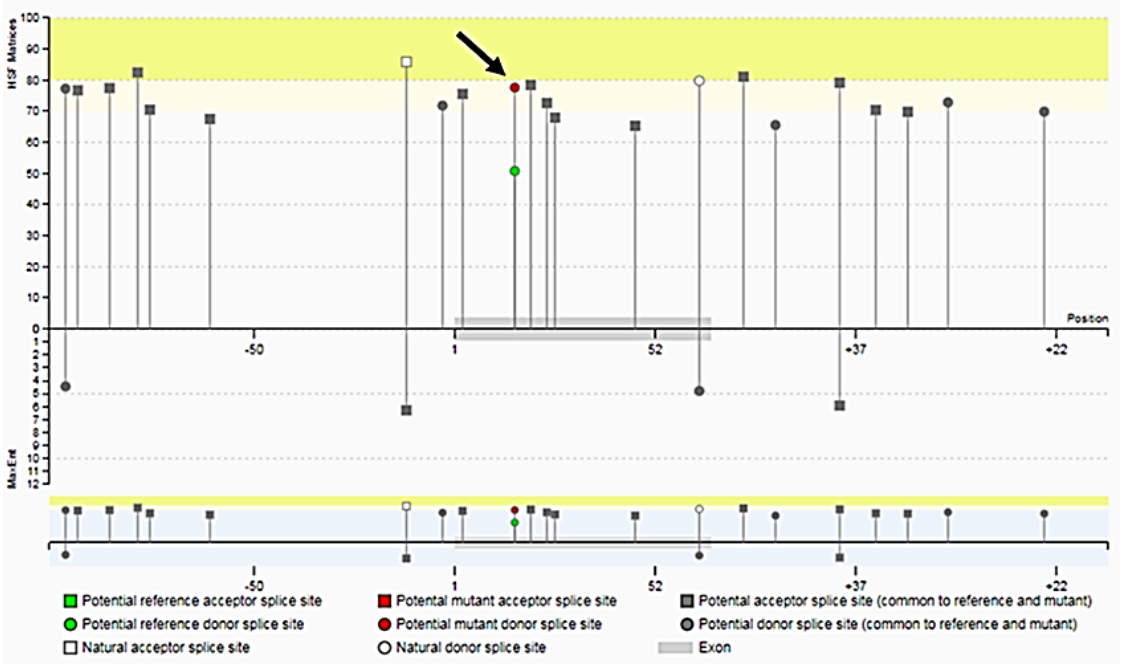


Figure 3-10 Predicted exonic cryptic donor site associated with a c.48A>G synonymous CCDC40 mutation in family (G086)

Analysis used the HSF Matrices algorithm in the Human Splicing Finder 3.0 tool which identified the predicted exonic cryptic splice site (red circle).

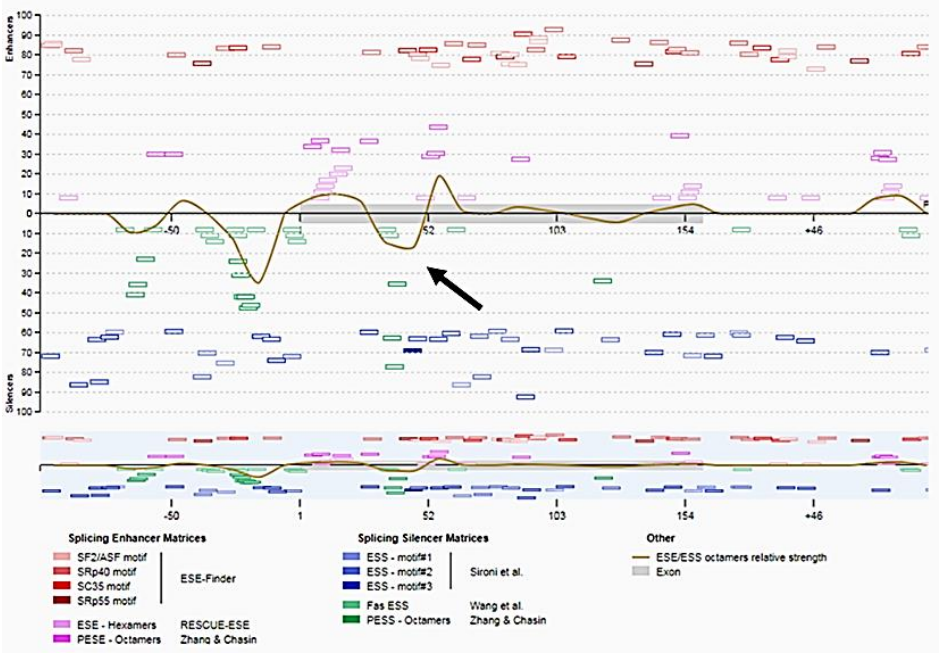


Figure 3-11 Predicted creation of a new exonic splicing silencer due to c.5157C>T; p.Phe1719Phe synonymous mutation in DNAH5, in family (G093)

Analysis was done using PESS Octamers and IIEs algorithms within the Human Splicing Finder 3.0 tool. This predicted the creation of an exonic splicing silencer site.

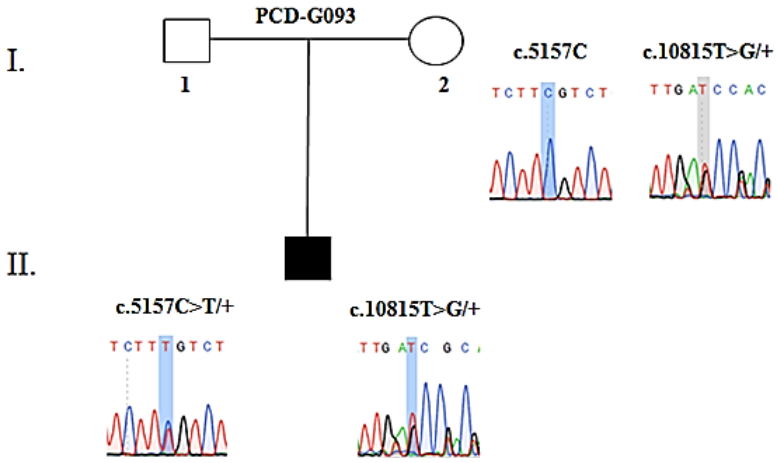


Figure 3-12 DNAH5 compound heterozygous mutations in a European family (G093) with ODA loss

One mutation (c.5157C>T; p.Phe1719Phe) is a synonymous missense variant predicted to affect splicing. The other variant (c.10815T>G; p.Asp3605Glu) is a missense mutation. The mother was found to be a heterozygous carrier of the missense mutation only (father's DNA not available).

3.2.7 Copy Number Variations (CNVs) can cause PCD

ExomeDepth software was used to call CNVs in each batch of sequencing. It uses read depth data from each sample and compares it to a matched aggregate reference set which combines data from the same batch. CNVs were detected in 6 families in the current study. There were homozygous deletions found in 5 families including patients from two families with a previously reported 3.5 kb deletion in *DYX1C1* in addition to single cases of homozygous deletions in *CCDC40, DRC1* and *DNAAF1*. One heterozygous deletion encompassing exons 68-75 of *DNAH11* was found in one family where another missense mutation (**c.13040T>C, p.Leu4347Pro**) was also identified in the same gene. (**Figure 3-13**)

3.2.8 Targeted NGS is a powerful tool for diagnosis and characterization of PCD patients

Clinical diagnostic TEM analysis results were available in 134 of the 161 families (83%) and ciliary ultrastructural defects were identified in 97 families (60%). Inconclusive TEM results were found in 37 families. Bi-allelic mutations in known PCD genes were identified in 18 (67%) out of 27 families in which the affected children had a clinical history suggestive of PCD but without TEM analysis. (**Figure 3-14**)

Grouping of the genotyped patients according to the different functional/structural classes of the causal genes (dynein arm genes, ODA docking complex, radial spokes, central complex, nexin link components and dynein assembly genes) showed that there is a well-defined correlation between genotypes and TEM defects as shown in **Table 3-2**. However, there was some variability in the TEM results, for instance, inconsistent TEM results could be found associated with *DNAH5* mutations which were detected in both ciliary ODA loss and combined loss of IDA+ODA patients. A degree of variable ciliary ultrastructural phenotype was also noted in individuals with mutations in dynein

assembly genes (**Table 2-3**). This finding supports the previously published reports that the TEM clinical test lacks sensitivity.(61, 78)

To investigate TEM and genetic defect correlation inconsistencies that may be present, especially between different diagnostic centres or samples analysed at different times, quantitative TEM re-analysis was conducted in a group of patients ascertained from the same clinical diagnostic service (Royal Brompton Hospital, London) with different mutations in *DNAH5* compared to a number of dynein assembly genes (*LRRC6, HEATR2, DYX1C1, DNAAF3, DNAAF1*). This can evaluate the relative value of both techniques.

Mutations in all five dynein assembly genes were consistently found to cause combined IDA+ODA loss in the majority of cilia cross sections examined. Also, *DNAH5* mutations consistently cause mainly ODA loss with significantly less combined IDA+ODA loss or normal cilia ultrastructure. (**Figure 3-15)**

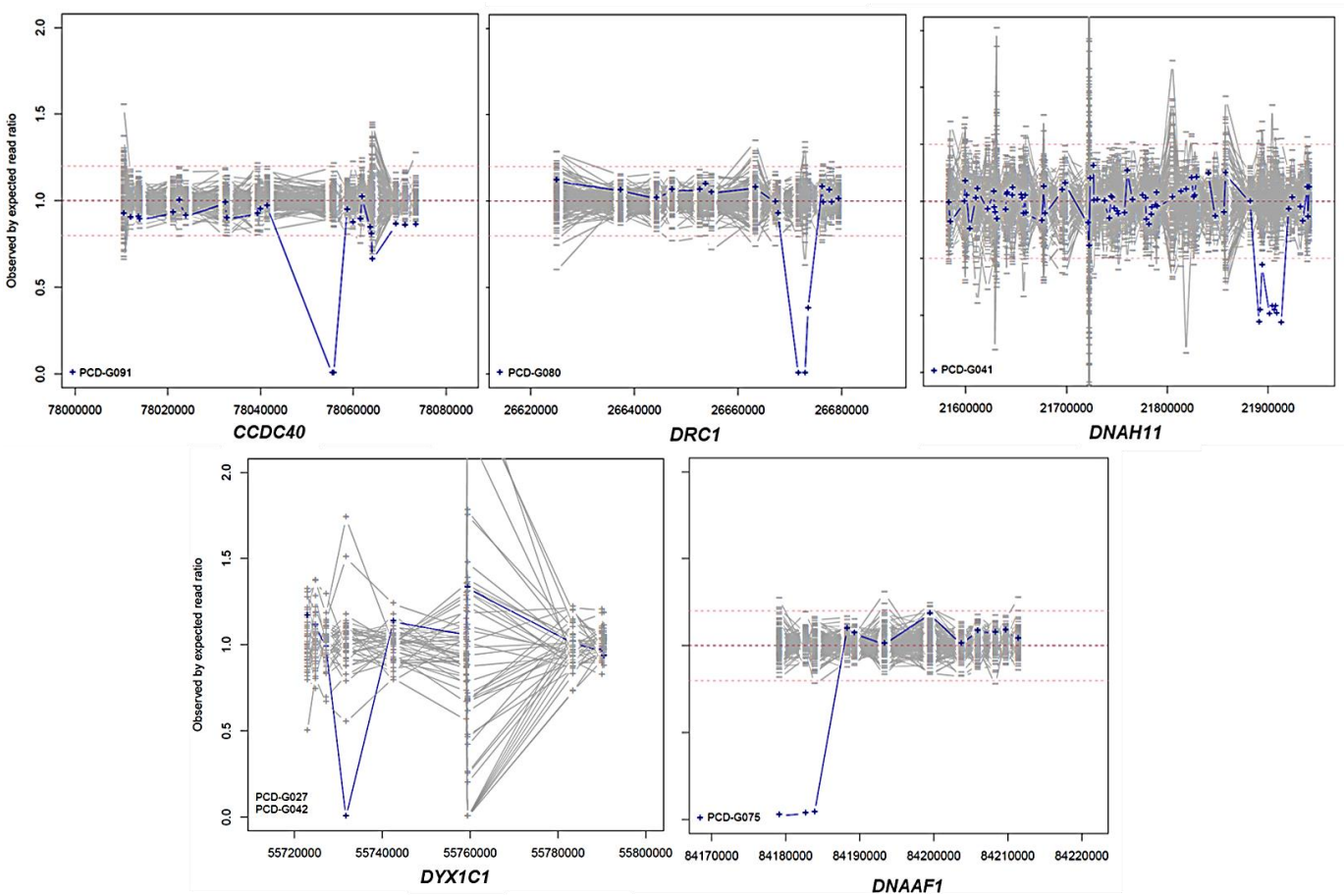


Figure 3-13 ExomeDepth software used for CNV analysis of the sequencing data

ExomeDepth analysis was able to identify homozygous deletions in *CCDC40, DNAAF1, DRC1* and *DYX1C1 (the latter affects two families).* Also a heterozygous deletion of exons 68-75 of *DNAH11* was detected in the family (G041).

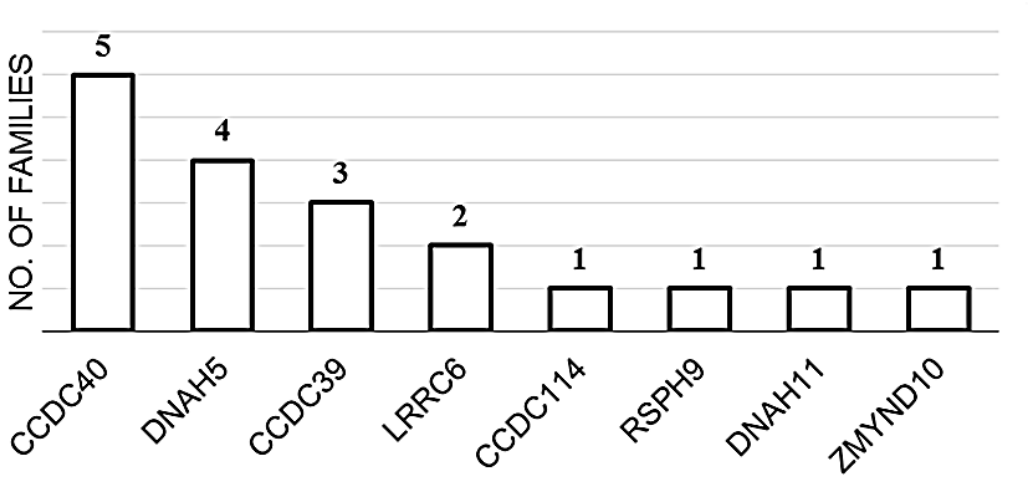


Figure 3-14 Mutant PCD genes in patients without TEM analysis

Of 161 families with PCD confirmed or highly suggested through diagnostic workup in this study, 17% (27 families) did not have TEM results (not available or insufficient sample obtained from nasal brushing). Bi-allelic mutations in known PCD genes were identified in 18 of these families (67%).

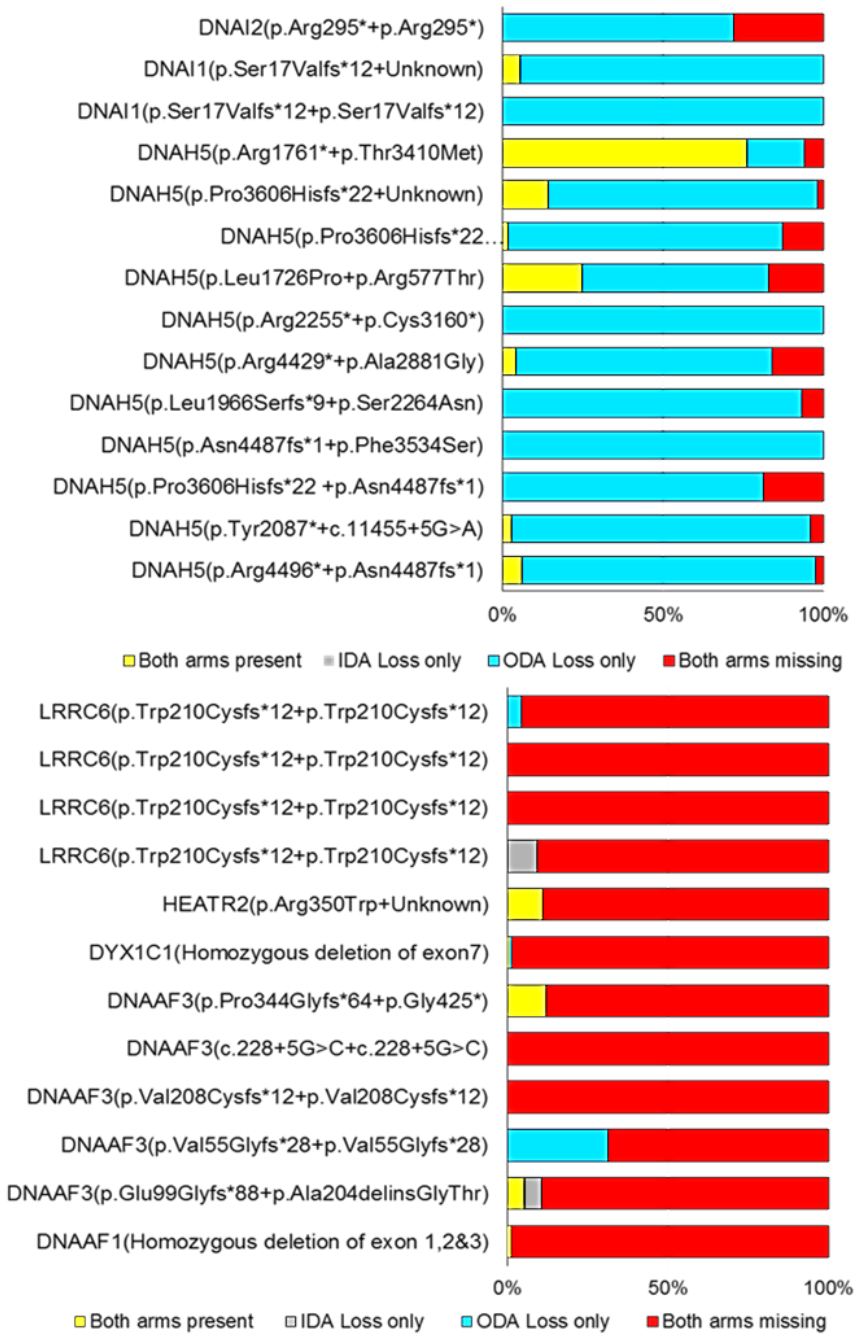


Figure 3-15 Genetics can better characterize PCD patients

The percentage of cilia cross sections showing a loss of either or both the inner and outer dynein arms was recorded in the routine TEM diagnostic setting at Royal Brompton Hospital with reference to the underlying genetic defects. Patients carrying bi-allelic mutations in dynein assembly genes showed a combined loss of both arms in the majority of cross sections examined. Patients carrying bi-allelic mutations in outer dynein arm components showed mainly an isolated loss of outer dynein arms. Combined IDA+ODA loss was also noted. Interestingly, quite variable numbers of cross sections showed a normal ultrastructure of the cilia.

N.B. DNAH5 (P.Pro3606Hisfs*22..) is a homozygous mutation.

3.3 Discussion

PCD causing mutations have been reported in 36 genes so far. They have been previously estimated to account for variants identified in about 70% of PCD patients.(124) Such extensive genetic heterogeneity makes Sanger sequencing an outdated, time-consuming and expensive approach for genetic diagnosis of PCD patients. Thus, the use of multi-gene panels including all the known PCD genes, especially with recently developed library preparation techniques and advanced data analysis software, offers a valuable, efficient tool for genetic diagnosis of PCD, which will save time and cost compared to Sanger sequencing.(122) Targeted sequencing is also more effective than whole exome sequencing (WES) and whole genome sequencing (WGS) in terms of the level of sequence coverage of the target regions in a short time and reduced costs.(123)

In the current study, a customized targeted next generation sequencing panel was designed for genetic screening of a large cohort of PCD patients from various ethnicities (European, Arab, South-Asian and others). To the best of my knowledge, this can be considered the largest cohort of PCD patients of a multi-ethnic background to be screened in one study using this sequencing approach. The gene panel was designed to include all the known PCD genes, at the time of design, and a large list of other candidate genes for potential novel gene discovery. The design of the panel was performed after extensive literature review to include all genes linked to PCD and also other candidate genes with possible roles in cilia motility based on genomic and transcriptomic studies of collaborators working on *Chlamydomonas, Drosophila*, mouse node and zebrafish with additional curation of genes. This approach is different from other targeted sequencing studies conducted on PCD patients since usually they have screened only known PCD genes, apart from the recently reported study where screening of 310 known and candidate genes in Dutch PCD patients was conducted.(280, 282, 283) This genetic screening helped to better characterize and further expand the genetic landscape of PCD, elucidate peculiar ethnic differences and to confirm diagnosis in cases with inconclusive clinical diagnostic results.

Patients from 161 unrelated families with a confirmed or possible diagnosis of PCD were enrolled in this study. The diagnosis of PCD is usually difficult and involves a complex diagnostic testing workflow. PCD patients have recurrent upper and lower respiratory tract infections that can start early in life with neonatal respiratory distress and they can have later complications with the development of bronchiectasis. Patients also manifest with recurrent sinusitis and middle ear effusion. About 50% of PCD patients have laterality problems (malpositioning of the internal organs). In adult life, male patients and to a lesser extent female patients, have infertility and females are also thought to have an increased risk of ectopic pregnancy. Associated with these variable clinical symptoms, the clinical diagnosis of PCD is usually delayed due to overlap of these symptoms with many other disorders.(44)

As there is no gold standard diagnostic test for PCD, identifying ciliary ultrastructural defects by TEM analysis of cilia cross sections was considered as a hallmark for confirming the diagnosis of PCD according to the European Respiratory Society Guidelines.(58) Evident clinical presentation highly suggests PCD diagnosis, even with normal or inconclusive findings of TEM analysis. As there is no standard diagnostic testing criteria to exclude PCD in any individual, a combination of results from different tests to be interpreted in the light of clinical symptoms is essential to augment the accuracy of PCD diagnosis.(58)

For complete genetic diagnosis of PCD, bi-allelic mutations in an autosomal recessive gene or a hemizygous mutation in an X-linked gene is required. Highly complete genetic diagnostic output is provided in this study by using a targeted gene panel (with CNV analysis), which offers a confident genetic diagnosis rate of about 72% for patients with confirmed (by clinical functional tests) or highly suggested PCD. This is higher than the recently reported 67.6% diagnostic yield using targeted sequencing in 74 Dutch PCD patients.(283) This is also much higher than 43% diagnostic yield of using a targeted NGS panel of 24 PCD genes to screen 44 unrelated individuals in Italy.(282) Moreover, this is higher than the 67% output of combining Sanger sequencing of two PCD genes, *DNAH5* and *DNAI1* with targeted NGS of 19 genes in 33 unrelated families from the Czech Republic.(280) However, It is

similar to the 76% diagnostic success rate achieved by the use of WES and targeted CNV analysis in patients with no genetic diagnosis after Sanger sequencing of 12 known PCD genes in 52 individuals in Canada despite the fact that the number of patients in the current study is over triple the number of patients enrolled in the Canadian study.(114, 279)

A number of limitations have been encountered in using targeted exome sequencing for PCD. These include difficulties when only one mutant allele is identified in a known PCD gene, giving an incomplete genetic diagnosis. There are also technical drawbacks of variable coverage depth and bioinformatics challenges to identify CNVs. Additional potential candidate genes may not have been included in the panel design, hence its comprehensiveness is unknown until all genes involved in cilia motility are defined.

For 13 out of the 161 families in the study, mutations in novel candidate genes were identified in the affected individuals, in the context of no mutations having been found in any of the known PCD genes. Further work at the functional level is required to confirm disease causality of the identified novel alleles and for 6 of the 10 putative novel PCD genes, the confidence around them remains affected by low patient numbers or uncertain significance of the variants e.g. if they are missense changes. For the other 4 putative new PCD genes (*C11orf70, DNAH9, IFT74* and *WDR19*), the functional work performed to evaluate their disease-causing status is summarized in **Chapters 4-6**.

In the other 20 families of which 11 families had a confirmed diagnosis of PCD by identifying ultrastructural defect by TEM, no putative sequence variants could be identified. These results reflect the project limitations and in these cases the disease may be due to mutations affecting splicing in the deep intronic regions as they are not covered by the NGS panel design. This class of mutations has not yet been readily studied in PCD. To confirm these possibilities, WGS to cover sequencing of the intronic regions (122) and RNA-Seq offer deeper and more comprehensive sequencing that will be required to detect possible aberrant expression or mono-allelic expression of the mutant allele.(284) Also, since some genomic regions may not be fully

covered by NGS, complimentary Sanger sequencing of these regions may be required to identify any other missed mutations.

Mutations in 28 known PCD genes were identified in 116 out of the total 161 PCD families, achieving a complete genetic diagnosis (72%) with bi-allelic or hemizygous mutations. A total of 155 separate disease alleles were identified, 78 of which (50%) are novel to this study and not previously described. This number of genes with identified mutations is of a much higher number and spread than any of the previous targeted sequencing studies done to screen PCD patients. For example, in the recent Dutch study screening 310 genes, they only identified mutations in 12 known PCD genes.(283) This reflects the genetic heterogeneity identified through this study, that has most likely arisen from screening a large cohort of PCD patients from various ethnic backgrounds. In about half of these known PCD genes, mutations in each gene were identified in 3 or more families. In contrast however, in 10 genes there was only one family with mutations identified in each gene.

There is some degree of correlation between genes that have a difficult to define clinical diagnosis through cilia function tests (e.g. *DNAH11* and *HYDIN* genes that have subtle structural defects not well detected by conventional TEM, or HSVM since some cilia motility is retained) (161, 196) and a lack of recurrent alleles (no *DNAH11* or *HYDIN* recurrent alleles out of a total of 21 detected in this study), supporting the possible lack of detection and under-diagnosis of these kinds of ciliary defects, in contrast for example to *DNAH5* disease which has robustly detectable TEM and HSVM results with three recurrent mutations have been identified in the current study. Since these are all large genes it seems less likely that this finding arises from their having more technically difficult to detect disease alleles, such as for example deep intronic variants or ones less well covered by the NGS method, although the *HYDIN2* copy pseudogene is additionally a known technical block to efficient *HYDIN* mutation screening.(196)

Notably, it is interesting to find that one third of mutations identified in the outbred Caucasian European families screened in this project, were

homozygous despite the low consanguinity rate between Europeans (only one European family reported consanguineous marriage). This reflects other mechanisms than consanguinity, perhaps unrecognised endogamy, or mutational hotspots more connected to protein functions. It raises the possibility of the existence of a founder effect underlying the inheritance of some of these variants. In 14 of these families, the identified mutations were recessive alleles that had been previously reported in the literature.

Although *DNAH5* is considered to be the most commonly mutated gene causing PCD, (143) and overall within the cohort this was found to be true, this was not the case within different ethnicities. *DNAH5* mutations were identified in 37% of European families, confirming it as the most common mutant gene in Europeans. On the other hand, mutations in *LRRC6* and *CCDC103* were the most prevalently mutated genes in South-Asian families with only one family with *DNAH5* mutations. In Arab families, mutations in *DNAH5* were identified in only two families, whilst mutations in *CCDC39* and *CCDC40* were the commonest. These data highlight a significant ethnic impact on the genetics of PCD patients and ensure the importance of including PCD patients from various ethnicities to elucidate the full genetic landscape of PCD. This conclusion will be helpful to prioritize certain genes for genetic testing in certain ethnicities especially in countries with limited genetic resources.(76)

The majority of mutations identified in known PCD genes (more than 75%) were protein truncating mutations. This is consistent with types of mutations reported before in PCD genes by our group and others.(80) About 50% of all the mutations detected in known PCD genes could be classified according to the guidelines of the American College of Medical Genetics and Genomics (ACMG) as likely pathogenic, not reported before variants, expanding the genetic landscape and mutation spectrum of PCD. Despite the widely accepted belief that most PCD mutations are private, (58, 124) as was clear in the 10 genes with only one family with mutation identified in each gene, certain mutations in this study have been detected in more than one of the unrelated families. These mutations appear to be more frequent in certain populations and thus can be used for carrier screening panels in certain

populations. This can also be useful to design a preliminary targeted allele-specific genetic testing in countries where NGS facilities cannot be afforded or not widely available.

Synonymous mutations can be easily overlooked and are not commonly reported in PCD. Here two synonymous mutations were identified in two unrelated families which predicted an alteration of gene splicing. This highlights the importance of looking for potential synonymous variants in unsolved cases. The interpretation of such variants is very challenging and it is essential for the right genetic diagnosis to be consistent with the expected ciliary ultrastructural defect in the patients. This imposes another limitation for the genetic testing to be the first gold standard test for PCD diagnosis.

The identification of ciliary ultrastructural defects by TEM analysis remains one of the most important steps in the PCD diagnostic workflow, though failure of identifying defects by TEM does not exclude PCD.(58) However TEM analysis is expensive and is additionally a highly skilled and operator dependent technique. It is estimated that up to 30% of PCD patients may have normal or inconclusive TEM result (161, 285) as was the case for 37 families in this study. In this study, pathogenic variants were identified in known PCD genes in 80% of patients with either normal or inconclusive TEM findings. It can therefore be proposed that genetic testing should be considered as a pivotal step in the diagnosis of this group of patients where the ultrastructure of the cilia is apparently normal or there are inconclusive findings by conventional clinical TEM.(161, 285, 286) This is estimated to be the case for up to 20% of PCD patients in the UK national PCD service.(78)

Copy number variations remain an important cause of many Mendelian disorders.(287, 288) Identification of CNVs and large structural variants while using whole exome sequencing or custom gene panel sequencing is very challenging.(289) Although whole genome sequencing has the potential to overcome this challenge, the cost and burden of data management remain the major obstacles within the clinical diagnostic context.(287) CNV calling from targeted sequencing data has not been widely used, however CNVs were reported before in PCD patients identified by other genetic screening

methods.(224, 233, 279, 290) In the current study, CNV analysis using ExomeDepth software of targeted sequencing data has newly identified CNVs within PCD genes that added to the diagnostic yield of the panel sequencing and expanded the mutation spectrum in PCD patients. There are many determinant factors that should be considered for successful identification of CNVs when applying algorithms such as ExomeDepth in targeted sequencing data including amplification bias (bias of using amplification-based library preparation methods), the distances between targeted exons, the variability of sequencing coverage within samples and the extent of read depth.(261, 289) A lack of calling of CNVs using this approach does not exclude CNVs in the screened genes, but its use is still of good positive diagnostic value to be considered in the clinical context.

Genetics helps in the diagnosis of PCD in cases where TEM analysis is not available but there is a strong clinical history (mutations were identified in 70% of this group). While there is an increasingly well-defined genotype ultrastructural phenotype correlation underlying PCD (78), variability in the TEM diagnostic analysis results are well documented even with mutations in the same gene. TEM versus genotype re-analysis here showed that most of *DNAH5* mutation cases display cilia ODA defects, but there can also be combined IDA+ODA loss recorded or inconclusive TEM results. This variance may largely be due to difficulties in evaluating the presence or absence of IDAs by TEM.(291) The percentage of arm loss in *DNAH5* mutations was quantitatively assessed compared to mutations in dynein assembly genes. This showed that mutations in *DNAH5* lead to a loss of the majority of ODA arms, while dynein assembly mutations lead to combined loss of most of IDA and ODA arms. This finding highlights the value of genetic testing for better characterization of PCD patients.

3.4 Summary

In summary, I have described the use of targeted multi-gene panel sequencing for genetic screening of PCD patient. A high diagnostic yield of about 72% was achieved among patients with confirmed or highly suggestive PCD diagnosis. This improved the diagnostic workflow outcome by confirming PCD in patients with inconclusive TEM results and helped in the diagnosis of patients where TEM analysis was not available. A notable impact of ethnicity on the genetic variability of PCD was identified. Despite the established concept that most of PCD mutations are family-specific private mutations, a list of mutation that are more frequent in certain ethnicities or more universally present in multiple ethnicities, were identified. Unusual sequence variants including synonymous variants potentially affecting splicing as well as CNVs were newly reported within this study.

The following three chapters describe further work on four putative candidate PCD genes of interest identified through this NGS sequencing cohort study (**Table 3-5**), to investigate their likely roles in cilia motility and disease aetiology.

Chapter 4 Mutations in *C11orf70* disrupt the dynein assembly process and cause PCD

4.1 Introduction

Axonemal dyneins are large multi-protein complexes acting as the motor components of the axoneme that drive cilia motility.(292) Cilia contain inner (IDA) and outer (ODA) dynein arm motors that work together to power and regulate the ciliary beating and waveform.(293, 294) They are composed of molecular subunits (heavy, intermediate and light dynein chains) that are pre-assembled in the cytoplasm and then transported into the cilia to be docked to the axonemal peripheral microtubules.(15, 219) This multi-step process requires cofactors and chaperones in addition to essential transport agents, which are not components of the mature axonemal dyneins.(217)

As the ciliary axoneme is highly conserved across species from the unicellular eukaryotes to mammals (218), delineation of the assembly mechanisms of the axonemal dyneins in the cytoplasm has been attempted using human genetics and PCD model organism studies.(217) Most of the current knowledge about dynein assembly and transport came from studies on *Chlamydomonas*, a protist with motile flagella.(15, 112, 217, 226, 294-296)

Most of PCD causing mutations were identified in genes encoding structural components of the cilia. As shown in the previous chapter, this is dominated overall by mutations affecting components of axonemal dynein arms like *DNAH5* and *DNAH11* as subunits of outer dynein arm motors. On the other hand, there is a growing number of loci linked to PCD that do not encode for axonemal structural proteins but rather encode for proteins involved in the process of assembly of dynein motor complexes in the cytoplasm and subsequently their transport into the cilia.(124)

Mutations in genes implicated in dynein assembly, were identified to cause PCD with combined loss of both inner and outer dynein arms. Most of these proteins are found exclusively in the cytoplasm where they act as co-chaperones essential for proper folding and maturation of axonemal dynein complexes. This group of proteins include *DNAAF1*, (94, 224, 227) *DNAAF2*, (112) *DNAAF3*, (232) *DYX1C1*/*DNAAF4*, (97, 233) *HEATR2*/*DNAAF5*, (238, 240) *LRRC6* (242, 244, 297) *PIH1D3* (298, 299) *ZMYND10* (248, 252) and

C21orf59.(98, 253) These proteins act as assembly factors in a similar way to chaperones, since when mutated they affect the assembly of motor dyneins.

Once the axonemal dyneins are assembled in the cytoplasm, they need to be transported to the cilia passing the transition zone barrier.(300) This transport process requires intraflagellar transport (IFT).(301, 302) Transport of such cargoes along the cilia requires cargo adaptors to mediate and stabilize the interaction with IFT trains. *Chlamydomonas* ODA16 (WDR69/DAW1 in human) was identified as an IFT cargo adaptor for ODAs as it interacts with the N-terminus of IFT46 protein.(96, 296, 303, 304)

This chapter summarizes the use of panel sequencing for novel gene discovery in PCD relating to the *C11orf70* gene (*CFAP300*). Mutations in *C11orf70* cause loss of both inner and outer dynein arms, cilia immotility and consequently result in PCD. Various genetic approaches have been used to identify mutations in human genes implicated in the dynein assembly process through screening of affected individuals with combined loss of inner and outer dynein arms.(112, 232, 233) Here, studies in human cells and in two model organisms: *Paramecium* and *Chlamydomonas,* reveal a highly conserved role for C11orf70 in dynein assembly.

4.2 Results

4.2.1 *C11orf70* mutations in PCD patients

Mutations in the *C11orf70* (*CFAP300*) gene were found in 3 patients from two unrelated families where PCD diagnosis was confirmed by other clinical testing modalities at Royal Brompton Hospital, London (summarized in **Table 4-1**). The affected individuals from both families presented with classical PCD symptoms including a positive history of neonatal respiratory distress. Dextrocardia and situs inversus were present in affected children from the two families (one sibling from family G133 has levocardia). Nasal NO was measured in only the eldest affected sibling of family G133, as his age was appropriate, showing extensively reduced levels (28 nl/min) that are below the diagnostic cut off level of 77 nl/min.(305) High speed video microscopy (HSVM) examination of nasal brushing biopsies of the patients showed completely immotile cilia. Transmission electron microscopy (TEM) analysis of cilia of the respiratory epithelial cells obtained by nasal brushing biopsies revealed a predominant loss of both inner and outer dynein arms (**Figure 4-1**). Immunofluorescence (IF) analysis confirmed the absence of DNAH5 (ODA component) and DNALI1 (IDA component) from the patient's respiratory cilia (**Figure 4-2**).

Table 4-1 Summary of the clinical data of patients carrying C11orf70 mutations

ID	NRD	Wet cough	Rhinitis	Cardiac situs	nNO (nl/min)	HSVM	TEM
G133 II:2	YES	YES	YES	Levocardia	28	immotile	IDA+ODA loss
G133 II:3	YES	YES	YES	Dextrocardia	N/A	immotile	IDA+ODA loss
G137 II:1	YES	N/A	N/A	Dextrocardia	N/A	immotile	IDA+ODA loss

NRD; Neonatal Respiratory Distress, nNO Nasal Nitric Oxide, HSVM; High Speed Video Microscopy, TEM; Transmission Electron Microscopy, N/A; Not applicable.

A homozygous missense mutation (**c.776A>G, p.His259Arg**) was identified in two affected siblings in a Pakistani consanguineous family (PCD-G133) in which the parents were first cousins. Sanger sequencing was performed to confirm the homozygous status of the mutation in the patients, carrier status of the parents and in another unaffected sibling. This variant was found only once in a heterozygous carrier in South Asian population in ExAc database with total allele frequency of 8.242e-06. It is absent from the EVS, 1000G and dbSNP databases. Phylogenetic analysis reveals that this mutation affects a highly conserved amino acid **His259** within the DUF4498 domain (a domain of unknown function) of the C11orf70 protein. Using Polyphen-2 software, this mutation was predicted to be probably damaging with a score of 0.997 (sensitivity: 0.27; specificity: 0.98). It was also predicted to have a damaging effect on the protein by both SIFT tool and Mutation Taster. Moreover, it had a CADD score of 23.9. (**Figure 4-3**)

The affected child of the other Indian non-consanguineous family (PCD-G137) had two nonsense mutations, one (**c.154C>T, p.Gln52***) inherited from the father and the other (**c.361C>T, p.Arg121***) inherited from the mother. Sanger sequencing confirmed the mutations in the patient's DNA and the carrier status of the parents. In the ExAc databases, the p.Gln52* mutation (rs767760877) was found in 2 heterozygotes in the South Asian population with total allele frequency of 1.653e-05. The p.Arg121* mutation (rs561237622) was found only once in a heterozygous state in the South Asian population with a total allele frequency of 8.254e-06. (**Figure 4-4**)

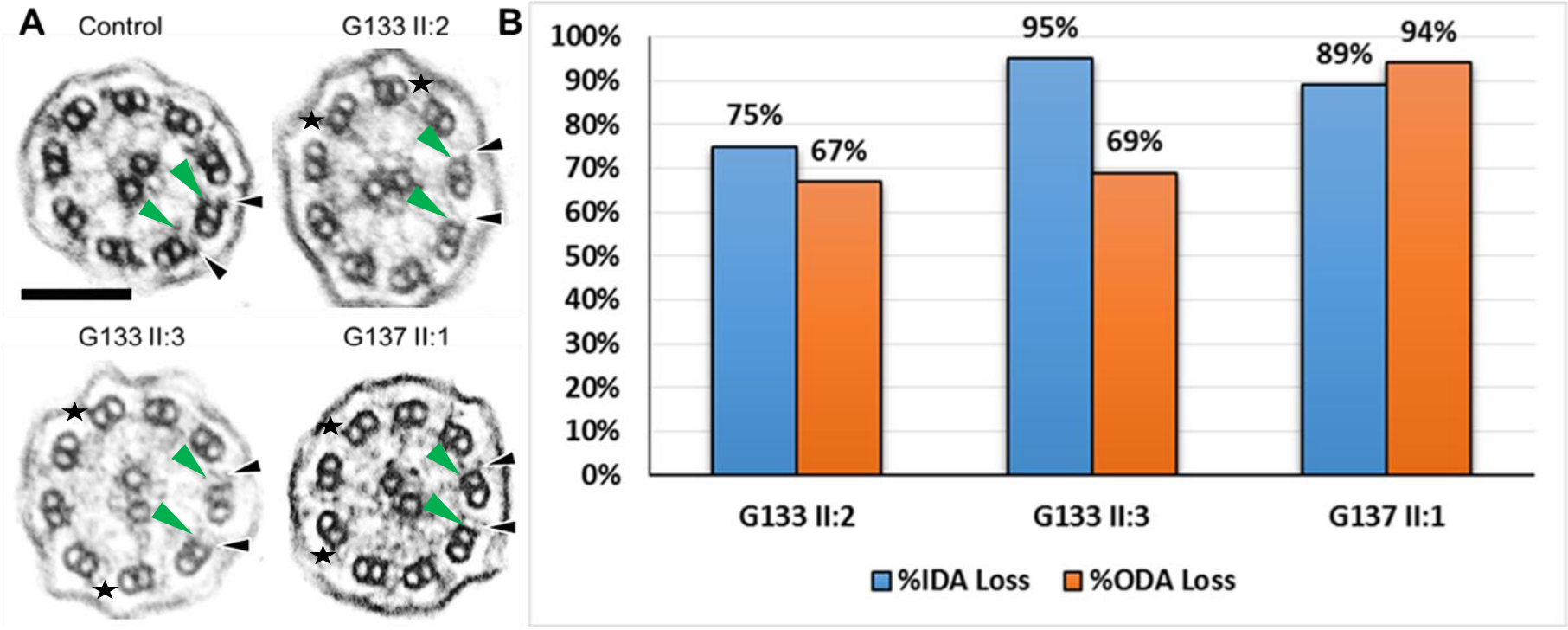


Figure 4-1 Representative images of TEM cross sections of C11orf70 patients' respiratory cilia

A) A loss of both the outer (black arrowheads) and inner (green arrowheads) dynein arms from patients of both families was shown compared to a healthy control. Black stars indicate remnants of short outer dynein arms that were observed to be retained in some cross sections. Scale bar, 100 nm. B) Quantitative TEM analysis showing the percentage of loss of IDA and ODA in the cilia cross sections of the affected individuals following a standard diagnostic assessment (typically 50-70 cilia per individual).

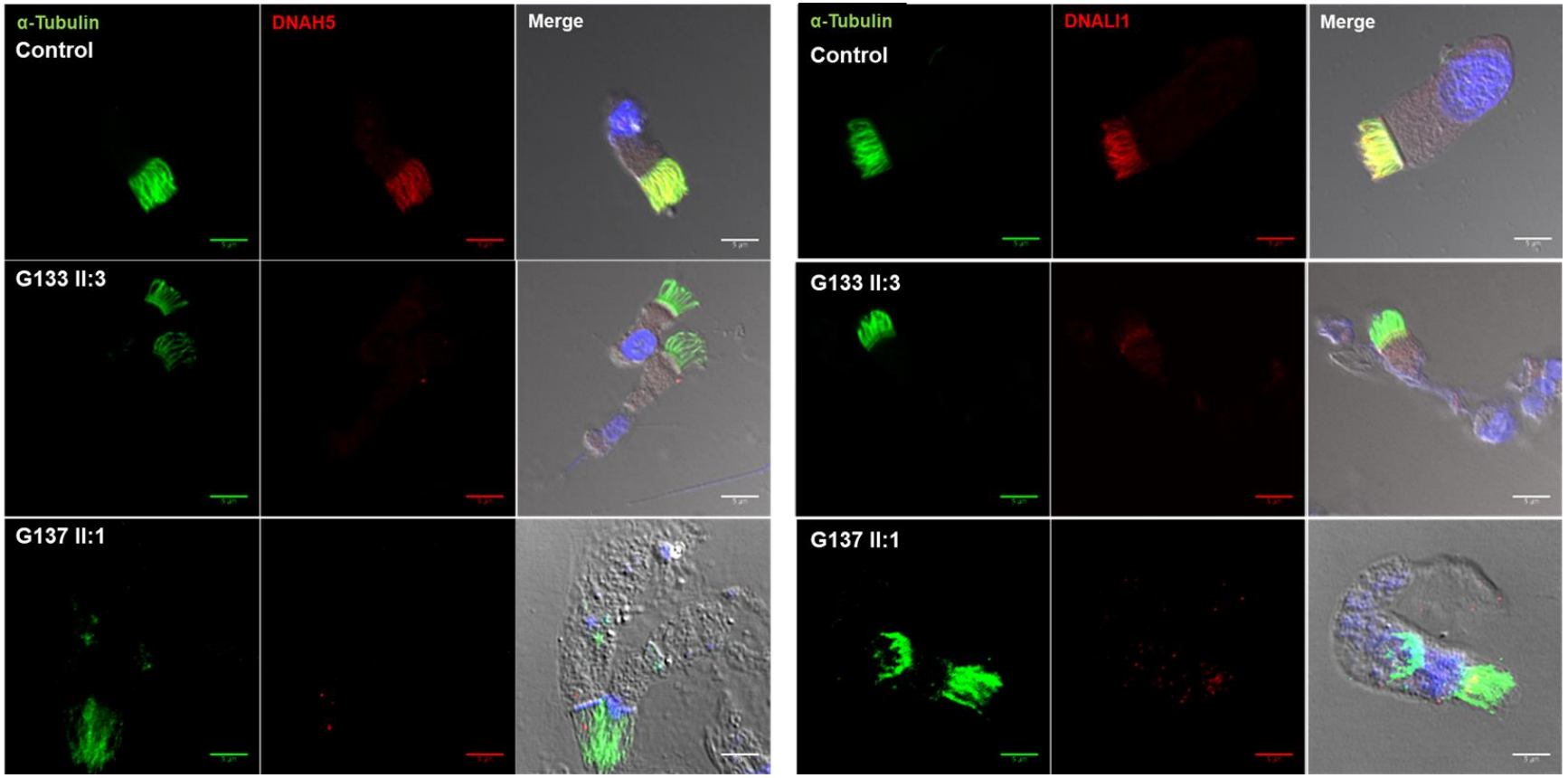


Figure 4-2 Immunofluorescence staining of respiratory epithelial cells of the C11orf70 patients

Left, the outer dynein arm antibody marker DNAH5 (red) shows highly reduced staining for DNAH5 in the respiratory cilia of the affected individuals from both families indicating a loss of the outer dynein arms. Right, antibody staining for the inner dynein arm antibody marker DNALI1 (red) shows a highly reduced staining in the affected individuals from both families for DNALI1 in the cilia, indicating also a loss of the inner dynein arms. Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 5 µm.

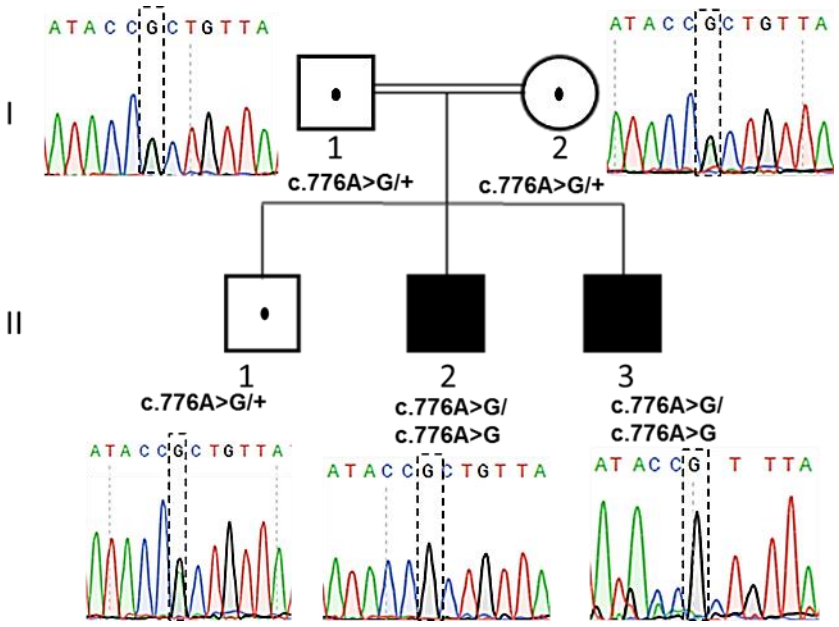


Figure 4-3 Pedigree structure of affected family (PCD-G133)

Electropherogram shows variant regions in the family members; individual I-1 and I-2 (carrier father and mother), II-1 (carrier sibling), II-2 and I-3 (affected siblings). The affected siblings carry a homozygous missense mutation c.776A>G (p.His259Arg).

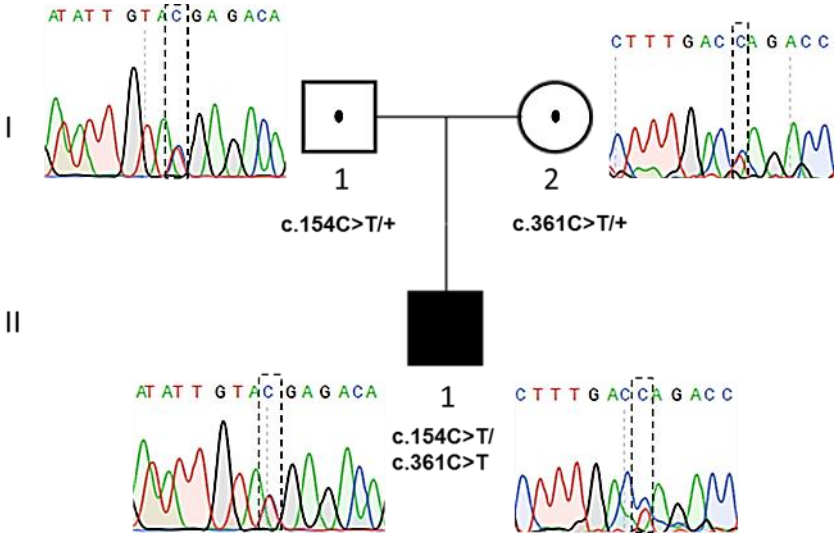


Figure 4-4 Pedigree structure of affected family (PCD-G137)

Electropherogram shows variant regions in the family members; individual I-1 and I-2 (carrier father and mother), II-1 (affected child). The affected child carries compound heterozygous nonsense mutations c.154C>T (p.Gln52*) and c.361C>T (p.Arg121*).

4.2.2 *C11orf70* expression is enriched with multiciliogenesis

To explore the contribution of *C11orf70* to cilia motility, its transcriptional profile was analysed at different time points across 30 days in vitro, using cultures of normal human bronchial epithelial cells grown at an air-liquid interface (ALI) (see **Section 1.4.2**). Real time quantitative PCR of mRNA extracted from ALI cultured cells from a healthy control was used to compare *C11orf70* expression to the expression profile of the known PCD gene, *DNAH5* at the same time points of multiciliogenesis. The expression of *C11orf70* has a similar profile to that of *DNAH5* expression that started to increase around day 10, rising sharply around day 19, followed by a plateau phase around day 24 (**Figure 4-5**). In the GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2), the expression of *C11orf70* is highly enriched in multiciliated tissues as shown in (**Figure 4-6**).(306)

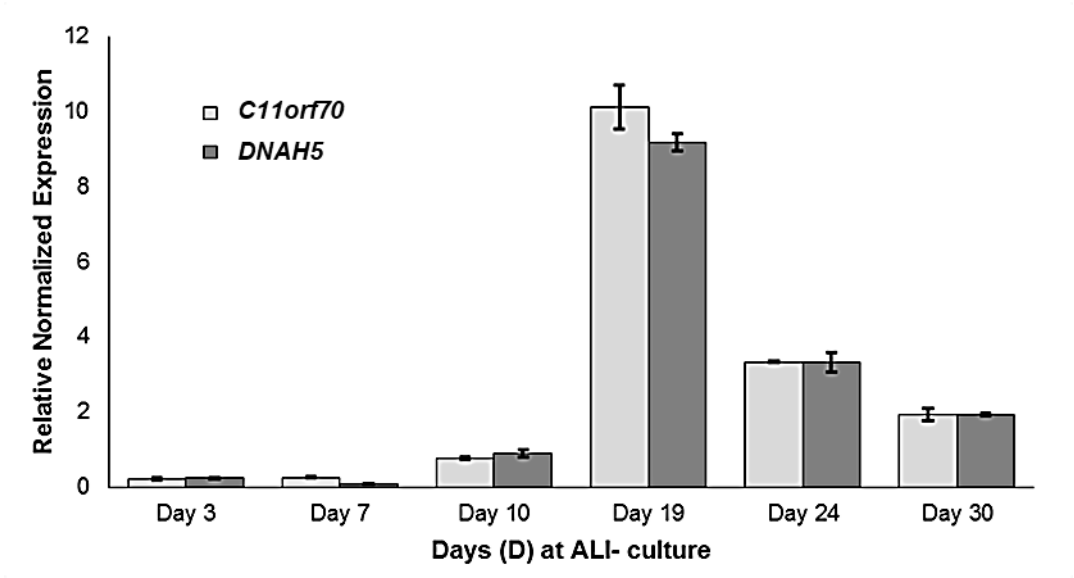


Figure 4-5 Transcriptional profile of C11orf70 during multiciliogenesis

C11orf70 expression was analysed by qPCR during multiciliogenesis in ALI cultures from a control individual, over 30 days in culture compared to the expression of *DNAH5*. The relative expression of both *C11orf70* and *DNAH5* genes were normalized to the expression of *GAPDH* as a housekeeping gene. Both genes showed a similar transcriptional profile over time in ALI culture, increasing around day 10 with a peak at day 19 followed by a plateau from day 24. Error bars indicate SEM (Standard error of the mean) each PCR reaction was done in technical triplicates.

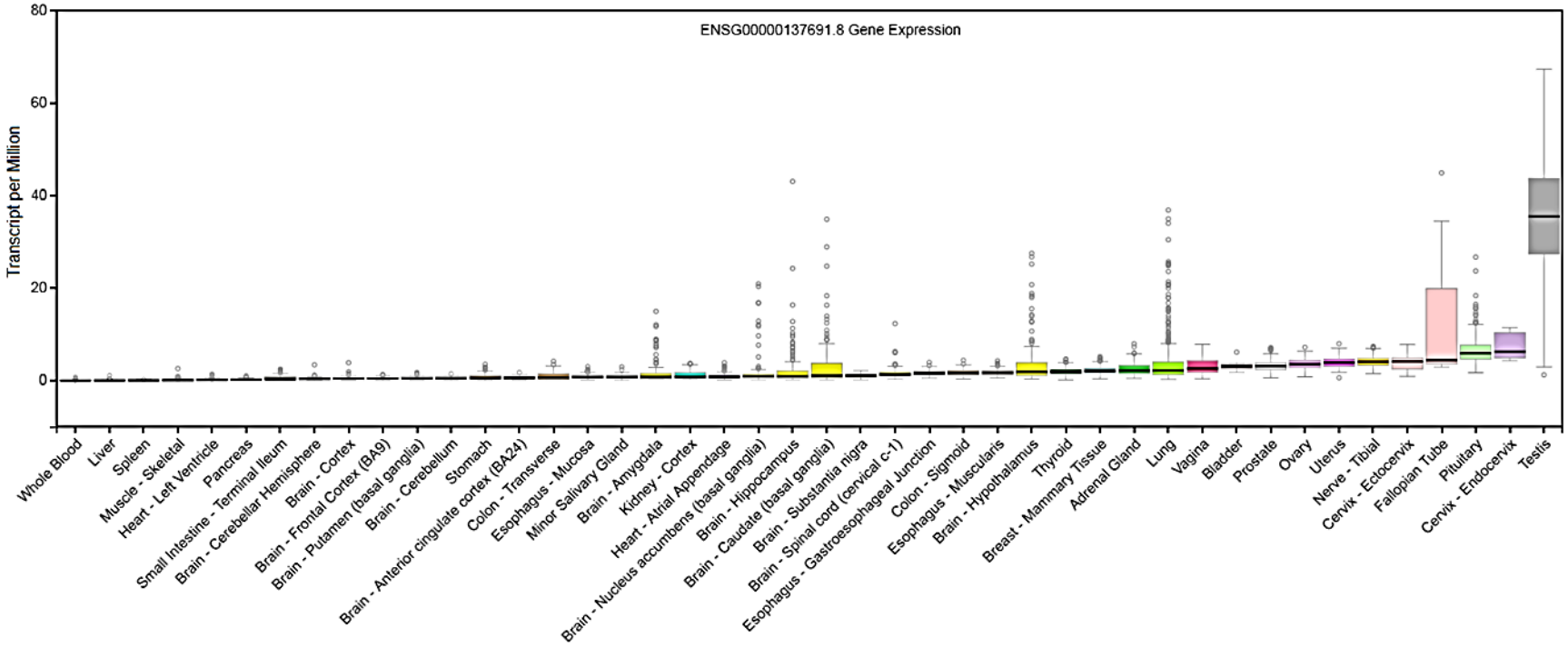


Figure 4-6 Snapshot of the expression profile of C11orf70 gene in GTEx portal

The expression of *C11orf70* is highly enriched in tissues with multiciliated cells mainly testes. This expression pattern is similar to other genes essential for cilia motility. (Data obtained from GTEx portal)(306)

4.2.3 *C11orf70* is highly conserved across species that have motile cilia and IFT-dependent dynein assembly

The phylogenetic distribution of the *C11orf70* orthologs across multiple species was analysed by Prof David Mitchell (SUNY Upstate Medical University in Syracuse, New York) comparing it to the presence or absence of motile cilia in each species, axonemal dynein arms and IFT-dependent dynein assembly proteins. This showed that *C11orf70* orthologs are retained in species with motile cilia with few exceptions. *C11orf70* orthologs were not found in the genomes of species lacking outer dynein arms such as nematodes and lower plants (e.g., *Caenorhabditis, Selaginella,* and *Physcomitrella*). They were also lost from the genomes of organisms that have outer dynein arms but that do not use IFT for dynein assembly/transport as they are lacking ODA16 as an IFT-interacting dynein assembly protein (e.g., *Plasmodium, Toxoplasma,* and *Thalassiosira*). In *Drosophila melanogaster*, *C11orf70* ortholog was also lost despite it retains dynein arms and IFT. (**Figure 4-7**)

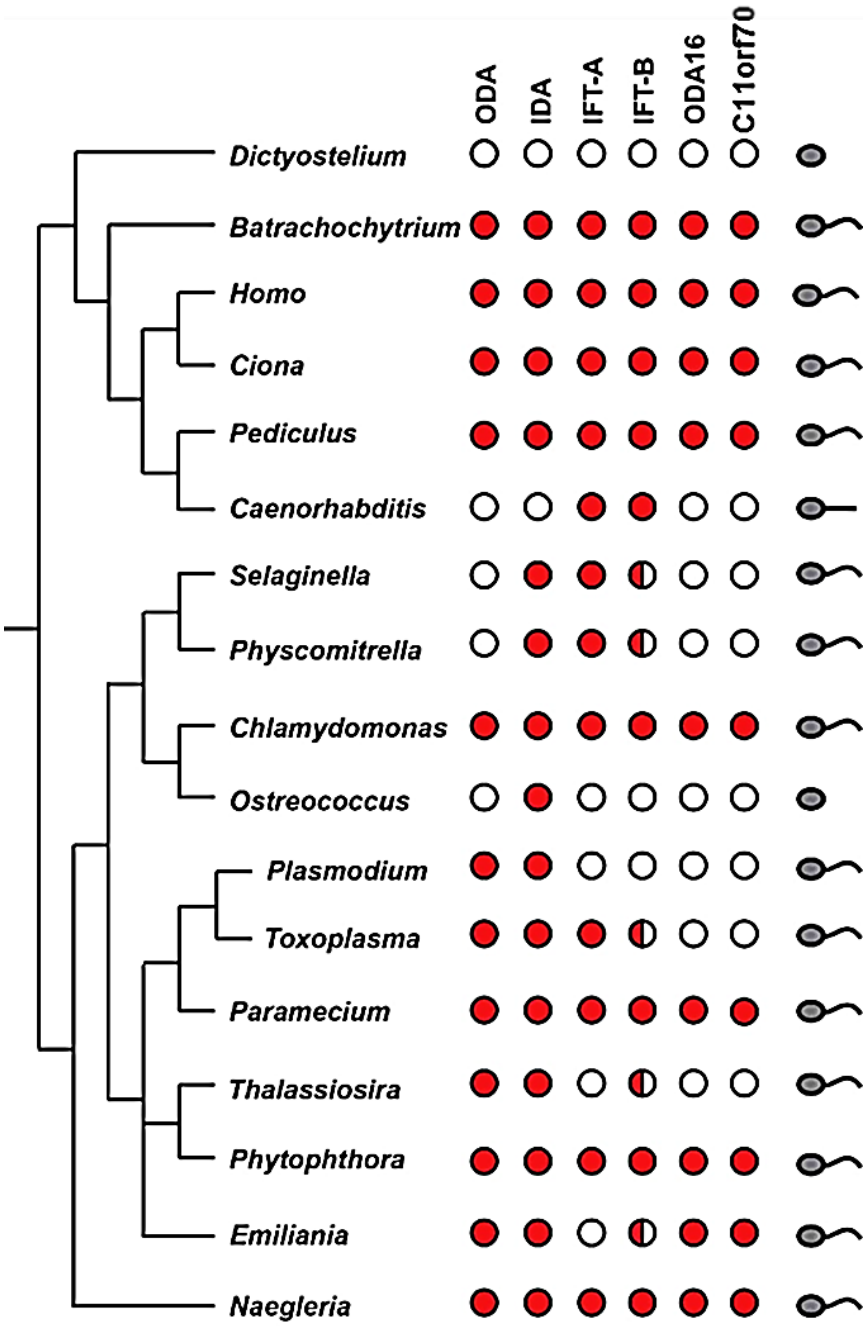


Figure 4-7 Phylogenetic analysis of C11orf70 orthologs distribution compared to the presence of axonemal dyneins and dynein trafficking proteins

Red filled circles indicate system or protein present; white empty circles indicate absent; half-filled circles indicate partially retained. A single *C11orf70* copy is retained in organisms with motile cilia (last column, curved organelle) that use outer dynein arms (ODA column) in the presence of IFT-B complexes and the IFT-interacting dynein assembly factor; ODA16. *C11orf70* orthologs are absent when IFT-B complexes are also completely missing or compromised by missing subunits (half-filled circles). (Data obtained from Prof. David Mitchell, SUNY UMU, USA).

4.2.4 *C11orf70* knockdown in *Paramecium* disrupts axonemal dynein assembly

To study the role of the *C11orf70* gene in cilia motility and its involvement in causing PCD, silencing of the *C11orf70* ortholog (*GSPATT00011350001*) in *Paramecium* was performed followed by assessment of the resulting phenotype.

In **Figure 4-8**, the C11orf70 protein multiple sequence alignment shows a high degree of conservation across species from human down to *Paramecium*. The silencing experiment was done in *Paramecium* following a well-established protocol of RNAi knockdown (KD) by feeding.(102) The level of knockdown was assessed by qPCR showing an 80% reduction in the expression of the *Paramecium C11orf70* ortholog in a KD-*C11orf70* strain compared to a KD-*ND7* strain which is typically used as a control for knockdown experiments, used in a way similar to an empty vector control (since *ND7* gene has no role in cilia motility).(275) (**Figure 4-9**)

RNAi knockdown of the *C11orf70* ortholog in *Paramecium* resulted in an evident phenotypic defect secondary to defective cilia motility. This was found not to be a consequence of any gross change in cell size, cilia length or number per cell, assessed visually. (**Figure 4-10**) As a consequence of the gene abrogation, there was severe reduction (p = 6.19E-11) to about 68% in the swimming velocity of *Paramecium* in the KD-*C11orf70* (0.58 ± 0.22 mm/s) compared to that of the control KD-*ND7 Paramecium* cells (0.85 ± 0.20 mm/s) after 72 hr of RNAi feeding. (**Figure 4-11**)

HSVM analysis of the *Paramecium* cilia beating pattern and frequency was performed. A reduction ($p < 0.05$) in cilia beating frequency (CBF) was noted in KD-*C11orf70* cells (20.3 ± 4 Hz) compared to that of the control KD-*ND7* cells (24.3 ± 2 Hz). (**Figure 4-12**) Despite the modest decrease in overall CBF recorded to arise from reduced *C11orf70* expression, in the movies of the KD-*C11orf70* strain, it was clear by eye that some cilia were completely immotile. There was also a variable beating pattern, ranging from a mild reduction in the amplitude of the cilia waveform to a half beat.

TEM analysis of the *Paramecium* cilia cross sections revealed combined loss of IDA and ODA which was consistent with the ultrastructural phenotype seen in patients carrying *C11orf70* mutations. Quantitative assessment of IDA and ODA loss revealed statistically significant reduction of both IDAs and ODAs from the cilia cross sections in KD-*C11orf70* cells compared to KD-*ND7* cells. The numbers of IDAs and ODAs that were present per cilia cross section in KD-*C11orf70* cells (IDA, 4.9 ± 1.8) and (ODA, 7.2 ± 2.2) were significantly reduced (IDA, p = 2.17E-8 & ODA, p = 0.0019) compared to the control KD-*ND7* strain (IDA, 7.8 ± 1.1) and (ODA, 7.9 ± 1.3). The modest reduction of ODA loss, despite being statistically significant, may indicate a less essential role for C11orf70 in ODA assembly compared to its more profound role in IDA assembly. (**Figure 4-13**)

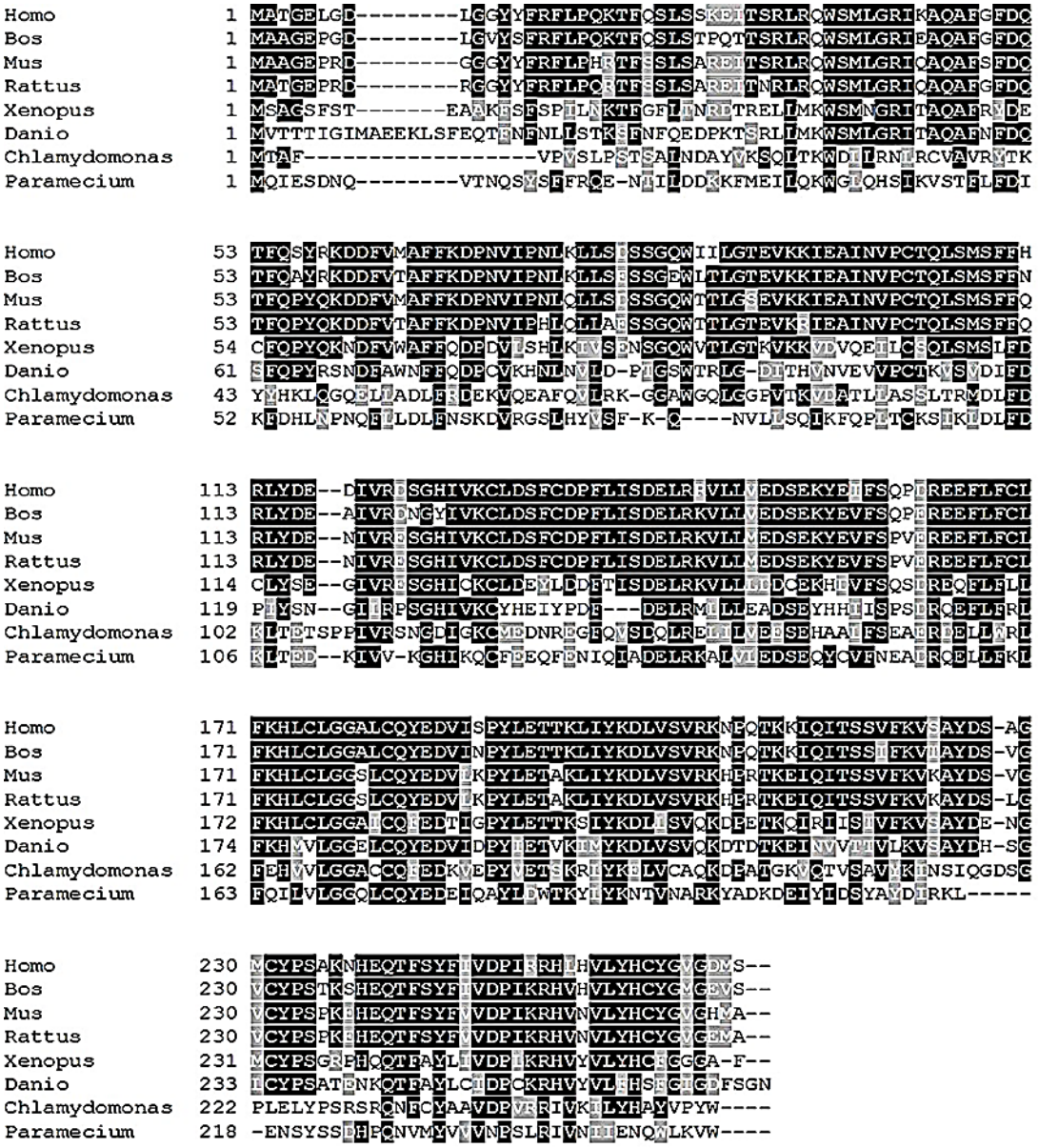


Figure 4-8 Cross-species protein alignment for C11orf70

C11orf70 is a highly conserved protein across species. There is only one copy of the human *C11orf70* ortholog in the *Paramecium* genome despite the *Paramecium tetraurelia* has undergone at least three whole-genome duplications.

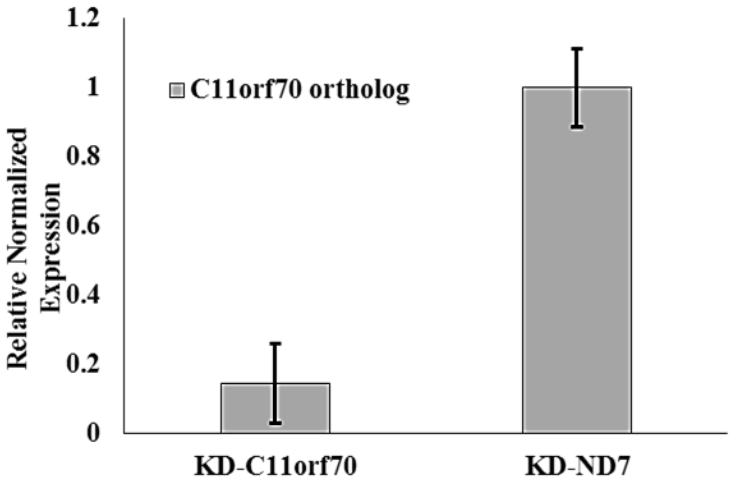


Figure 4-9 Real-Time PCR assessment of the level of C11orf70 knockdown in Paramecium

There was a significant reduction in the *C11orf70* ortholog mRNA level of about 80% after knockdown in KD-*C11orf70* compared to KD-*ND7 Paramecium*. All error bars indicate SEM.

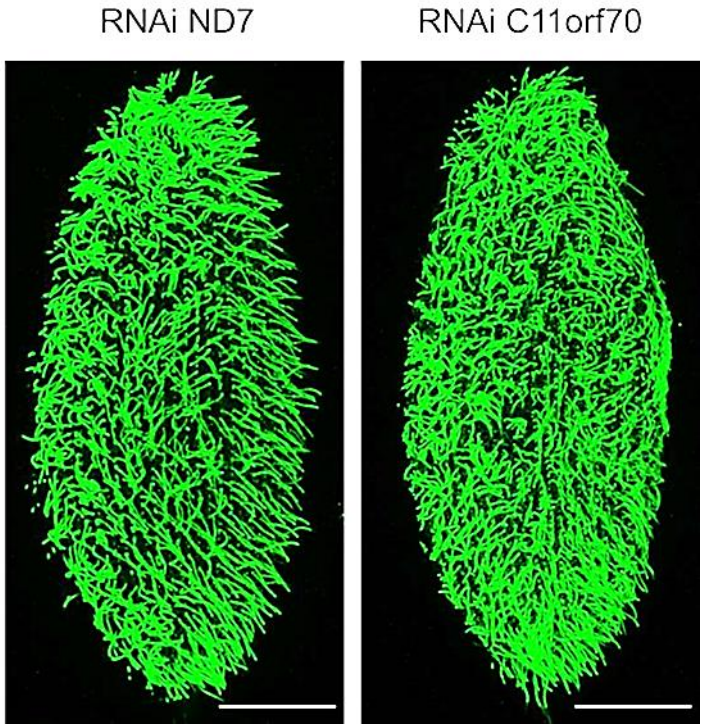


Figure 4-10 RNAi knockdown experiment in Paramecium did not grossly affect the cell size or cilia morphology or numbers

Silencing of *C11orf70* does not affect cilia number or length when compared to control *ND7* silencing. Anti -polyglutamylated tubulin antibodies (PolyE) were used to stain the cilia. Scale bars, 20 µm.

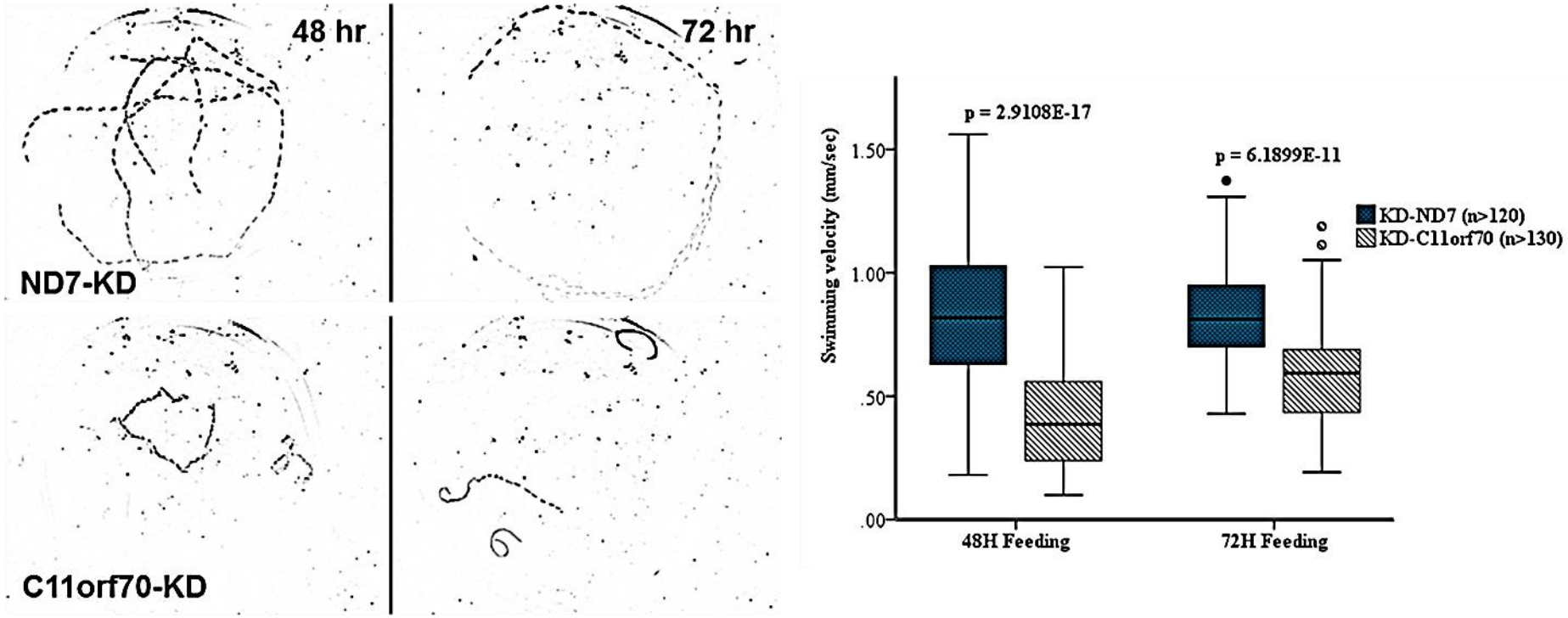


Figure 4-11 Paramecium swimming velocity assessed by dark-field microscopy

Z-projection of track recordings was captured after 48 and 72 hr of RNAi feeding. Wider spaced dots indicate fast moving *Paramecium*, dots tightly close indicate slower moving or immobile *Paramecium*. KD-*C11orf70* cells (bottom) show a severe motility phenotype compared to the control KD-*ND7* cells (top) at 48 hrs and even more strikingly at 72 hrs. Statistical analysis of the swimming velocity of *C11orf70* and *ND7* knockdown cells was done by independent sample t test showing significant reduction in the swimming velocity after *C11orf70* silencing compared to control *ND7* silencing. n= number of *Paramecium* cells analysed per condition.

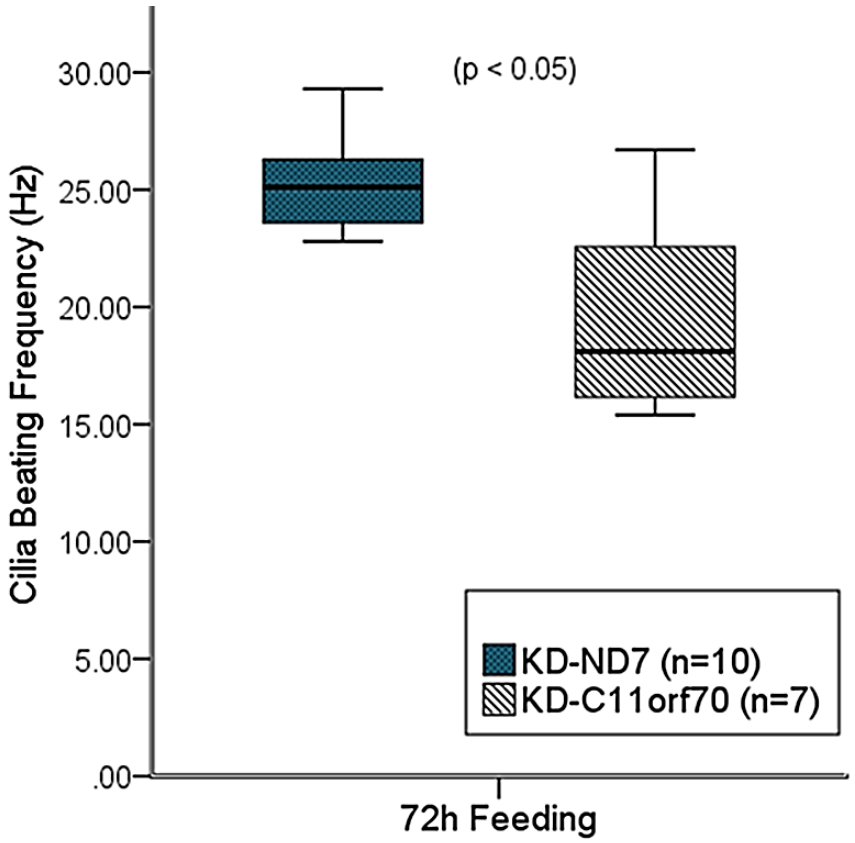


Figure 4-12 Cilia beating frequency quantified by HSVM

After 72 hr of RNAi feeding, the ciliary beat frequency of KD-*C11orf70* cells (20.3 ± 4 Hz) was reduced by about 20% ($p < 0.05$) compared to KD-*ND7* control cells (24.3 ± 2 Hz). 7–10 *Paramecium* cells were evaluated per condition; *p* value was measured by Mann-Whitney U test. n= number of *Paramecium* cells analysed per condition.

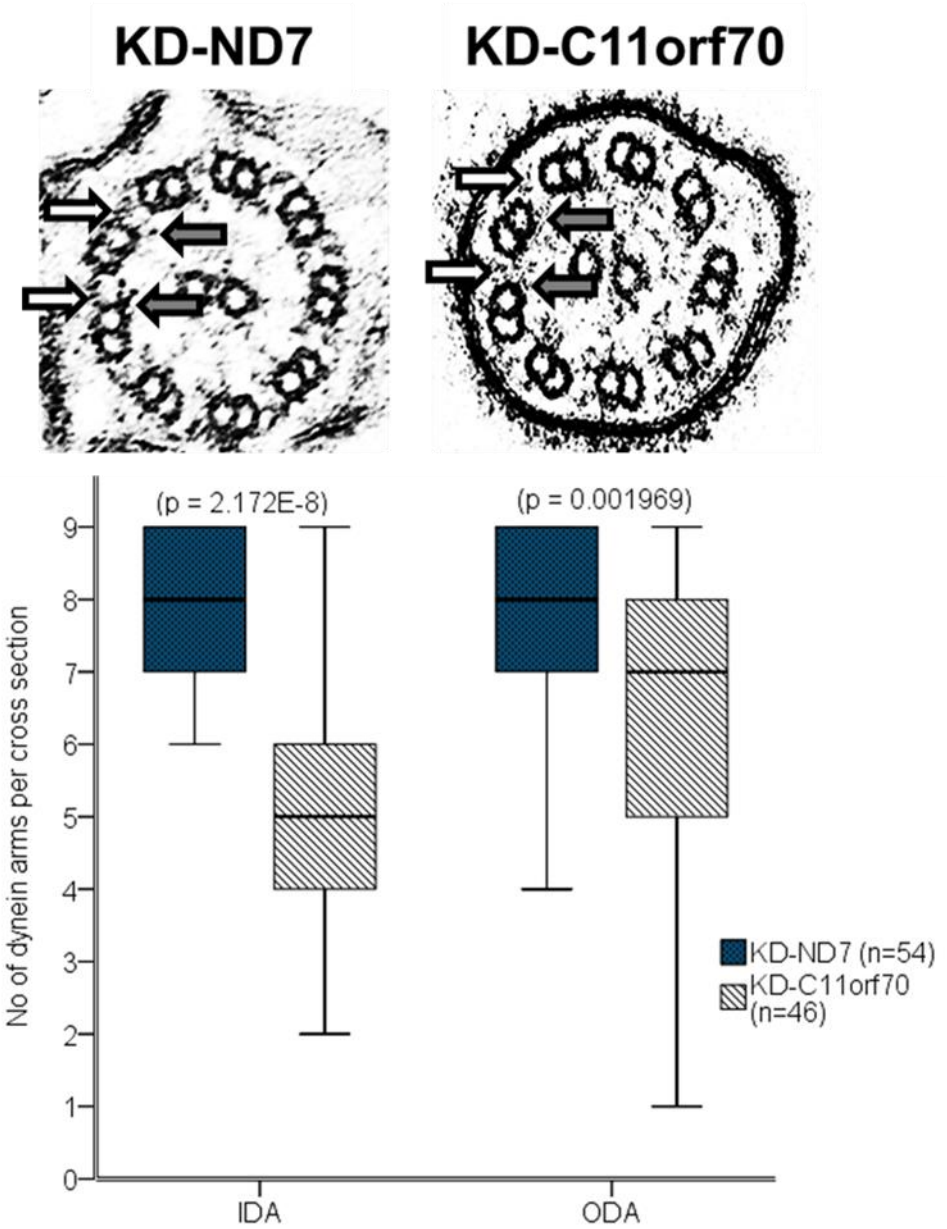


Figure 4-13 TEM analysis of Paramecium cilia cross-section

Absence of the outer (white arrows) and inner (grey arrows) dynein arms in the KD-*C11orf70* cells (right) compared to the control KD-*ND7* cells (left). Quantification of dynein arms showed significant reduction of numbers of ODA and IDA in KD-*C11orf70* cells compared to KD-*ND7* control cells. *p* value was measured by independent samples t test. n= number of cilia cross sections analysed per condition.

4.2.5 C11orf70 has a distinct intracellular distribution pattern similar to IFT-associated proteins

The C11orf70 protein is consistently present in species where motile cilia, IFT complexes and IFT-interacting dynein assembly factors are present as previously shown in **Figure 4-7**. To further study the intracellular distribution of C11orf70 and how it can interact with the IFT system to play its role in axonemal dynein assembly, the behaviour of the *Paramecium* C11orf70 protein was assessed in the context of both anterograde (IFT-B related) and retrograde (IFT-A related) protein trafficking in cilia. First, the effect of knockdown of one of the members of the retrograde IFT-A complex (*IFT139*) on C11orf70 was assessed. Injection of GFP-tagged *Paramecium* C11orf70 into the *Paramecium* macronucleus was performed by Dr Anne-Marie Tassin and Pierrick Le Borgne (*Paramecium* Lab, CNRS). Expression of GFP-tagged C11orf70 generated several transformant clones with variable levels of C11orf70-GFP expression but with a normal growth phenotype. The results were the same in all different clones, with C11orf70 predominantly present in the cytoplasm with a small amount present in the cilia (**Figure 4-14A**). Knockdown of the *Paramecium* ortholog of *IFT139* by RNAi feeding in C11orf70-GFP transformed *Paramecium* resulted in a strikingly abnormal accumulation of C11orf70 protein at the tips of the cilia, suggesting a transport defect and indicating that IFT-A trafficking is involved in the transport of C11orf70 within the cilia (**Figure 4-14 A**).

The behaviour of C11orf70 during cilia regrowth in *Paramecium* was assessed in comparison to the behaviour of IFT46 (a member of the anterograde IFT-B complex which is essential for the transport of ODA). *Paramecium* transformed with C11orf70-GFP, IFT46-GFP and GFP alone were de-ciliated and then assessed after 15 mins of cilia re-growth, as the IFT transport system is active at this time. Whilst IFT46 accumulates in the tips of growing cilia, C11orf70 is present in both cilia and cytoplasm, with slight accumulation at the tips of the cilia, though no strong difference compared to GFP alone. (**Figure 4-14B**) This suggested that C11orf70 could have a role in IFT-based dynein transport, although further experiments would be required

to determine whether it is actively transported into the cilia and then recycled back to the cell body'.

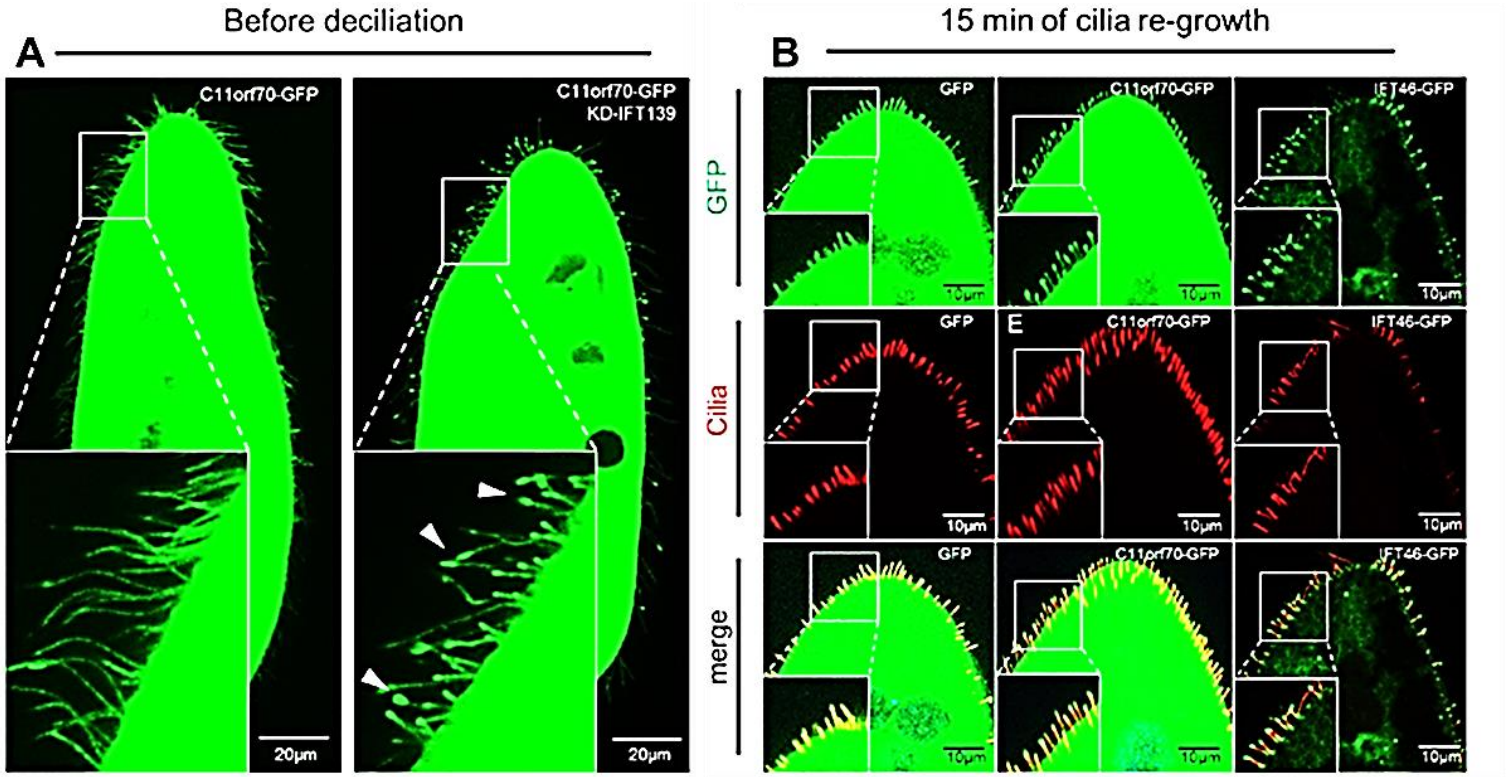


Figure 4-14 C11orf70 is actively transported in cilia in a manner dependent on the IFT transport system

(A) C11orf70-GFP transformed *Paramecium* cells show a high level of C11orf70 in the cytoplasm with small amounts are present in the cilia (left panel). Knockdown of *IFT139* results in defective IFT transport shown by the aberrant accumulation of C11orf70 at the tips of the cilia (right panel). Scale bars, 20 μm. (B) The behaviour of C11orf70 during cilia re-growth was shown by direct GFP fluorescence (green, top) and monoclonal antibody staining against polyglutamylated tubulin (cilia marker, red, middle; merge, bottom). Scale bars, 10 μm.

4.3 Discussion

Targeted next generation sequencing, using a multi-gene panel designed to include all known PCD genes and other potential candidate genes involved in cilia motility, was used for genetic screening of PCD patients. Mutations were found in the previously uncharacterized *C11orf70* gene, in 3 patients from two unrelated families. All affected individuals had a clinical presentation consistent with PCD and situs inversus was reported in both families. PCD diagnosis was confirmed in these patients by various clinical diagnostic testing results before attempting genetic screening.

The affected individuals had immotile cilia demonstrated by HSVM analysis of cilia of the patients' respiratory epithelium obtained by nasal brushing biopsies. Patients' cilia showed a combined loss of both inner and outer dynein arms from the majority of cilia cross sections by TEM analysis. This highlighted the potential power of using a targeted NGS approach with a strong list of biologically relevant candidate genes for novel gene discovery linked to PCD. Interestingly, Inga Hoben et al (307) also published a report of loss of function mutations in *C11orf70* identified in PCD patients from 5 unrelated families, at the same time as this study. These patients have the same phenotypic characteristics of patients identified in the current study. They showed complete immotility of the respiratory cilia and sperm flagella and also loss of both inner and outer dynein arms from the cilia axoneme.

As the ciliary axonemal structure is highly conserved across species, various model organisms have been used to study cilia motility and functionally characterize potential candidate disease genes essential for cilia motility that might be linked to human PCD.(97, 112, 232, 233, 237) C11orf70 was shown to be a highly conserved protein across anciently divergent species. *C11orf70* orthologs are retained today only in the genomes of organisms with motile cilia/flagella and lost in organisms that do not have an IFT-associated axonemal dynein assembly process. Phylogenetic analysis of C11orf70 distribution across species showed a distinct distribution pattern across different species, consistently linked to the presence of IFT-associated

dynein assembly factors like ODA16 (WDR69).(96, 296, 303) These findings support the importance of C11orf70 in axonemal dynein assembly and IFT-B-based transport into the motile cilia. Moreover, Hoben et al have shown direct interactions between C11orf70 and DNAAF2 (another dynein assembly factor, mutated in PCD) using a yeast two-hybrid screen, which was also confirmed by co-immunoprecipitation (co-IP).(307)

Previous gene expression studies of both mRNA and proteins in various tissues showed that C11orf70 is more enriched in multiciliated tissues than non-ciliated tissues.(308, 309) However, it was not reported to be present in the proteomic profiling of human respiratory cilia.(110) The relatively low abundance of the C11orf70 protein in the ciliary axoneme can explain why it was not detected in the motile cilia proteome since the study only analysed the proteomic profile of the isolated cilia from the cells and C11orf70 distribution is mainly cytoplasmic. Microarray-based transcriptional profiling of mucociliary differentiation has shown upregulation of *C11orf70* expression.(108) These findings are consistent with what was shown here that the *C11orf70* expression pattern is similar to that of other PCD genes during ciliogenesis, with a peak around the time of cilia emergence followed by a plateau when the cilia become more mature. On the other hand, the *Paramecium* C11orf70 ortholog, GSPATP00011350001, was also not detected within the cilia proteome (310) or during cilia regrowth after deciliation of *Paramecium*.(311) By *in situ* hybridization analysis of 8.25 dpc (days post-coitum) mouse embryos, *C11orf70* ortholog (*9230110C19Rik*) was expressed in the left/right organizer, the only site of motile cilia at this stage. By RNA-Seq, higher expression of *C11orf70* was detected in ALI- cultured human bronchial epithelial cells and in nasal brushing biopsies compared to its expression in blood or immortalized lymphocytes.(307)

For functional characterization of *C11orf70* and to study its role in cilia motility, its ortholog was silenced in *Paramecium*, followed by assessment of the phenotypic consequences of its knockdown on cilia motility and structure. At first, the level of knockdown was assessed at the RNA level and then the cilia phenotype was evaluated. *C11orf70* RNAi silencing resulted in a significant decrease in cilia beating frequency leading to a reduction in

Paramecium swimming velocity. Ciliary ultrastructural analysis showed a partial loss of IDA and small but significant reduction in ODA, similar to ultrastructural changes observed in patients' cilia. This indicates the conserved role C11orf70 plays in the process of dynein arm assembly/transport.

Expression studies in *Paramecium* showed that C11orf70 has a mainly cytoplasmic distribution with a small fraction present in the cilia. Slight axonemal enrichment of C11orf70 after knockdown of retrograde *IFT139* and an expression in the elongating axonemal tips during regeneration marked by anterograde IFT46, indicates the potential role of C11orf70 in the IFT-associated dynein assembly process.

In a supporting parallel study of the function of C11orf70 in *Chlamydomonas* done by Prof. David Mitchell (SUNY Upstate Medical University in Syracuse, New York), a single well-conserved C11orf70 ortholog was identified, FBB5. It was found that FBB5 was upregulated after deflagellation though it was absent from any published ciliary/flagellar proteomes.(312-314) The *Chlamydomonas* FBB5 was shown to be predominantly present in the cytoplasm with only a tiny amount present in the flagellar fraction in a pattern similar to other IFT-associated dynein assembly and transport factors e.g. ODA16 and ODA8.(294, 296, 315) The flagellar portion of FBB5 was found to be reduced in the *oda16* mutant strain which lacks the IFT-associated dynein transport factor ODA16.(96, 296, 304) The expression and localization of FBB5 was however not affected by an N-terminus deletion of IFT46 which is essential for the interaction between IFT46 and ODA16 for dynein transport process indicating that the N-terminus of IFT46 is not essential for the transport of FBB5 in the cilia.(296, 303, 304)

4.4 Summary

In this chapter, I have reported the use of a custom ‘motile ciliome’ gene panel which was designed to include a good list of candidate genes for novel gene discovery in PCD. I have identified disease-causing mutations in a previously uncharacterized gene, *C11orf70* in PCD patients. These mutations cause loss of both IDAs and ODAs and cilia immotility leading to PCD.

I have shown the outcome of using a new model organism, *Paramecium* for functional characterization of this gene and to study its role in cilia motility. Knockdown of the *C11orf70* ortholog in *Paramecium* reproduced the ciliary ultrastructural defects seen in the patients. This added to the current understanding of the dynein assembly and the IFT cargo transport process, expanding our understanding of the genetic background of PCD in humans.

Chapter 5 Mutations in *DNAH9* cause a distinct PCD phenotype

5.1 Introduction

The motile ciliary axoneme is the core structure of cilia, comprising a circle of nine peripheral doublets of microtubules and a central pair of single microtubules termed the '9+2' array. Most species have motile cilia of the embryonic left-right organiser, lack the central pair of microtubules and have a '9+0' pattern.(124, 316) Dynein motors comprise the inner "IDA" and outer "ODA" arms. They are lateral extensions attached to the peripheral microtubules and are responsible for cilia beating through ATP hydrolysis.(317)

In mammalian species, each complex outer dynein motor structure is attached to the microtubule by a docking complex and is composed of multiple heavy, light and intermediate dynein chains.(318, 319) These dynein chains comprise multiple dynein isoforms produced by different genes, rather than resulting from alternative splicing of one or a limited number of genes. The ODA structure is highly evolutionarily conserved and it repeats every 24 nm along the length of the axoneme. This spacing is controlled through the nexin dynein regulatory complex and radial spokes to produce a regular ciliary waveform.(320, 321)

There are two types of ODAs described in human cilia. Type 1 ODAs are localized to the proximal part of the cilium nearer to the cell body, in the region of the microvilli. They contain the two dynein heavy chains (HC) DNAH5 (orthologous to the ODA γ-HC found in *Chlamydomonas*) and DNAH11 (orthologous to the *Chlamydomonas* ODA β-HC). Type 2 ODAs are located more distally in the cilium and contain the two heavy chains DNAH5 and DNAH9 (also orthologous to *Chlamydomonas* ODA β-HC).(162, 322) This regional distribution of DNAH9, restricted to type 2 ODAs of the distal part of the human motile cilium was shown by immunofluorescence studies in normal ciliated epithelial cells. In contrast, DNAH9 was shown to be present along the whole length of the sperm flagellum (11), although expression databases suggest that the paralogue DNAH17 is likely a testes-specific cross reactive isoform responsible for this pattern.(17)

DNAH9 was the first dynein heavy chain gene to be studied in human, with determination of its genomic structure and protein domains.(323) Mutational analysis of 31 families affected with PCD did not reveal any disease causing variants in *DNAH9*, but this was an early study in 2001, and did not exclude *DNAH9* as a potential candidate for PCD.(323) Disease causing mutations in the other structural dynein proteins; DNAH5, DNAH11, DNAI1, DNAI2 and DNAL1 have been identified in PCD patients. Mutations in *DNAH5, DNAI1* and *DNAH11* are amongst the commonest causes of PCD.(80, 161, 324)

DNAH5 mutations result in loss of the γ-HC DNAH5 and subsequently loss of the β-HC DNAH9 leading consequently to a classic ODA-loss ultrastructure phenotype when assessed by TEM and PCD. Localization of DNAH9 remains unchanged in the sperm flagellum of patients with homozygous causal mutations in *DNAH5 as* DNAH5 is not present in the sperm flagellum and consequently DNAH9 is not affected.(11) DNAH9 is also lost from the ciliary axoneme of a patient with *DNAI1* truncating mutations.(11)

Mutations in *DNAH11* lead to an apparently normal ultrastructure by conventional TEM with the apparently undisturbed presence of ODAs despite the absence of DNAH11.(161) There is in fact a subtle structural ODA defect in cases with *DNAH11* mutations, which can be identified by higher resolution ultrastructural analysis (electron tomography). These small DNAH11 defects were found to affect only the proximal ODAs, causing a 25% reduction of ODA volume visible as a small hole in the ODAs.(163) It was also shown in DNAH11-deficient ciliated cells, by immunofluorescence, that DNAH9 can translocate to reside in the place of DNAH11 in the proximal part of the ciliary axoneme.(162)

Here, I identified mutations in *DNAH9* in two unrelated families with PCD. The patients respiratory cilia showed distinctive ultrastructural and motility defects. Knockdown of *DNAH9* orthologs in *Paramecium* confirmed the highly conserved role of DNAH9 in cilia motility as a structural component of ODAs.

5.2 Results

5.2.1 Mutations in *DNAH9* identified in patients with PCD symptoms

Mutations were identified in *DNAH9* in three affected individuals from two unrelated families. The clinical picture and the results of the PCD diagnostic testing are summarized in **Table 5-1**. The first family (G131) is a UK based Somali family with non-consanguineous parents. There were two affected siblings (male and female) who manifested with rhino-sinusitis and situs inversus with no history of neonatal respiratory distress, otitis media or bronchiectasis. Their nasal NO and cilia beating frequency were close to the normal range. They also had lung function within the normal range (FEV1%; Forced Expiratory Volume in one second % predicted in healthy children based on their age, height, gender and ethnicity). The second family (G141) was Turkish in origin with consanguineous parents. The only male child had rhino-sinusitis, cough, situs inversus and complex congenital heart problems including an unbalanced atrioventricular septal defect with congenitally corrected transposition of the great arteries and inferior vena cava conduit stenosis that required stenting. There was no history of neonatal respiratory distress, otitis media or bronchiectasis. His nasal NO level and cilia beating frequency were found to be within normal range. The cilia beating of all three affected children showed a distinct pattern with a subtle defect in the bend of the distal portion of the cilia.

Using the targeted next generation sequencing gene panel to screen patients with PCD phenotype, compound heterozygous variants were identified in both siblings of family (G131): a missense variant (**c.10193G>T, p.Arg3398Leu**) and a consensus splice acceptor site variant (**c.8708-2A>G**) at exon 46. Both mutations were confirmed in the affected siblings along with the heterozygous carrier state of the missense variant only in two unaffected siblings, confirming a consistent inheritance pattern in the absence of any available parental samples. (**Figure 5-1**) The missense variant (**c.10193G>T,**

p.Arg3398Leu) which is absent from the ExAc, dbSNP and EVS control databases, is predicted to be probably damaging by Polyphen-2 with a score of 0.997 and it has a CADD score of 34. The splice acceptor variant (**c.8708-2A>G**) is found in dbSNP (rs143007518) and is present once in the ExAc control database in a heterozygous state with a total allele frequency of 3.234e-5. It has a CADD score of 23.3 and is predicted to cause a protein frameshift.

A homozygous missense mutation (**c.12367G>A, p.Asp4123Asn**) was found in the affected child of family (G141). Sanger sequencing confirmed the mutation in the child and showed that the father was a heterozygous carrier. The DNA sample from the mother was not available. (**Figure 5-2**) This variant is not found in the ExAc, dbSNP and EVS control databases. It is predicted as probably damaging using Polyphen-2 with a score of 1.0 and it has a CADD score of 33.

In support of both the identified missense changes pathogenic nature, multiple protein sequence alignment across species (orthologs) and across several dynein heavy chains (paralogs) showed the highly conserved amino acid residues Arg3398 and Asp4123. (**Figure 5-3**) To study the functional consequence of the splice acceptor mutation (**c.8708-2A>G**), RT-PCR of cDNA synthesized after RNA purification from nasal brushing biopsies of one of the affected individuals and a healthy control was performed using primers in exons 41 and 47(the putative mutation affects the exon 46 splice acceptor). There were two PCR bands amplified in the patient sample compared to only one band amplified in the control sample. Sanger sequencing of gel-purified bands confirmed an exon 46 skipping event, predicted to lead to a frameshift of the protein producing a stop codon 53 residues downstream of the mutation [p.(**Glu2904Aspfs*53**)]. Quantitative RT-PCR showed reduced mRNA levels of *DNAH9* in the patient's nasal brushing sample compared to that of a normal control (normalized to *GAPDH* expression level). (**Figure 5-4**)

Table 5-1 Summary of the clinical symptoms and diagnostic findings in patients with DNAH9 mutations

Patient	**G131.II.2**	**G131.II.3**	**G141.II.1**
***DNAH9* mutations NM_001372.3**	c.10193G>T; p.Arg3398Leu + c.8708-2A>G; p.Glu2904Aspfs*53	c.10193G>T; p.Arg3398Leu + c.8708-2A>G; p.Glu2904Aspfs*53	c.12367G>A; p.Asp4123Asn + c.12367G>A; p.Asp4123Asn
NRDS	No	No	No
Rhino-sinusitis	Yes	Yes	Yes
Productive cough	Yes	Yes	Yes
Situs inversus	Yes	Yes	Yes
Otitis media	No	No	No
Bronchiectasis	No	No	No
Nasal NO (normal>77nl/min)	46nl/min	100nl/min	105nl/min
CBF (normal 8-15 Hz)	8-9.3Hz	8.4-8.6Hz	8.4-10.1Hz
Other morbidities	No	No	Protein losing enteropathy and complex congenital heart disease

NDRS; Neonatal Respiratory Distress Syndrome, NO; Nitric Oxide, CBF; Cilia Beating Frequency, Hz; Hertz.

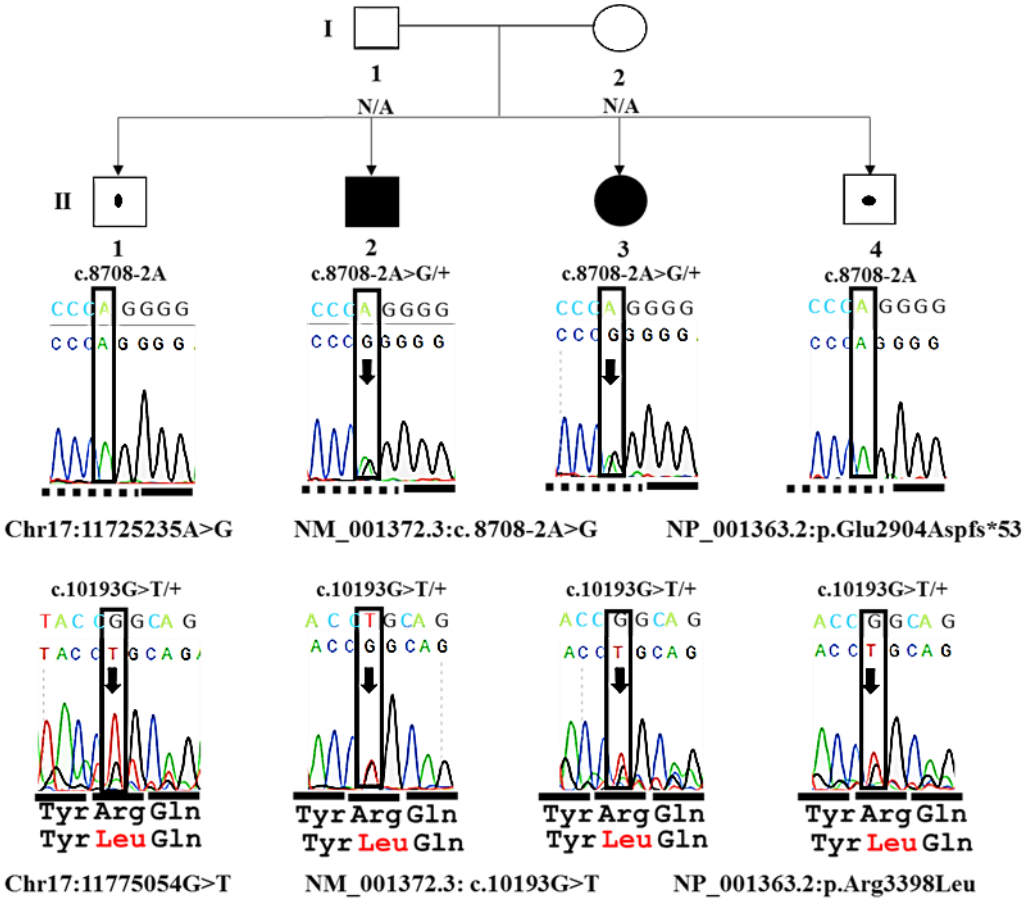


Figure 5-1 Pedigree structure of affected family (PCD-G131)

Sanger electropherograms show variant regions in the family members; II.1 and II.4 are normal siblings, II.2 and II.3 are affected children. The affected siblings carry compound heterozygous splice acceptor (c.8708-2A>G) and missense (c.10193G>T; p.Arg3398Leu) mutations. DNA samples from both parents were not available.

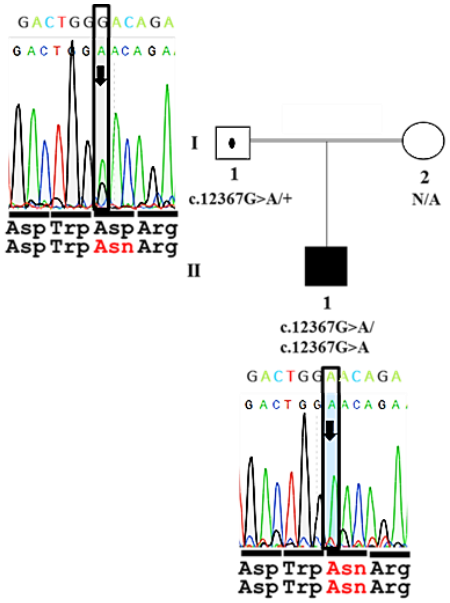


Figure 5-2 Pedigree structure of affected family (PCD-G141)

Sanger electropherograms show variant regions in the family members; II.1 is the only affected child. Sanger sequencing confirmed the homozygous status of a missense mutation (c.12367G>A; p.Asp4123Asn) in the patient and the carrier state of the father. DNA of the mother was not available.

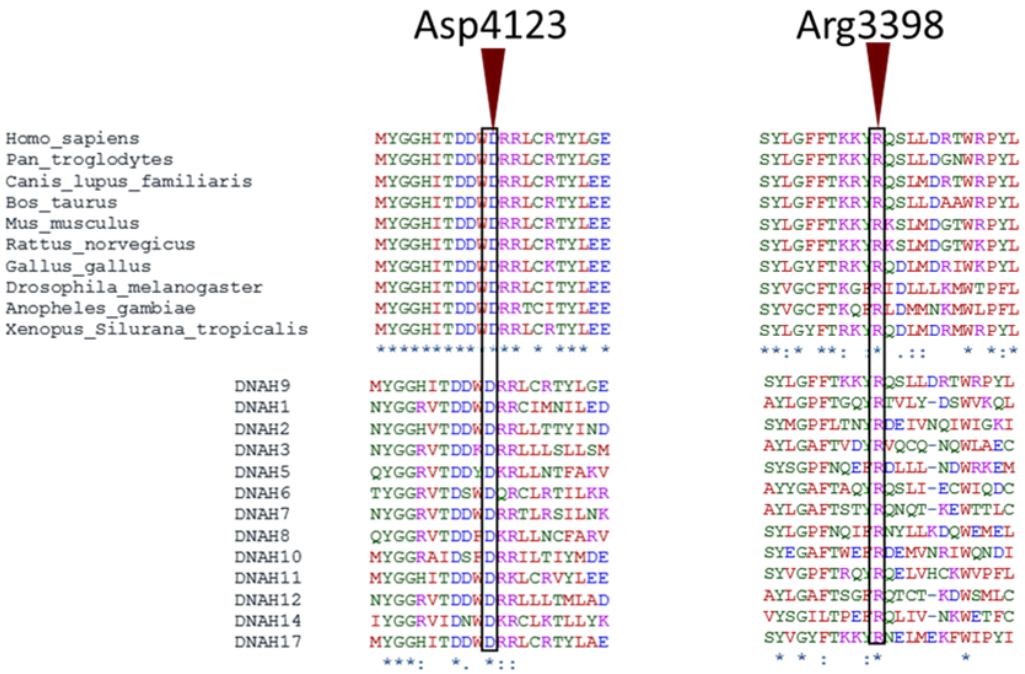


Figure 5-3 Phylogenetic analysis of Asp4123 and Arg3398 residues

Multiple sequence alignment (using ClustalW2 program) of DNAH9 between species shows that the identified missense mutations affect highly evolutionarily conserved residues that are also conserved across different dynein heavy chains in humans.

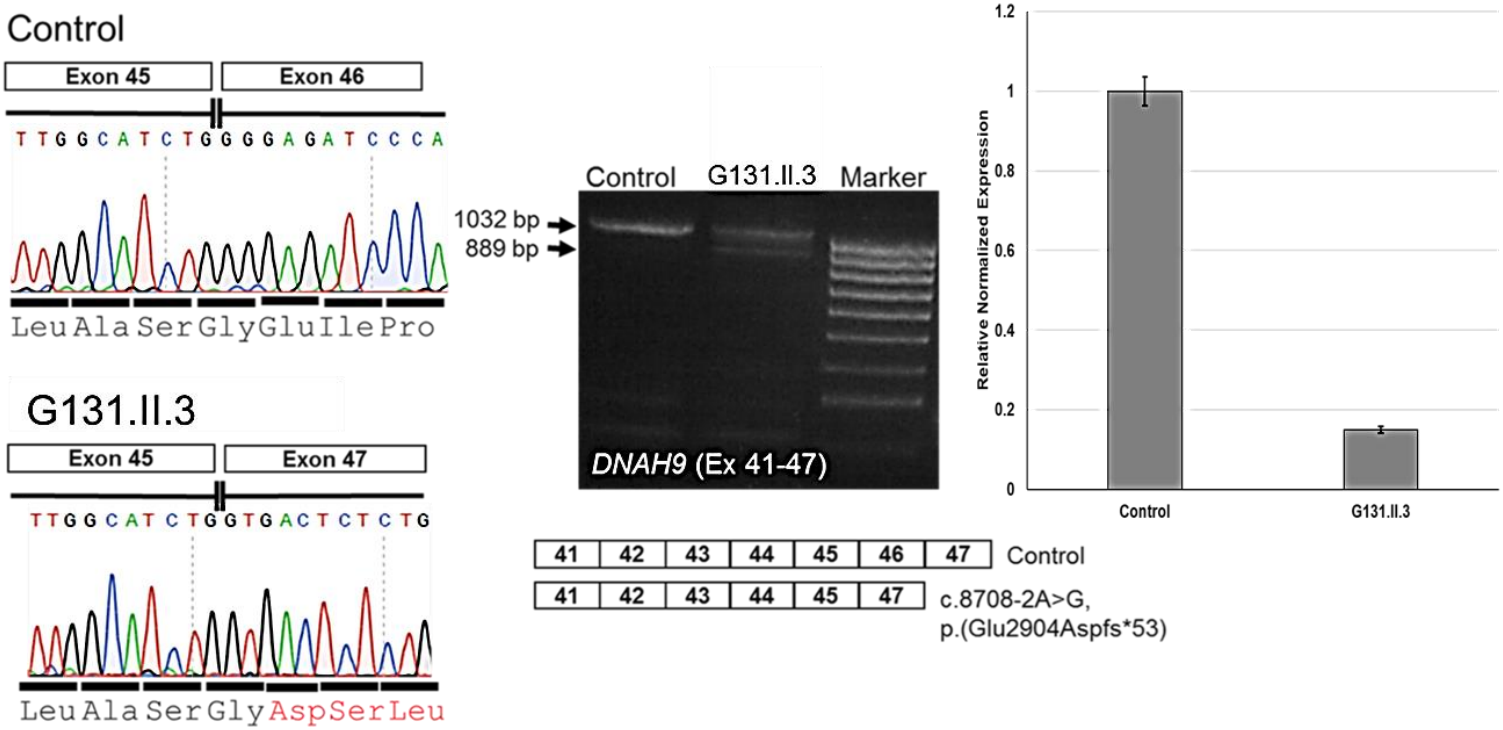


Figure 5-4 Functional impact of DNAH9 splice acceptor mutation

Patient's nasal brushing RT-PCR (middle panel) followed by Sanger sequencing of the amplified bands using primers in exons 41 and 47 (left panel), showed skipping of exon 46 in the proband of family G131, due to a splice acceptor variant c.8708-2A>G at exon 46, predicted to cause a frameshift in the protein. Right panel, RT-qPCR analysis of *DNAH9* expression normalised to that of *GAPDH* as a housekeeping gene, showed reduced levels in the affected patient (G131.II.3). The massive reduction of DNAH9 expression level identified by RT-qPCR (left panel) compared to that shown by RT-PCR (middle panel) can be due to the different numbers of multiciliated cells between the patient and control samples.

5.2.2 *DNAH9* mutations disrupt ODA components at the distal portion of the human cilia

Immunofluorescence staining was performed to study the consequences of *DNAH9* mutations on the localization of DNAH9 and other ODA components in the respiratory cilia of the affected individuals. Using specific anti-DNAH9 antibody and the axonemal marker anti- acetylated tubulin, showed the expected localization of DNAH9 only at the distal portion of the cilia in healthy controls. There was no staining of DNAH9 in the cilia of patients with *DNAH9* mutations indicating a complete absence of DNAH9 protein. (**Figure 5-5**)

Immunostaining of other components of ODA apparatus was sought to show the effect of DNAH9 loss on their localization along the ciliary axoneme. This included staining against DNAH5 (dynein heavy chain protein normally presents along the whole length of the motile cilia) and DNAI1 (dynein intermediate chain protein normally presents along the whole length of the motile cilia). Specific antibodies showed localization of DNAH5 and DNAI1 along the whole length of the cilia in a healthy control but the DNAH5 and DNAI1 staining was severely reduced to completely absent distally in the cilia of patients with *DNAH9* mutations. This indicates that the type 2 ODAs which are present in the distal part of the cilia, lack DNAH5 and DNAI1 in these patients. (**Figure 5-6 and 5-7**) This finding suggests that absence of DNAH9 often leads to loss of DNAH5 and DNAI1 in the distal part of the cilia i.e. the whole ODA is lost.

Staining against CCDC114, a component of the ODA docking complex normally present along the whole length of the cilia, was also attempted. It showed the presence of CCDC114 along the whole length of the cilia in both patients and a healthy control. This indicates no loss of ODA docking complex in patients with *DNAH9* mutations. (**Figure 5-8**)

TEM ultrastructural analysis of patients' respiratory cilia done by Dr Amelia Shoemark (Royal Brompton Hospital, London) showed consistent results with these immunostaining findings. Examination of cilia cross sections showed a normal composition of type 1 ODAs, located in the proximal

segment of the cilia, corresponding to the region of microvilli (apical extensions which are shorter and narrower than the cilia). In contrast, the distal segment of the cilia displayed a predominantly loss or truncation of the ODAs. (**Figure 5-9**)

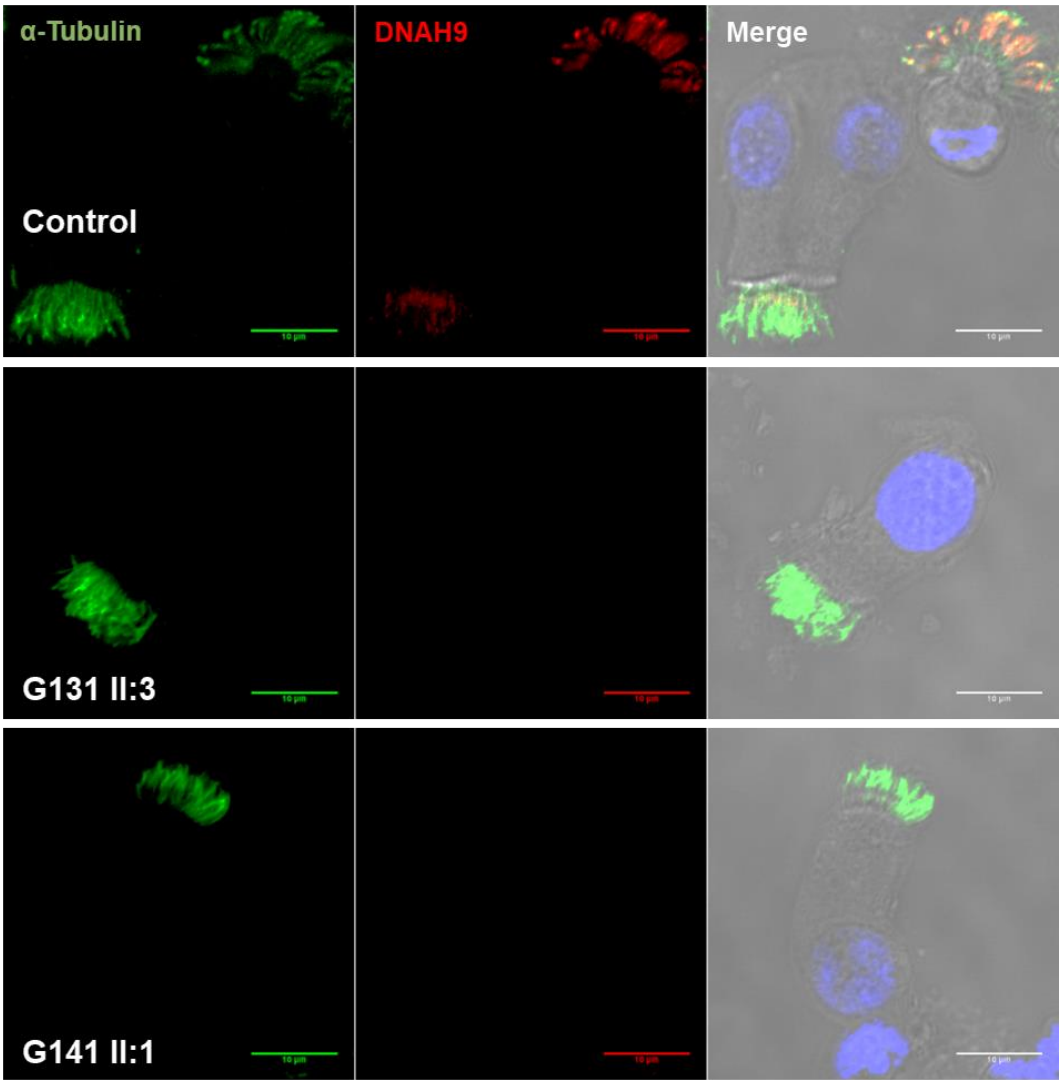


Figure 5-5 Loss of DNAH9 from the distal part of the cilia in patients from the two families with DNAH9 mutations

Using anti-DNAH9 antibody shows that DNAH9 (red), normally found at the distal part of control cilia, is absent from the cilia of affected individuals. Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 10 µm.

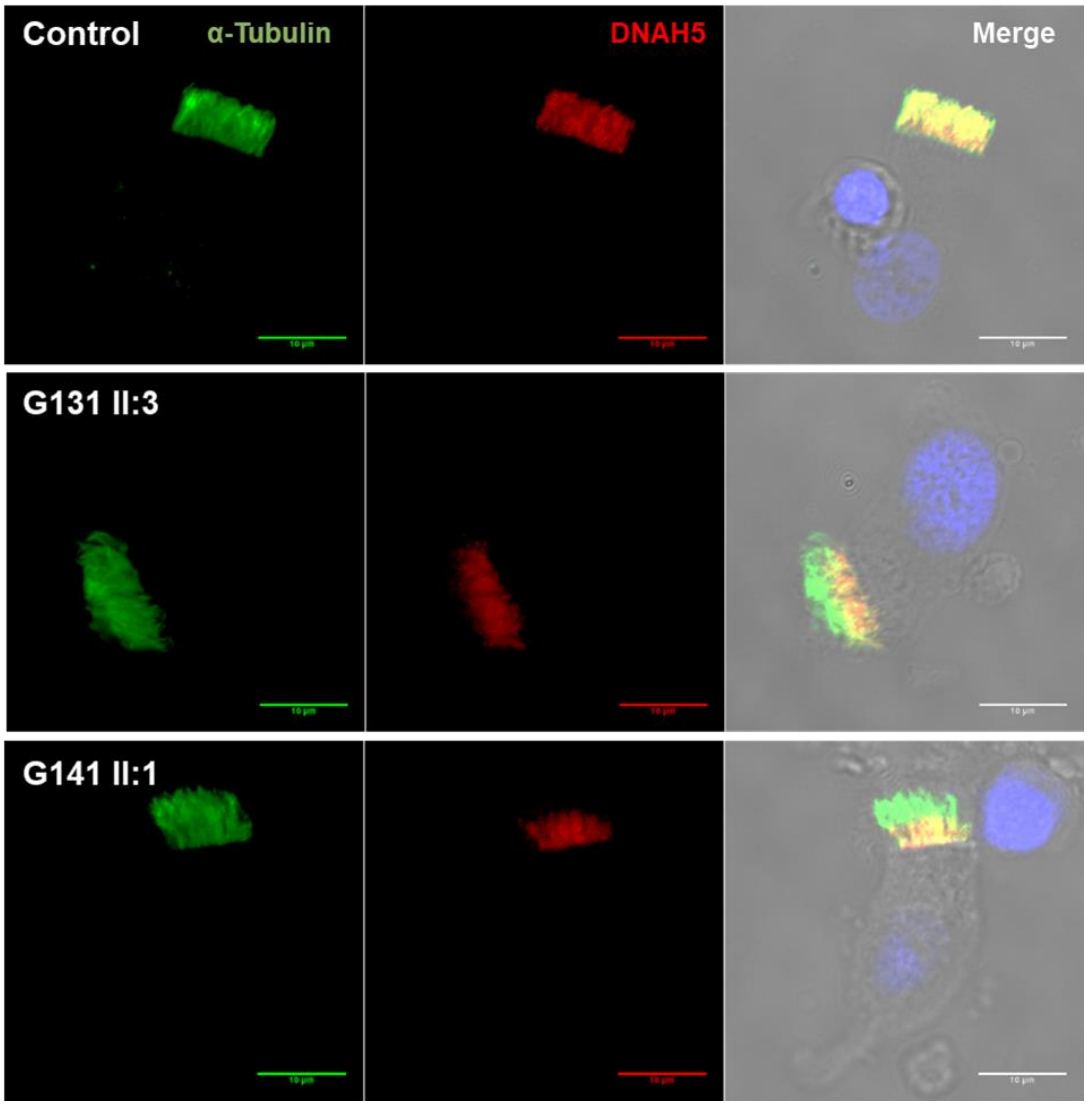


Figure 5-6 Loss of DNAH5 from the distal part of the cilia in patients from the two families with DNAH9 mutations

Using anti-DNAH5 antibody shows that DNAH5 (red), normally located along the whole length of the control cilia, is absent distally from the cilia of affected individuals. Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 10 µm.

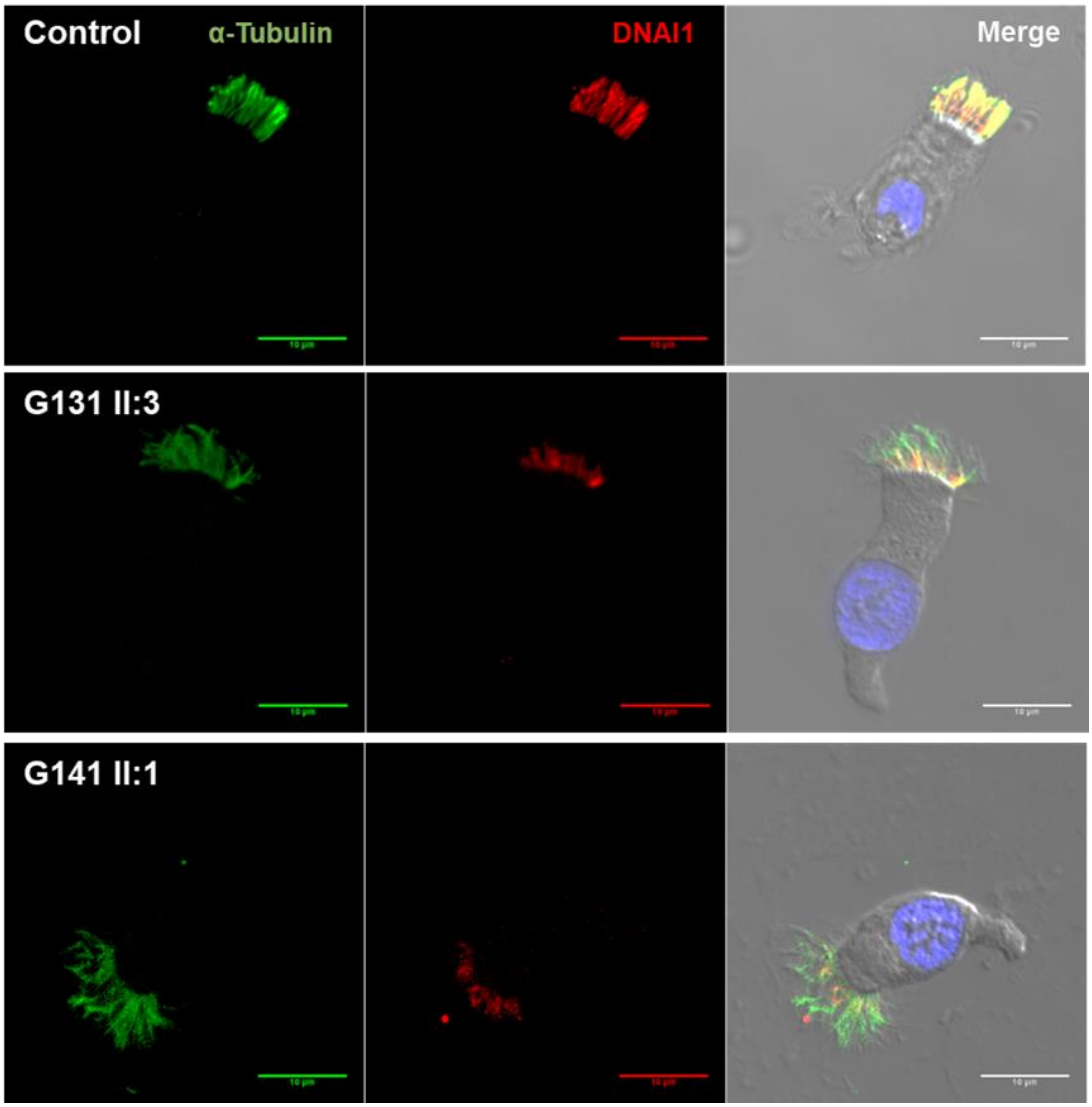


Figure 5-7 Loss of DNAI1 from the distal part of the cilia in patients from the two families with DNAH9 mutations

Using anti-DNAI1 antibody shows that DNAI1 (red), normally present along the whole length of control cilia, is absent distally from the cilia of affected individuals. Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 10 µm.

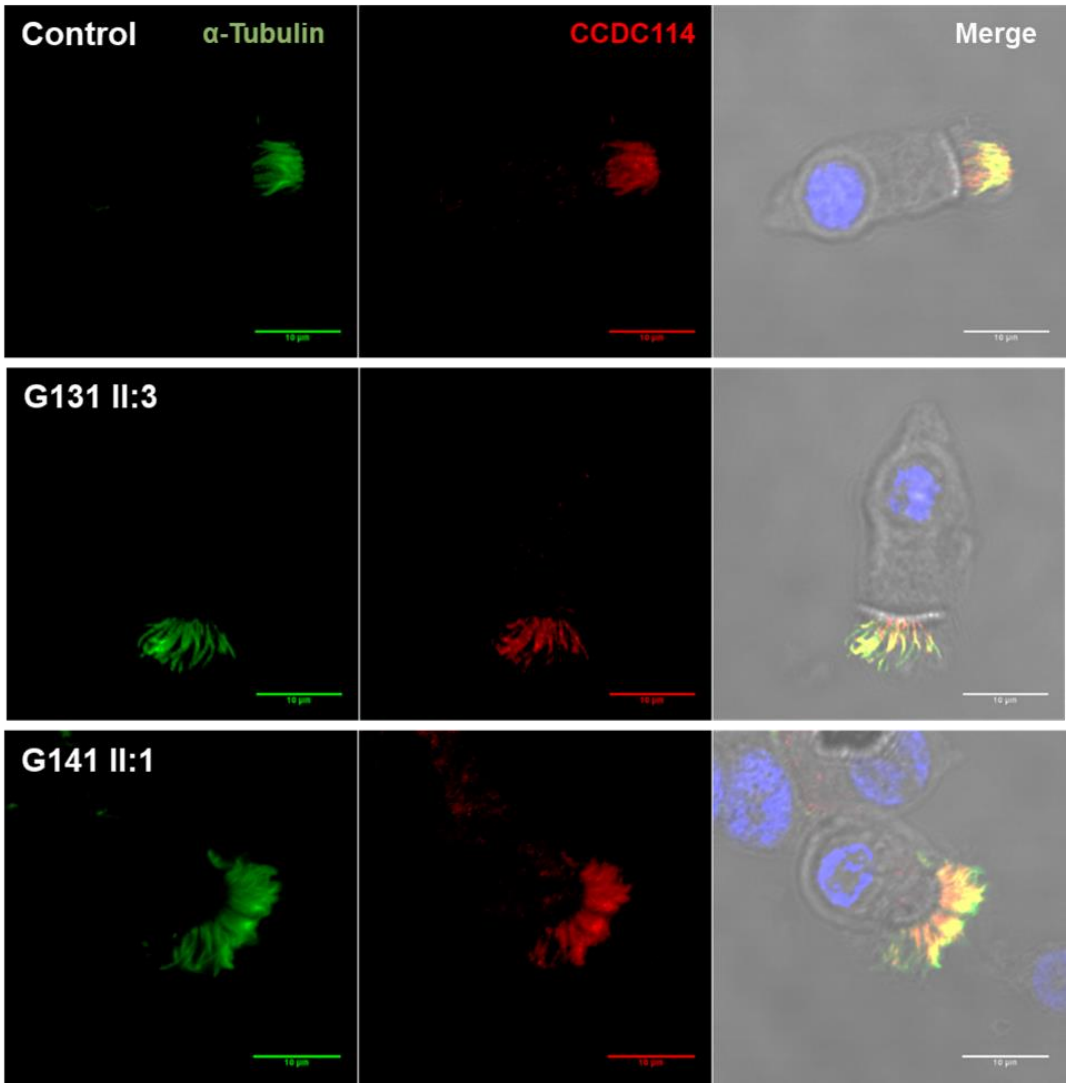


Figure 5-8 CCDC114 localization is undisturbed in patients' cilia with DNAH9 mutation

Using anti-CCDC114 antibody shows the distribution of CCDC114 (red) along the whole length of the cilia of both the patients and a control. Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 10 µm.

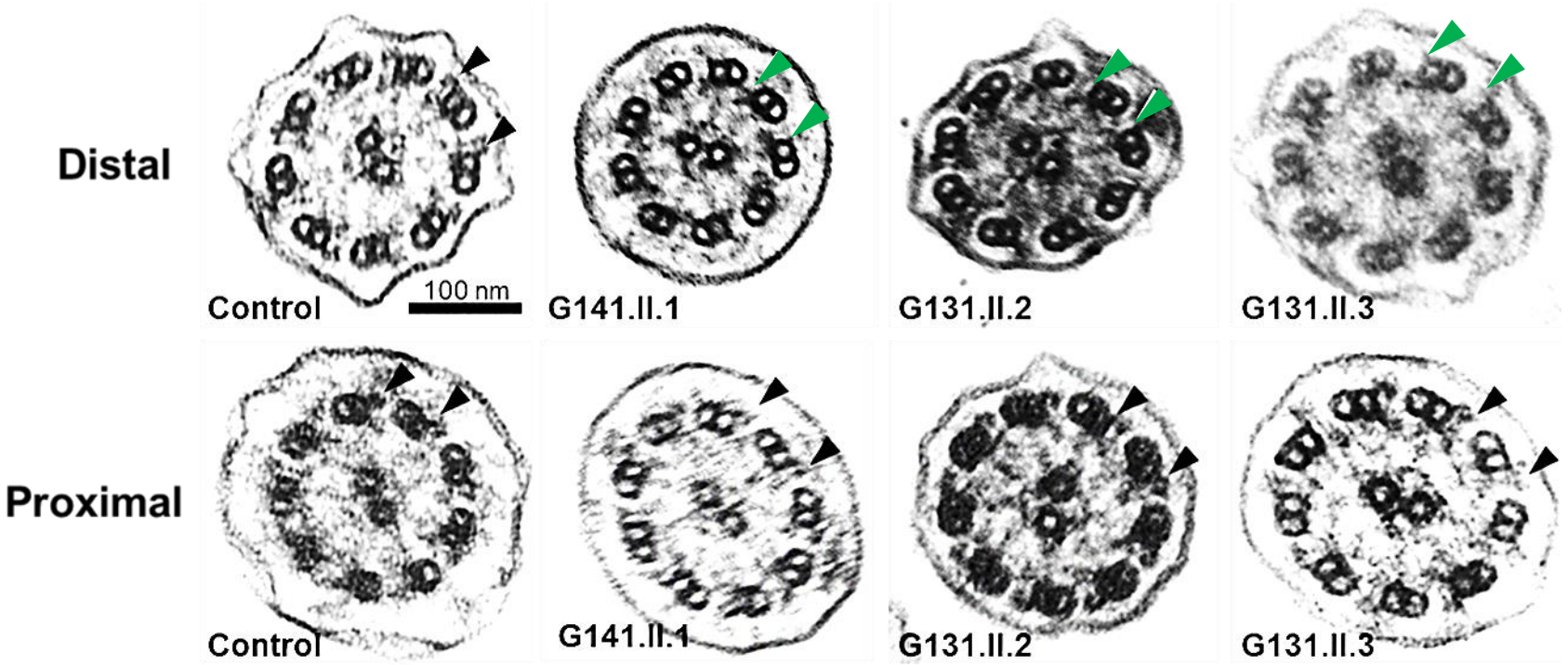


Figure 5-9 Representative TEM images of cilia cross sections comparing the proximal and distal parts of the cilia of DNAH9 patients against control

Black arrowheads highlight the presence ODAs and the green arrowheads highlight ODAs loss. ODAs are intact in the proximal and distal segments of the control cilia. They are present in the proximal parts of the patients' cilia but are lost distally. Scale bars,100 nm.

5.2.3 DNAH9 has a highly conserved role in cilia motility

To study the role of *DNAH9* gene in cilia motility and due to the highly conserved nature of ciliary proteins, silencing of *DNAH9* orthologs in *Paramecium* was attempted by the RNAi knockdown (KD), using the feeding method. This was followed by the assessment of the resulting knockdown phenotypes, as previously described in **Chapters 2 and 4**. Due to whole genome duplication, there are two *Paramecium* orthologs of *DNAH9* in the *Paramecium* genome. They correspond to the sequences (*GSPATG00038423001, GSPATG00038424001, and GSPATG00038425001*) and (*GSPATG00011643001, GSPATG00011644001, GSPATG00011645001*). (**Figure 5-10**) The RNAi construct used for knockdown of both orthologs was 100% identical to the sequence of both orthologs.

At first, the level of silencing was assessed by qPCR showing 80% reduction in the expression of both *Paramecium DNAH9* orthologs. The same as in the previous chapter, the KD-*ND7* strain was used as a control for the knockdown experiments (*ND7* gene has no role in cilia motility and its silencing is typically used as a control for gene silencing experiments in *Paramecium*). (**Figure 5-11**)

RNAi knockdown of both *DNAH9* orthologs in *Paramecium* led to an evident defective cilia motility phenotype. *Paramecium* cilia beating frequency was assessed by HSVM after 72 hrs of feeding. A significant reduction (p <0.001) in cilia beating frequency (CBF) was noted in KD-*DNAH9* cells (12.6 ± 3.5 Hz) compared to the control KD-*ND7* cells (24.3 ± 3.4 Hz). (**Figure 5-12**) This subsequently led to severe reduction in the swimming velocity of *Paramecium* as shown by dark field microscopy. (**Figure 5-13**) The swimming velocity was significantly reduced (p <0.001) in KD-*DNAH9* cells to about (0.23 ± 0.80 mm/s) compared to KD-*ND7 Paramecium* cells (0.85 ± 0.20 mm/s) after 72 hr of RNAi feeding. (**Figure 5-14**)

Ultrastructural analysis of the *Paramecium* cilia by TEM revealed a predominant loss of ODAs which was consistent with the ultrastructural

phenotype seen in the patients carrying *DNAH9* mutations. IDAs were also lost after *DNAH9* knockdown in *Paramecium*. Quantitative assessment of IDA and ODA loss revealed statistically significant reduction of both IDAs and ODAs from the cilia cross sections of KD-*DNAH9* cells compared to those of KD-*ND7* cells. The numbers of IDAs and ODs still present in the control KD-*ND7* per the examined cilia cross section (7.8 ± 1.1) and (7.9 ± 1.3) were significantly reduced to (5.6 ± 1.5) and (3.4 ± 2.1) in KD-*DNAH9* cells respectively (p <0.001). (**Figure 5-15**)

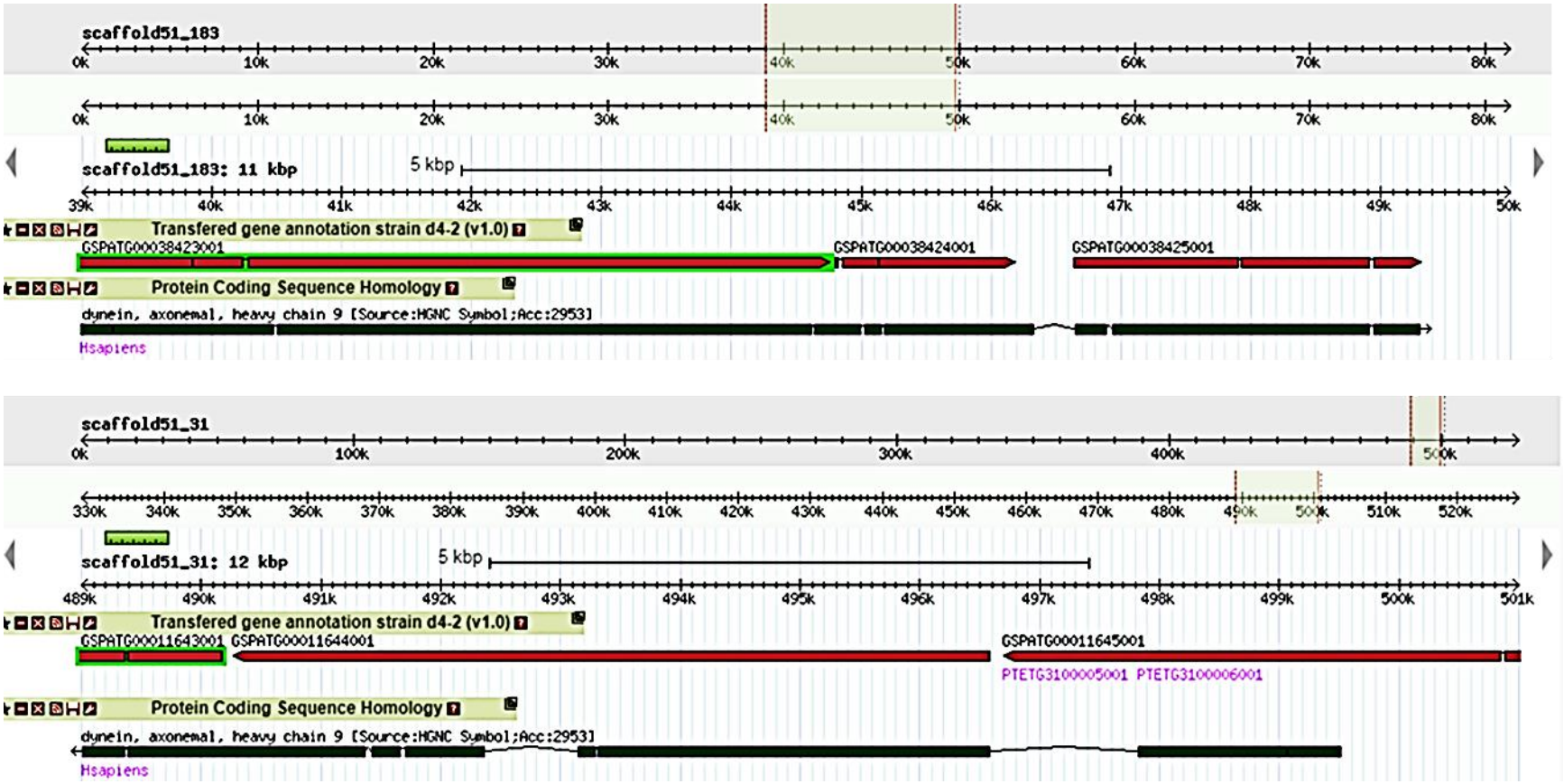


Figure 5-10 Two DNAH9 orthologs are found in the Paramecium genome

Screenshot of the GBrowse tool in the ParameciumDB database showing the two *DNAH9* orthologs found in the *Paramecium* genome and their coding sequence homology with the human protein.

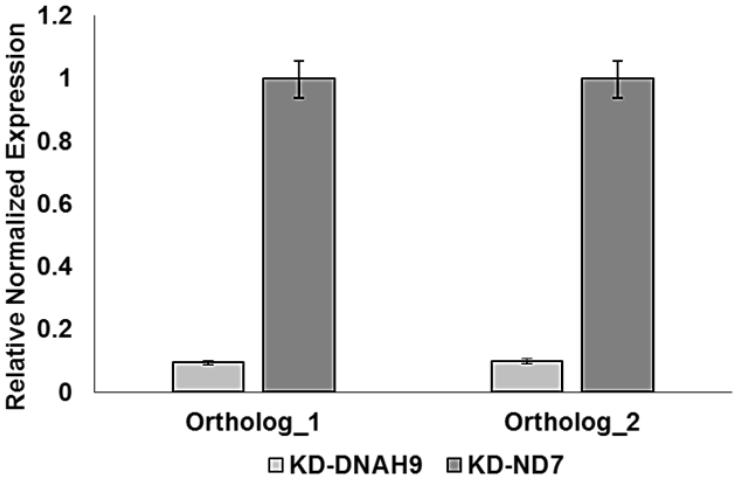


Figure 5-11 qPCR assessment of mRNA level of both the DNAH9 orthologs after knockdown in Paramecium

There was significant reduction in both the *DNAH9* ortholog mRNA levels of more than 80%, after RNAi by feeding in KD-*DNAH9* compared to KD-*ND7* *Paramecium*. All error bars indicate SEM.

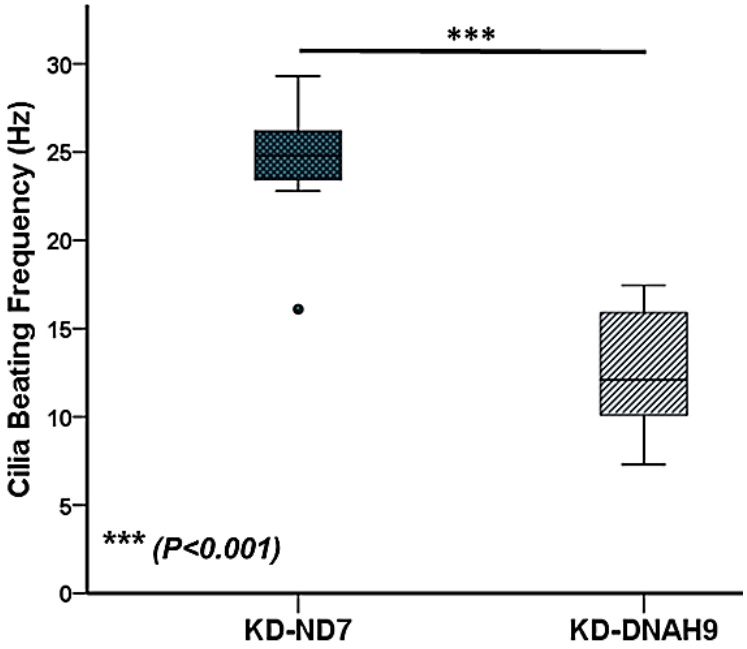


Figure 5-12 Cilia beating frequency was significantly reduced after DNAH9 knockdown

After 72 hr of RNAi feeding, the ciliary beating frequency of KD-*DNAH9* cells was significantly reduced compared to that of KD-*ND7* control cells. The *p* value was measured by Mann-Whitney U test.

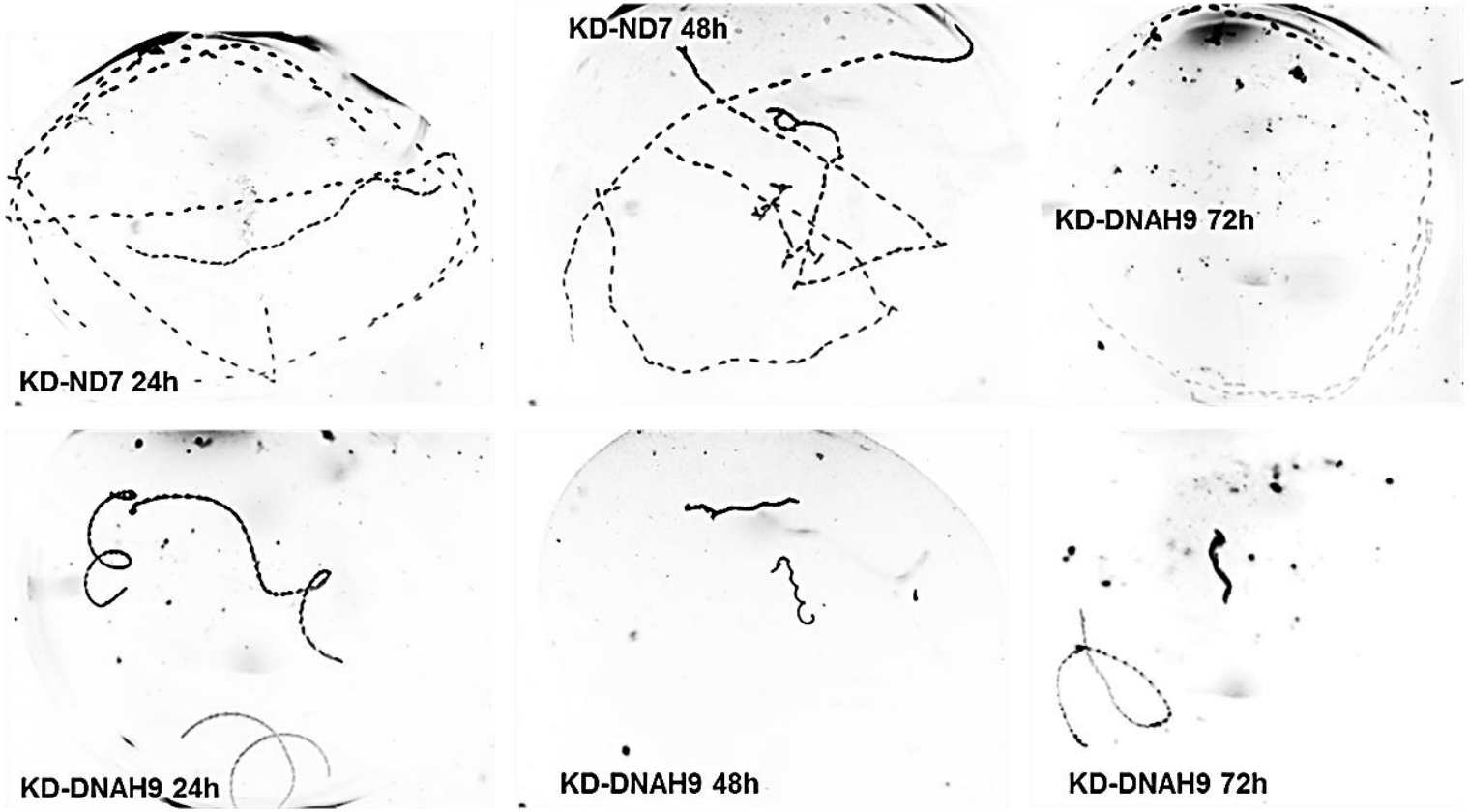


Figure 5-13 KD-DNAH9 Paramecium swimming velocity assessed by dark-field microscopy

Z projections of track recordings were captured after 24, 48 and 72 hr of RNAi feeding. Wider spaced dots indicate fast moving *Paramecium*, dots tighter together indicate slower moving or immobile *Paramecium.*

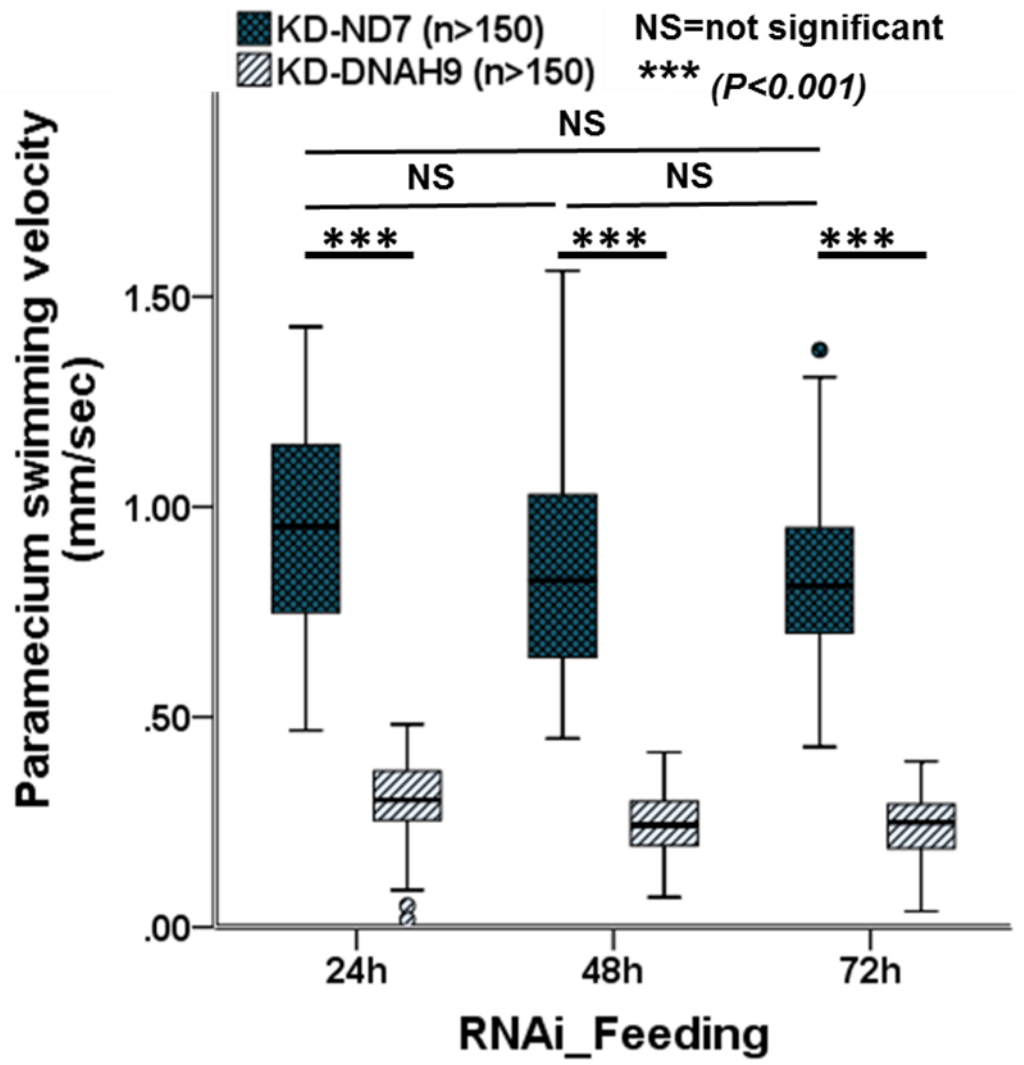


Figure 5-14 Significant reduction in the Paramecium swimming velocity after DNAH9 knockdown

Statistical analysis of the swimming velocity of *DNAH9* and *ND7* knockdown cells after 3 days of *DNAH9* RNAi by feeding analysed by independent sample t test. There was a significant reduction in the swimming velocity after *DNAH9* silencing compared to control *ND7* silencing.

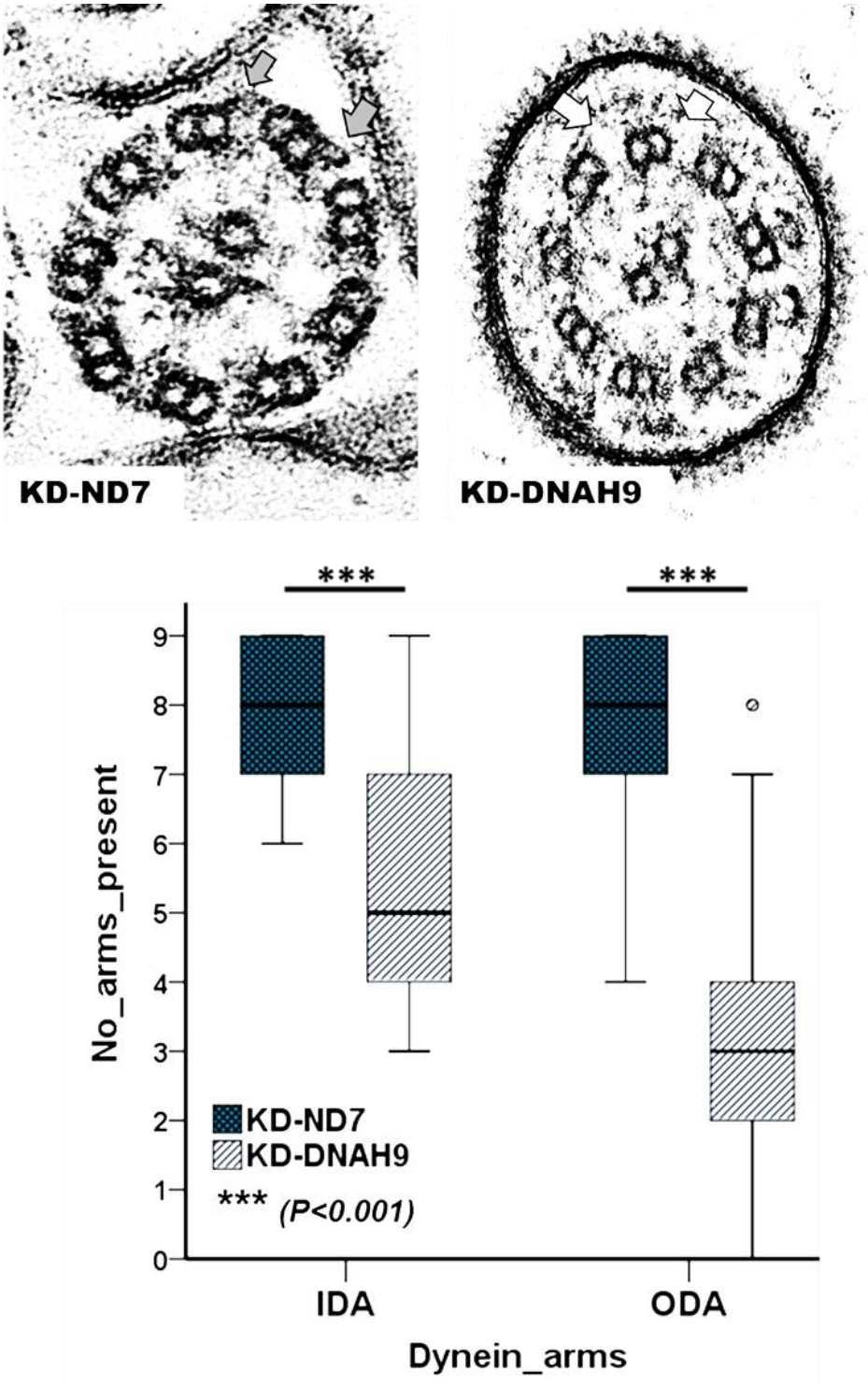


Figure 5-15 TEM analysis of KD-DNAH9 Paramecium cilia cross-section

Absence of the ODAs (white arrows) in the KD-*DNAH9* cells compared to the control KD-*ND7* cells where ODAs were present (grey arrows). Quantification of the dynein arms showed a significant reduction of numbers of ODA and IDA in KD-*DNAH9* cells compared to those in the KD-*ND7* control cells. *p* value was measured by independent samples t test.

5.3 Discussion

DNAH9 mutations were identified in 3 patients (from two unrelated families) attending PCD clinics. All three affected individuals have a distinct PCD phenotype that differs from the classical clinical course. All of the 3 patients identified in this study have situs inversus with upper respiratory symptoms but without any history of neonatal respiratory distress, otitis media or bronchiectasis. The latter are three classic disease markers. Moreover, the results of their clinical cilia functional diagnostic tests were inconclusive as follows. First, their nasal NO levels were either within normal or close to the normal range. The cilia beating frequency in all cases was close to the lower limit of the normal range in all affected individuals. As a note of caution, although the clinical phenotype of the patients is similar in the two families, it is still a low number of patients to conclude if there is a possible milder respiratory phenotype correlation with *DNAH9* mutations.

DNAH9 is one of the axonemal structural protein components of the outer dynein arm motors which are essential for cilia motility. ODA structure is highly conserved and is composed of multi-protein complexes. In *Chlamydomonas*, the ODA is composed of 3 heavy (400-500 kDa) chains (α, β, γ), 2 intermediate (45-100 kDa) chains and 8 light (8-55 kDa) dynein chains.(317, 325) Human ODA heavy chains have been isolated, INCLUDING DNAH5, DNAH9, DNAH11, DNAH8 and DNAH17. Phylogenetic analysis showed that their respective homologs in other organisms are components of outer dynein arms.(134, 135)

PCD causing mutations have been identified in 3 dynein heavy chain genes; *DNAH5*, *DNAH11* and here in this study, now *DNAH9*. DNAH5 is orthologous to the *Chlamydomonas* γ heavy chain dynein present along the entire length of the motile ciliary axoneme, with *DNAH5* mutation-related TEM defects easily seen as ODA loss from along the entire cilium associated with fully static cilia.(11, 86, 141, 143, 144) DNAH11 is the human ortholog of the *Chlamydomonas* β heavy chain. Its localization is restricted to the proximal part of the motile cilia, hence *DNAH11* mutation-related TEM defects and

associated ciliary dyskinesia are restricted to the proximal cilia; notably *DNAH11* mutations were previously thought to confer no TEM defects but these are now recognised to be subtle and also proximal cilium-restricted.(158, 161, 162, 326)

DNAH9 is human β heavy chain paralogue, a component of type 2 ODAs and it localizes distally in the motile cilia.(11, 162) Here, in patients with *DNAH9* mutations, ultrastructural analysis of the respiratory cilia showed predominantly a loss of ODAs restricted to the distal part of the cilia, with ODAs retained proximally. Because it was already known that DNAH9 is localised specifically to the distal cilium, we were aware that potentially only this portion would be affected, as turned out to be the case, but this kind of regional defect can be easily missed during TEM analysis as examining cross sections of the proximal part of the cilia only would lead to a conclusion of normal cilia ultrastructure or inconclusive results. This highlights the importance of genetic analysis for diagnosis and to better characterize PCD patients.

Immunofluorescence staining also showed loss of DNAH9 from the distal segment of the patients' cilia (where it is normally localized). Moreover, DNAH5 and DNAI1 were severely reduced in the distal portion of the DNAH9-mutant cilia but were still present in the proximal parts of the cilia. This is consistent with previous observations showing that DNAH9 was lost from the cilia of patients with *DNAH5* and *DNAI1* mutations.(11) Moreover, *DNAI2* mutations were shown to prevent the proper assembly of ODAs and subsequently DNAH9 was undetected in the motile cilia of patients with *DNAI2* mutations.(150) On the other hand, DNAH11 mutations interestingly led to a shift of DNAH9 distal localization, for it to redistribute along the whole length of the ciliary axoneme.(162) No evidence was found here for a parallel situation with *DNAH9* mutations, as DNAH11 does not appear to translocate into the distal ciliary segments when DNAH9 is absent. (Data from Dr Amelia Shoemark, Royal Brompton Hospital, not shown in this thesis)

Cilia are still beating in patients with *DNAH9* mutations with an apparently near normal cilia beat frequency. However, the cilia showed a

distinct beating pattern with subtle defective bending of the distal part of the cilia. Despite the finding that DNAH9 re-localizes to distribute itself along the ciliary axoneme when DNAH11 is defective or absent, *DNAH11* mutations are associated with stiffness and defective bending of the proximal part of the cilia leading to a hyperkinetic beating pattern.(162) This comparison showed that DNAH9 cannot functionally replace DNAH11, however it still presumably can heterodimerize with DNAH5 in the proximal parts of cilia.

Although two of the three *DNAH9* mutations described in the current study are missense mutations, loss of DNAH9 from the patients' cilia was shown by immunostaining that was equivalent to that of the -2 splice acceptor mutation, which is a predicted null frameshift allele from RT-PCR results. These missense mutations affected highly conserved residues within the DNAH9 protein that may possibly affect protein stability and subsequent protein degradation. It was previously shown that missense mutations can reduce the mutant protein half-life and lead to protein degradation despite normal mRNA expression.(327, 328) Moreover, mutations in the splice region can have a deleterious impact on mRNA splicing and eventually damaging effect on the protein level. These mutations can lead to either complete exon skipping, intron retention or introduction of a new splice site.(329) The splice acceptor mutation identified in the current study led to complete skipping of an exon and a predicted frameshift in the DNAH9 protein sequence. mRNA was not available from the patient (G141.II.1) with homozygous missense mutation, so the impact of this mutation on the transcriptional levels of *DNAH9* was not able to be assessed. However, the *DNAH9* mRNA level was reduced in the nasal brushing biopsy obtained from the patient (G131.II.3) with compound heterozygous mutations (missense and splice region) compared to that of a normal control. As the expression levels were normalized to *GAPDH* expression as a house-keeping gene, it is not clear if the reduced levels of *DNAH9* mRNA indicates a potential non-sense mediated decay process (330) or it is due to the unequal numbers of ciliated cells between the patient and control samples (which cannot be determined).

The fertility status of the *DNAH9*-mutation affected individuals could not be assessed as all of them were children. There is a high possibility that

DNAH9 mutations could lead to male infertility as DNAH9 distribution was shown along the whole length of the sperm tail.(11) Oligozoospermia with sperm immotility was shown in an Algerian adult male patient with two homozygous *DNAH9* missense mutations by a collaborating group (Data from Dr Marie Legendre, Hôpital Armand Trousseau, Paris).

Knockdown of DNAH9 orthologs in *Paramecium* led to loss of the dynein arms and a massive reduction in the *Paramecium* cilia beating frequency. This highlights the conserved role of DNAH9 in the structure of the ODA and in powering cilia motility across species. The loss of ODAs in *Paramecium* was not shown to be localized to certain part of the cilia however, most likely because the outer dynein arms in *Paramecium* are composed of 3 heavy chains, indicating the likely presence of DNAH9 along the whole length of the cilia.(331, 332)

5.4 Summary

In summary, mutations identified in *DNAH9* cause a distinct, variant PCD phenotype that likely arises from the limited location of DNAH9 in human cilia. A mild respiratory phenotype was noted in patients with *DNAH9* mutations raising a question about whether this phenotype can still be referred to as PCD at all. Loss of DNAH9 from the distal part of the cilia affects the localization of DNAH5 and DNAI1 distally and there is an abnormal pattern of cilia motility. DNAH9 has a conserved role in cilia motility as a structural component of the outer dynein arms.

Chapter 6 Motile cilia defects due to mutations in intraflagellar transport genes

6.1 Introduction

Ciliopathies are a broad range of human disorders caused by defective cilia structure or function. Dysfunction of ciliary and basal body proteins can affect primary and motile cilia either separately or together and result in first-order ciliopathies associated with multiple clinical features reflecting a range of defective functions of the basal body or ciliary compartment.(4, 333) Ciliary proteins may also have non-ciliary function and mutations in such proteins are linked to extra-ciliary dysfunctions e.g. disturbed intracellular collagen trafficking associated with *IFT20* mutations.(334) There is a wide array of human phenotypes of these disorders spanning skeletal anomalies, craniofacial defects, cystic kidneys, blindness, obesity, respiratory, laterality and fertility problems and other presentations.(26, 333, 335, 336)

Intraflagellar transport (IFT) is a highly conserved system within the cilia that transport structural (tubulin, ODAs, IDAs) and signalling components (Hedgehog, TGF-β, Wnt, PDGFRα, integrin and DNA damage repair signalling) along the axoneme in a bi-directional manner. It is essential for the assembly, stability and proper function of motile and primary cilia.(20, 302, 337) Anterograde IFT-B complexes carry cargos in the direction of the ciliary tip using kinesin-2 motors.(338) IFT-B is composed of a core component (IFT-B1) and several peripheral subunits which form the IFT-B2 sub-complex. The core component is composed of 9 proteins (IFT22, IFT25, IFT27, IFT46, IFT52, IFT70/TTC30, IFT74, IFT81, and IFT88). IFT56/TTC26 is another component of IFT-B complex potentially associated with IFT-B1 sub-complex.(339, 340) The peripheral subunits include (IFT20, IFT54, IFT57, IFT80, IFT172, and IFT38).(341, 342) IFT88 together with the N-terminal domain of IFT52, were shown to be essential for the interaction between the two sub-complexes.(342) (**Figure 6-1**) Retrograde IFT-A complexes convey the molecules back to the cytoplasm using the dynein-2 molecular motor complex. The IFT-A complex is composed of 6 proteins (IFT43, WDR35, IFT122, TTC21B, IFT140, and WDR19/IFT144).(343)

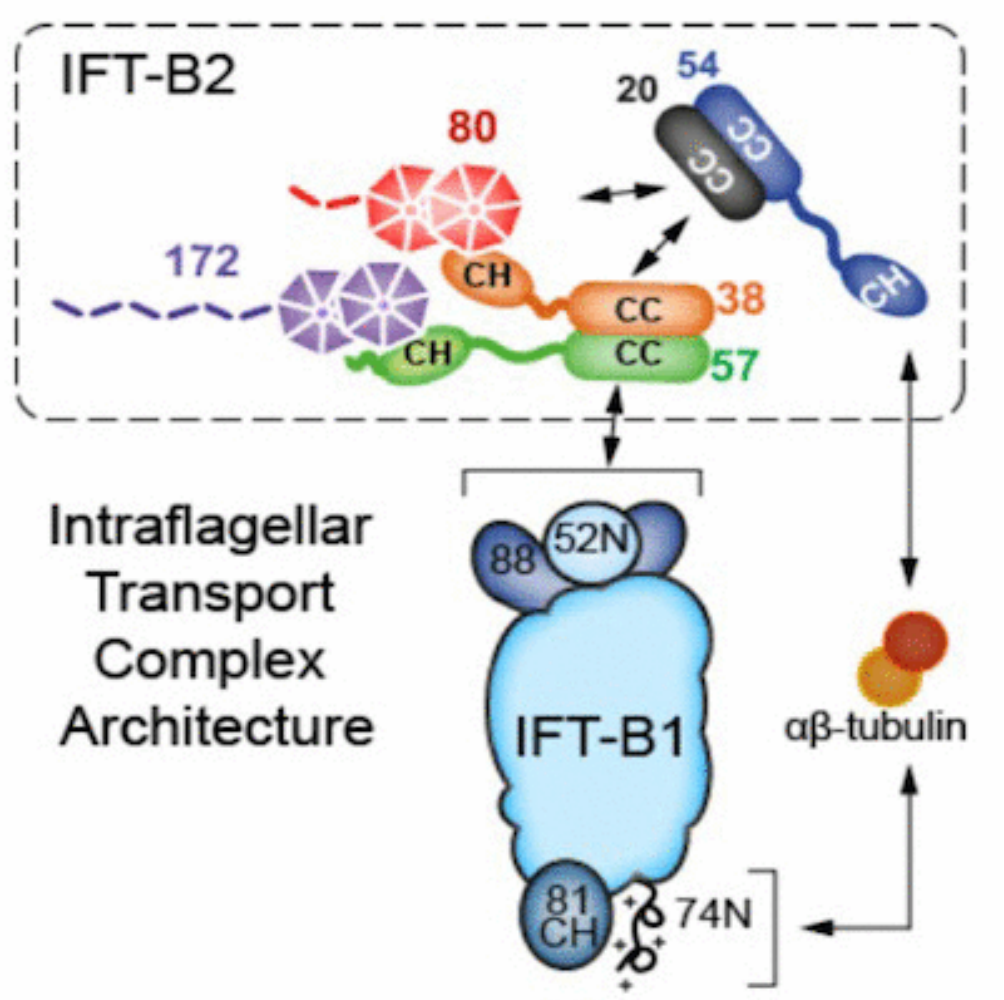


Figure 6-1 IFT-B complex is composed of two subcomplexes IFT-B1 and IFT-B2

The peripheral IFT-B proteins (IFT20, IFT54, IFT57, IFT80, IFT172 and IFT38) form a stable IFT-B2 sub-complex which interacts with IFT-B1 core sub-complex (through IFT88 and N-terminus of IFT52). The figure is taken from Taschner et al (2016).(342)

Phylogenetic analysis indicates that IFT-B is highly conserved and essential for cilia assembly. Mutations in the components of the IFT-B complex block cilia assembly leading to truncated or absent cilia in human and model organisms with consequences towards the severe end of the ciliopathy phenotypic spectrum.(31, 344-348) Recently, mutations in *IFT27* and *IFT74* were linked to Bardet-Biedl syndrome (BBS).(349, 350) On the other hand, defective IFT-A are associated with abnormal cilia architecture and bulging ciliary tips due to accumulation of IFT particles. Mutations in IFT-A components are associated with a broad spectrum of skeletal ciliopathies that extend across the non-lethal phenotypes: cranioectodermal dysplasia (CED, Sensenbrenner syndrome), more severe Jeune asphyxiating thoracic dystrophy (ATD) and the perinatal lethal short rib polydactyly syndromes (SRPS).(351-357)

Although IFT was first described in the motile flagella of *Chlamydomonas reinhardtii*, the motile cilia defects are mostly overlooked in patients with IFT mutations and respiratory symptoms are usually attributed secondarily to chest dysplasia and lung restriction.(302, 343) Lethal respiratory failure is usually linked to the severe end of the ciliopathy spectrum associated with perinatal lethal short rib polydactyly syndrome.(347, 348, 352, 354) Situs inversus was found in mouse embryos with defective IFT-A complex.(358) IFT140 has recently been shown to be important for spermatogenesis and male infertility in mice.(359) Motile cilia phenotypes including neonatal respiratory distress, recurrent respiratory infections, ciliary paucity, dyskinetic cilia motility and situs inversus totalis have been linked to mutations in *WDR35*.(38)

In this chapter, I will present the motile cilia defects associated with mutations in *IFT74* (a component of anterograde IFT-B complex) and *WDR19* (a component of retrograde IFT-A complex). I will furthermore illustrate the consequences of IFT74 inframe N-terminal deletion on its interactions with other IFT-B components.

6.2 Results

6.2.1 Targeted sequencing identifies an *IFT74* exon 2 deletion in patients with both PCD and a primary ciliopathy phenotype

Panel sequencing coupled with CNV analysis by ExomeDepth software revealed a homozygous genomic deletion spanning the whole length of *IFT74* exon 2 in two affected siblings (a girl and a boy) from a Palestinian family (PCD-G140) with consanguineous parents. (**Figure 6-2**) The variant analysis pipeline, carried out as described in **Chapter 2,** did not prioritize any other single nucleotide variant (SNV) of interest. This deletion was not found in CNV analysis of 180 samples screened in this project. It was also absent from the Database of Genomic Variants (DGV).(360)

To identify the breakpoints and the extent of the genomic deletion, a collaborator Dr Miriam Schmidts (Human Genetics Department, Radboud University Medical Center, Nijmegen, Netherland) has already done WGS for a case with a similar phenotype and exon 2 deletion identified by WES in previous work whilst at UCL then Nijmegen. The deletion spanned about 3 kb extending between the genomic coordinates (Chr9 (GRch38):26,959,923 - 26,962,979). By designing PCR primers outside the breakpoints followed by Sanger sequencing of the amplified bands, the same homozygous deletion was confirmed in both affected siblings and a carrier status of the parents and one of the unaffected siblings was also established. There was an insertion of 11 bp of unknown origin between the two breakpoints. (**Figure 6-3**) The same insertion sequence was also identified by Dr Miriam Schmidts in the additional IFT74 exon 2 deletion patients.

Both affected children in family G140 displayed features of skeletal dysplasia with short upper limbs, brachydactyly, mild bowing of the legs, dolicocephaly and exaggerated lordosis of the spine. Both of them had disproportionate short stature. The elder girl's height was at -5.8 SD (Standard Deviation) at the age of 8 years and the younger boy's height was at -6.8 SD

at the age of 3 years based on the World Health Organization (WHO) growth curves appropriate for each age and gender. There were no narrow thorax, polydactyly or renal problems in either but both patients had retinitis pigmentosa.

Both affected siblings had a history of neonatal respiratory distress, chronic productive cough, rhinosinusitis, otitis media, recurrent pneumonia and repeated hospital admissions. No situs inversus or cardiac abnormality were noted in either of them. The older affected girl showed low nasal NO level, but it could not be measured in the boy due to his young age. The girl also displayed signs of bronchiectasis in her lung by CT (computed tomography).

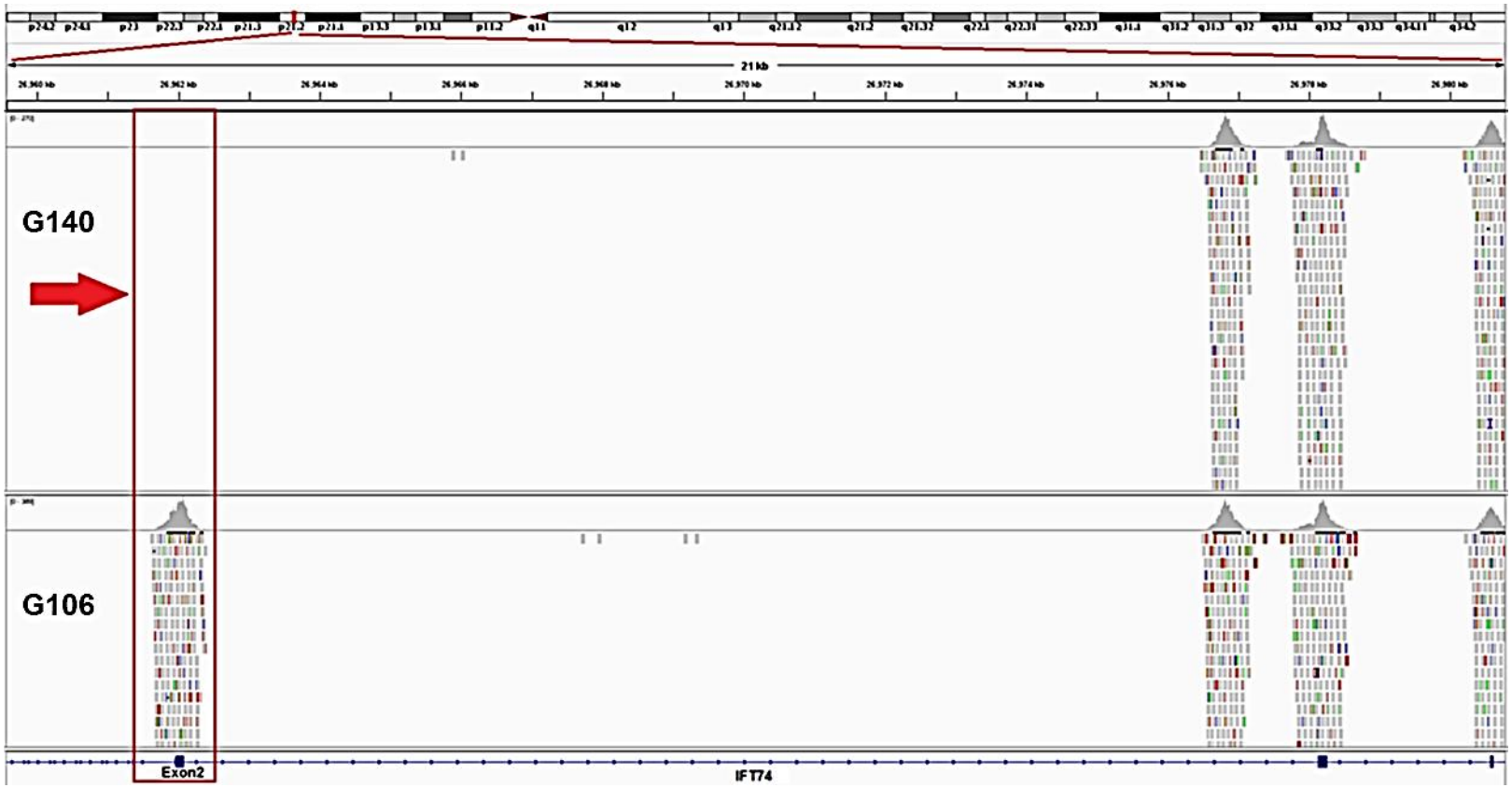


Figure 6-2 IFT74 exon 2 deletion identified by targeted sequencing

IGV (Integrative Genomics Viewer) screenshot of the BAM file of the affected girl in family G140 and the proband in family G106 (from the same sequencing run), aligned to the reference genome (GRCh37/hg19) shows that no reads were detected for exon 2 of *IFT74* in the affected girl compared to the other patient's sample in the same sequencing run. There was a uniform coverage of other IFT74 exons among all samples in the same run (full data not shown).

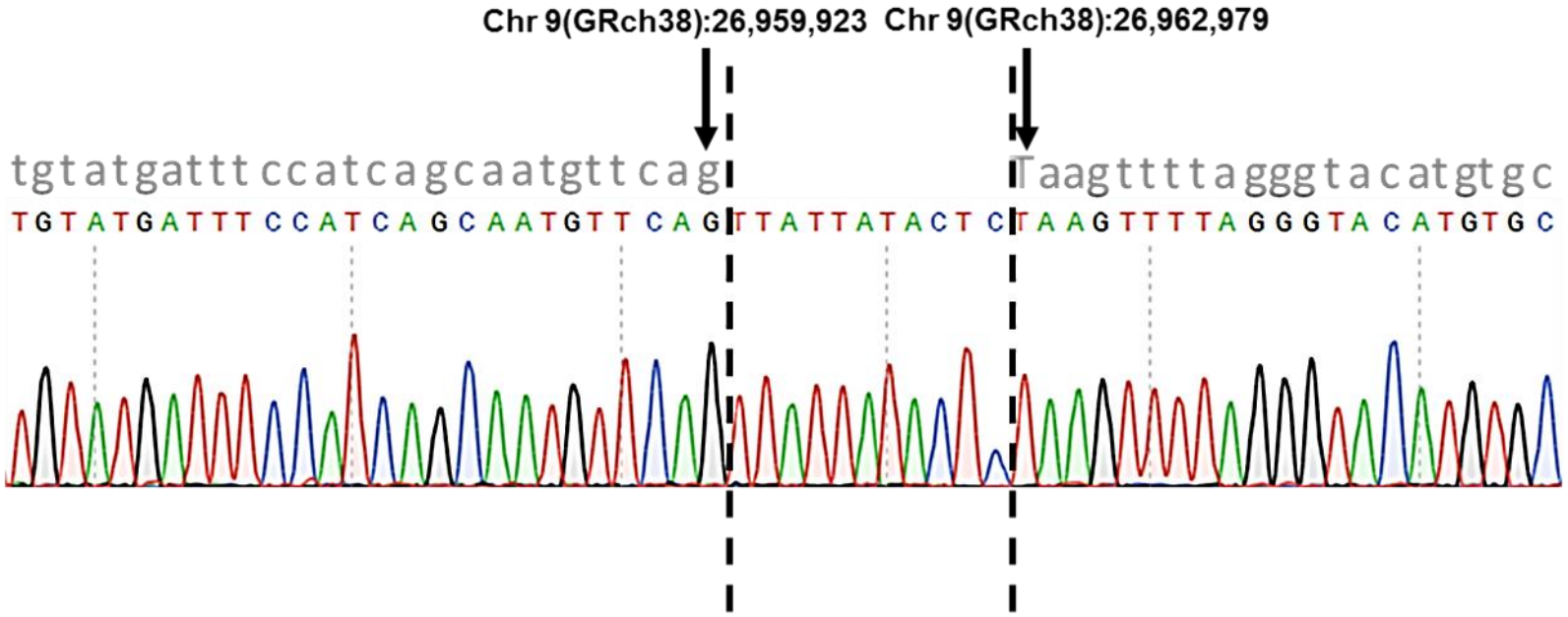


Figure 6-3 IFT74 genomic deletion of about 3 kb, confirmed by PCR and Sanger sequencing

Sanger sequencing following PCR using primers encompassing the genomic breakpoints of the deletion confirmed the homozygous states of the mutation in the affected children and carrier state of the parents and one unaffected sibling and revealed an insertion of 11 bp of unknown origin.

6.2.2 *IFT74* exon 2 deletion causes defective motile cilia

Investigating the consequences of *IFT74* exon 2 deletion on primary and motile cilia was pursued. A skin biopsy was taken from the patients but unfortunately, it was infected when delivered and fibroblasts couldn't be cultured. The analysis of the potential primary cilia defects was not possible.

TEM analysis provided by Dr Patricia Goggin from the PCD centre, University Hospital Southampton from 3 brushings (nasal and bronchial) from the affected girl and two brushings from the boy, showed short sparse respiratory cilia. Variable ciliary axoneme ultrastructural abnormalities were noted in the few cilia cross sections obtained, mainly affecting the microtubules showing reduced numbers of central pairs or peripheral doublets, microtubular translocations and loss of microtubular integrity. There was also disrupted orientation of the cilia basal feet which is determined by the integrity of the ciliary microtubular structure and the successful generation of cilia fluid flow.(41, 361) The ultrastructure of the basal bodies, where visible, appeared to be normal in longitudinal sections. (**Figure 6-4)**

Immunofluorescence staining of motile cilia obtained from the nasal brushing biopsies of the patients confirmed the TEM findings showing markedly sparse shorter cilia in the patients cells compared to their mother or a healthy control. Staining for DNAH5 (ODA component) showed reduced levels along the short cilia of the patients compared to normal distribution along the whole length of the ciliary axoneme of the healthy control. (**Figure 6-5**) Staining for IFT74 detected an abnormal distribution in the patients' cells compared to the control, since rather than its normal localisation at the base and along the ciliary axoneme, it was found to be mislocalized to all over the cytoplasm but more accumulated at the apical cytoplasm. (**Figure 6-6**) Staining for protein components of the IFT-B complex (IFT81, IFT88) showed a marked reduction in their axonemal staining in the patients' cilia. (**Figure 6-7 and 6-8**) Interestingly, staining for IFT140 (a component of the IFT-A complex) showed accumulation of the protein at the apical cytoplasm with reduced staining along the ciliary axoneme. Staining for DYNC2LI1, a

component of the IFT dynein-2 retrograde motor complex showed a similar pattern of reduced staining in the ciliary axoneme of the patients. (**Figure 6-9**) Whether the staining is genuinely reduced in the axoneme or is simply altered as a secondary result of the shortened, stubby cilia, is difficult to assess with this kind of non-quantified approach.

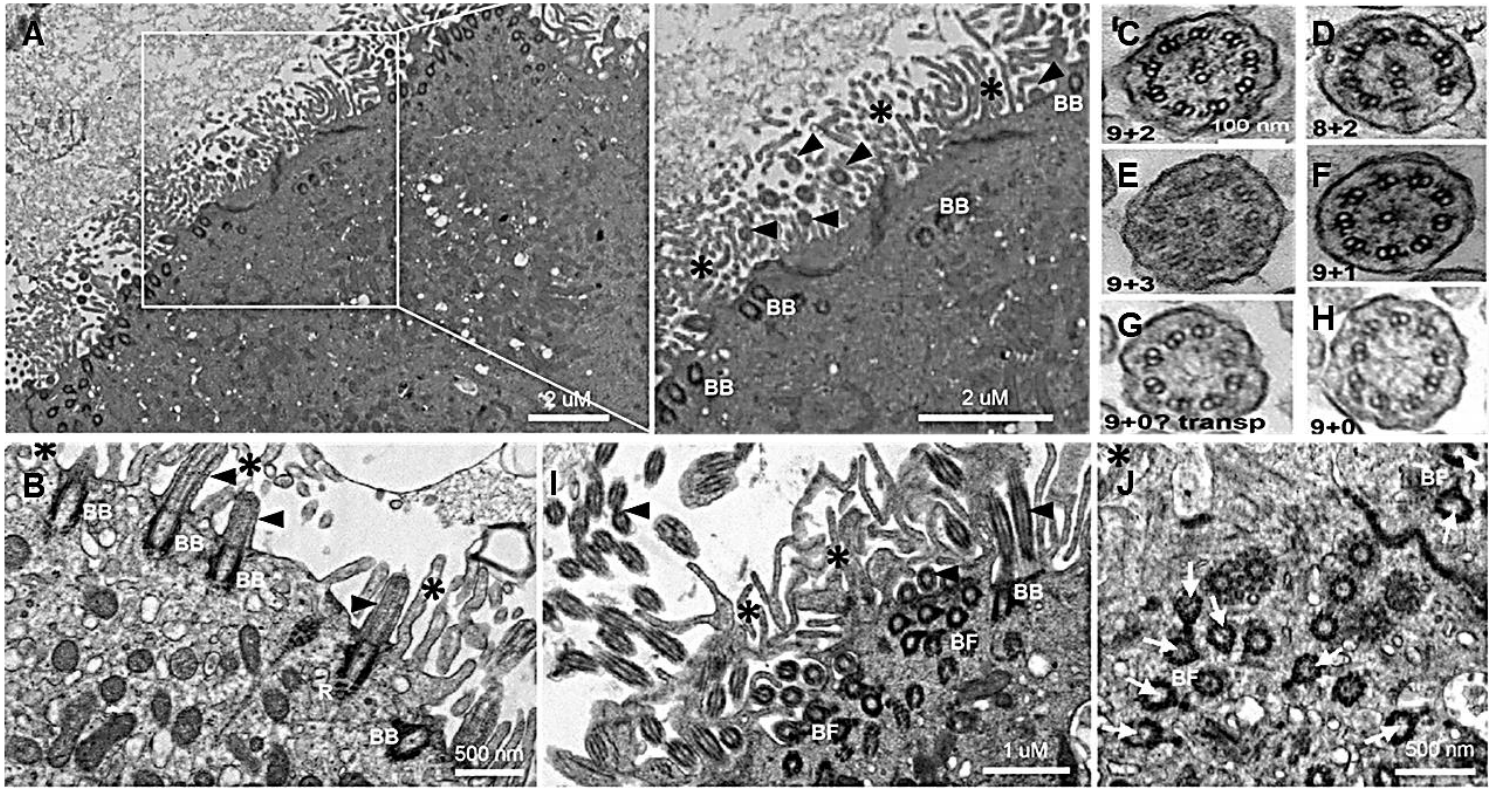


Figure 6-4 TEM ciliary ultrastructural findings in IFT74 exon2 deletion patients

The patients' cilia were markedly short and reduced in number. Normal basal body ultrastructure aligned along the apical surface of the cells was noted (A & B). A variable range of ciliary axoneme ultrastructural defects were noticed (C – H). Disorientation of the basal feet of the cilia was also shown (I & J). BB; basal body, BF; basal foot, R; root structure, Transp; transposition of the microtubules.

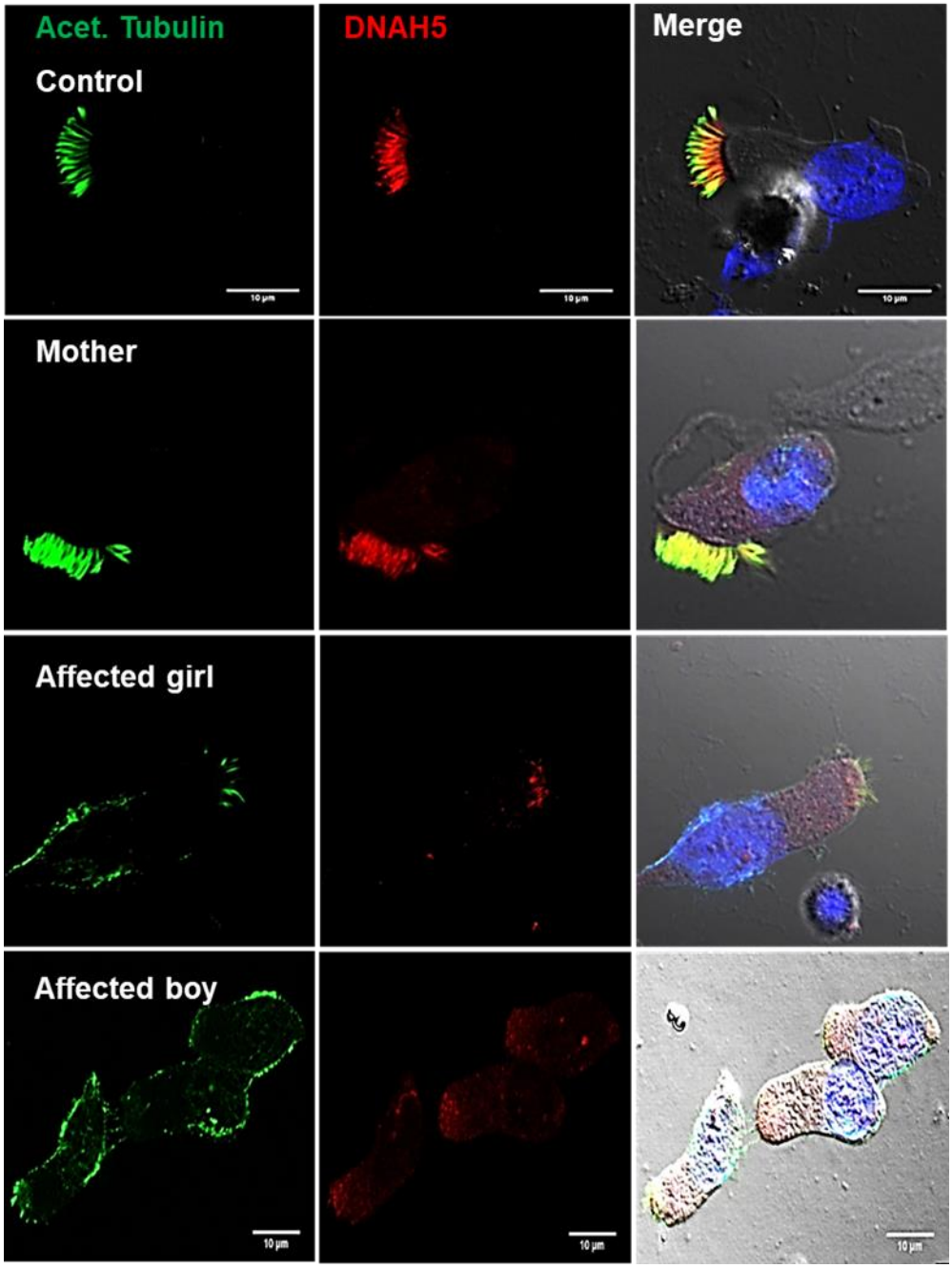


Figure 6-5 DNAH5 is reduced in patients from family G140 in the short sparse cilia and accumulates in the apical cytoplasm

Immunofluorescence staining using anti-DNAH5 antibody shows less DNAH5 (red) along the short sparse patients' cilia compared to its normal distribution along the axoneme of the respiratory cilia of the mother or healthy control. Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 10 µm.

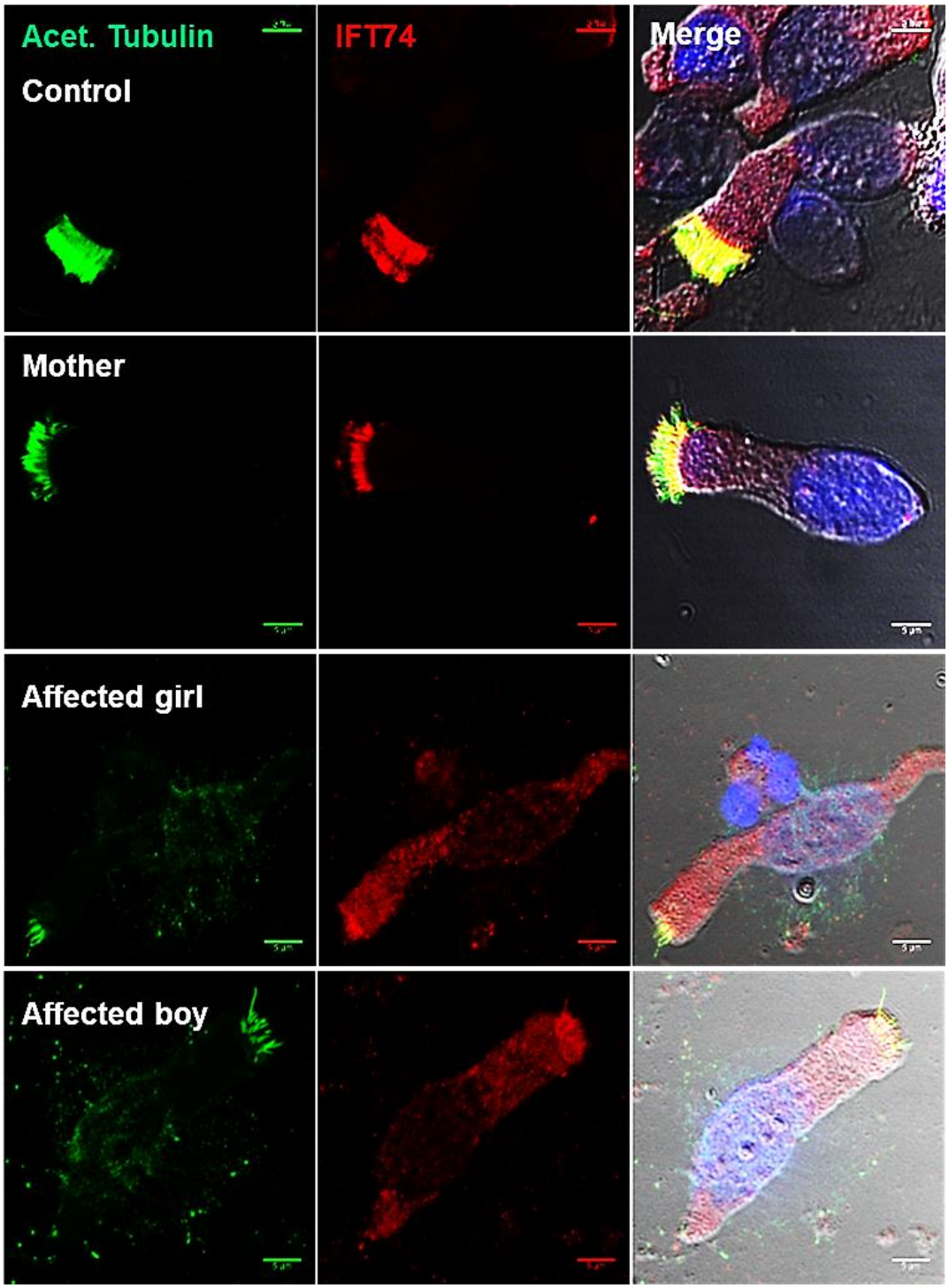


Figure 6-6 IFT74 is apparently present all over the cytoplasm and accumulates in the apical cytoplasm of family G140 patients' cilia

Immunofluorescence staining using anti-IFT74 antibody shows abnormal distribution and a reduced level of IFT74 (red) along the short sparse patients' cilia compared to its normal distribution along the cilia of the mother and healthy control. Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 5 µm. There was no change in the image acquisition settings used for the patients, mother and the healthy control samples.

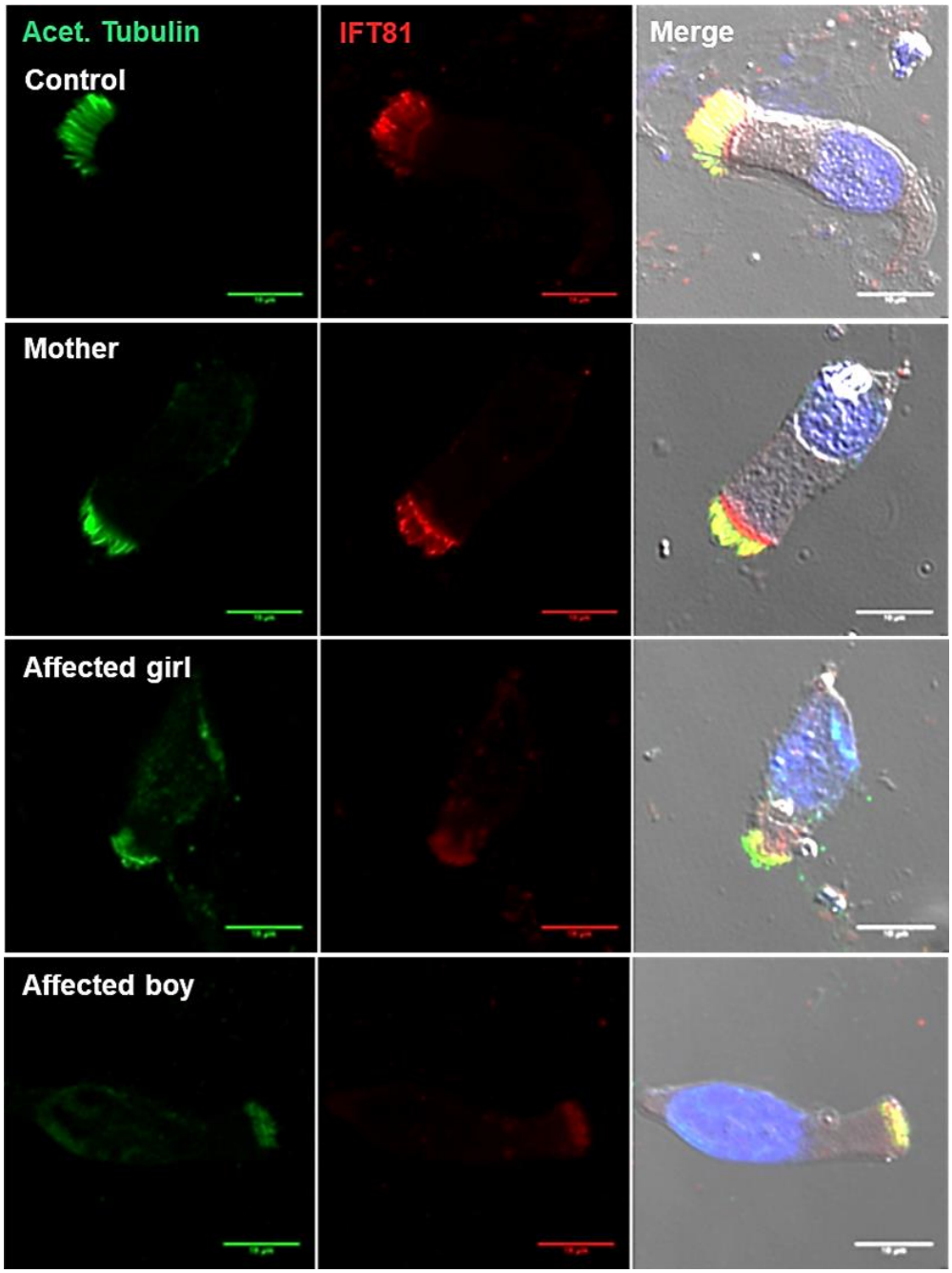


Figure 6-7 IFT81 is reduced in the family G140 patients' cells

Immunofluorescence staining using anti-IFT81 antibody shows reduced IFT81 (red) along the short sparse patients' cilia compared to its normal distribution along the cilia of the mother and healthy control. Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 10 µm.

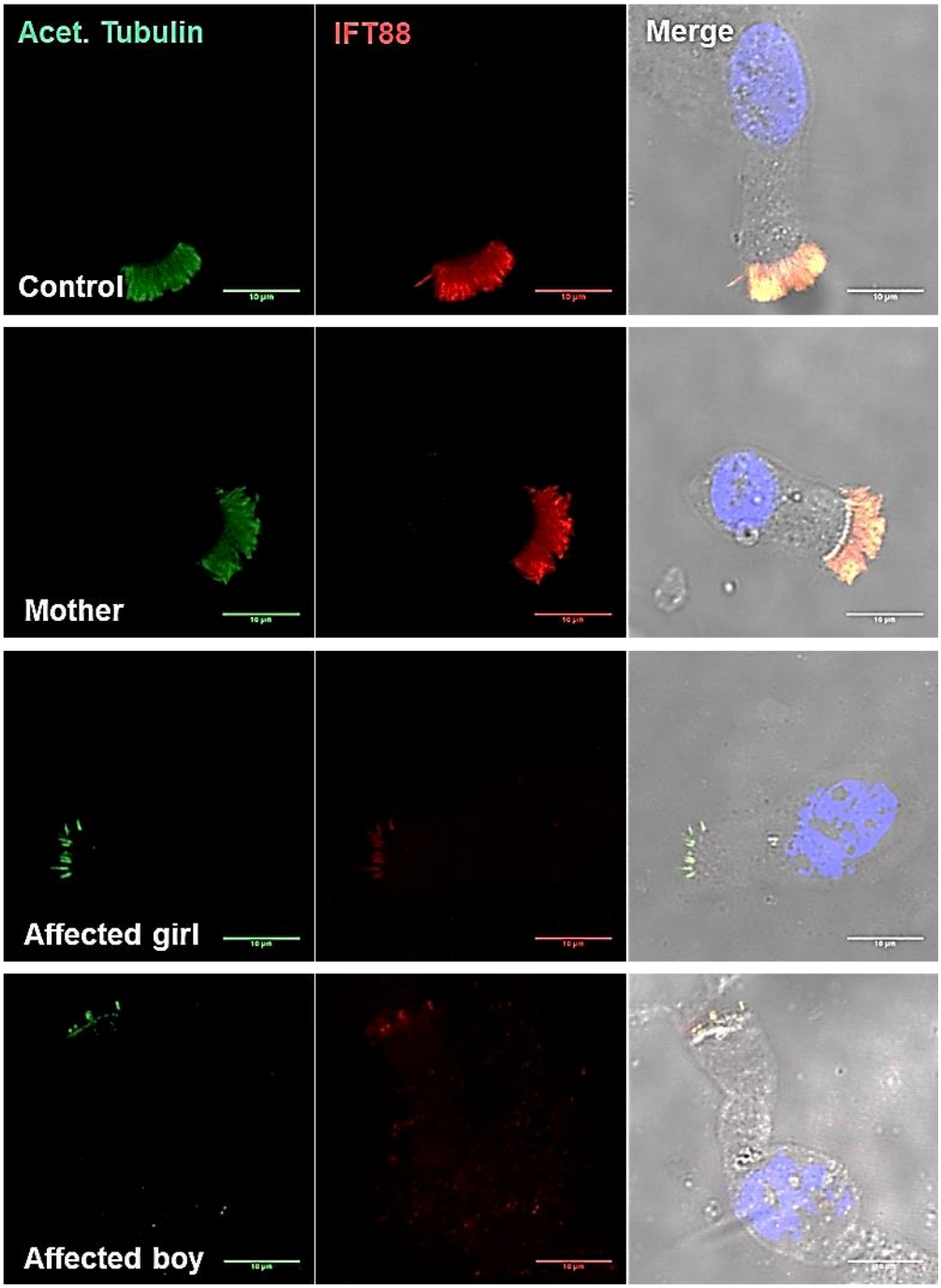


Figure 6-8 IFT88 is reduced in the family G140 patients' cells

Immunofluorescence using anti-IFT88 antibody shows reduced IFT88 (red) in the short sparse patients' cilia compared to its normal distribution along the whole length of the motile cilia of the mother and healthy control. Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 10 µm.

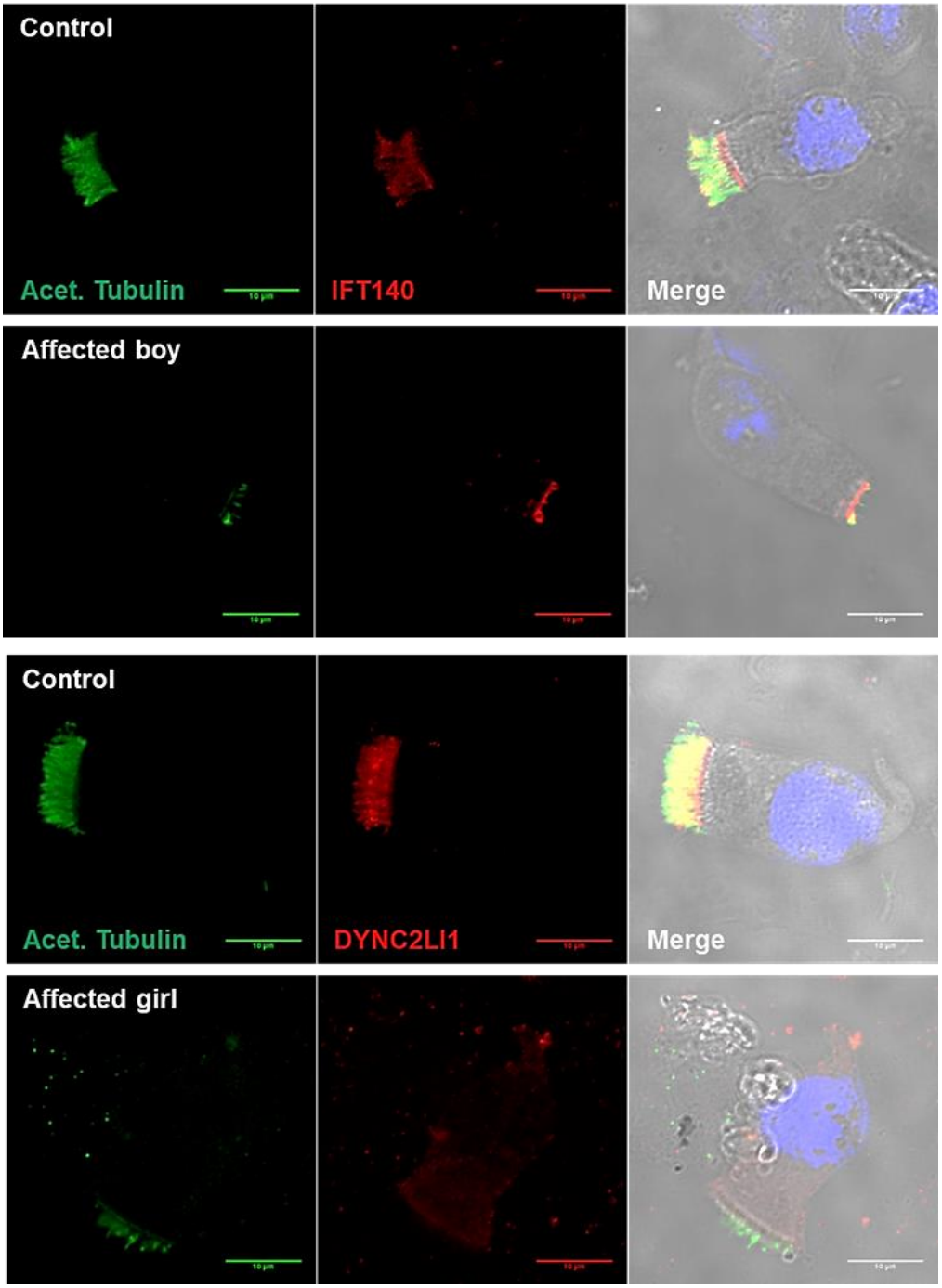


Figure 6-9 IFT140 and DYNC2LI1 are reduced in the family G140 patient's cells

Immunostaining using anti-IFT140 (top, red) and anti-DYNC2LI1 (bottom, red) antibodies shows reduced levels of both proteins in the short sparse patients' cilia compared the cilia of the healthy control. Accumulation of the IFT140 (red) was noted at the apical cytoplasm. Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 10 µm.

6.2.3 Consequences of *IFT74* exon 2 deletion on its mRNA and protein

To determine the effect of the deletion on the *IFT74* mRNA of the patients, RT-PCR was performed on mRNA extracted from nasal brushing sample of the patient, followed by Sanger sequencing of the amplified band. This revealed a complete loss of exon 2 only (NM_025103 transcript was used in analysis), which is the first coding exon of the *IFT74* gene. There was no effect on the upstream exon 1 (non-coding exon) or downstream exon 3 (which starts with an inframe ATG). (**Figure 6-10**) One caveat for the RT-PCR analysis presented here, is that it was directed by the specific primers chosen and does not rule out the possibility of alternative splicing effects that may result from this deletion. RNA-Seq, therefore, could be a more informative way to study the consequences of the deletion on all IFT74 transcripts expressed in multiciliated cells.

The transcriptional consequences of the exon 2 deletion on *IFT74* was assessed by qRT-PCR on mRNA extracted from the nasal brushings of one of the patients, mother and a healthy control of the same Palestinian-Arabic ethnicity and geographical region, using specific forward primers designed in *IFT74* exons 2, 3 and 4 and a reverse primer designed to span the corresponding downstream exon-exon junction. No quantification of *IFT74* transcripts was achieved in the patient's sample using the forward primer in exon 2 and the level of expression in the mother's sample (heterozygous carrier) was about half the level of expression of the healthy control, results consistent for the deletion inheritance pattern. In contrast, using primers in exons 3 and 4 showed higher expression levels of *IFT74* (nearly doubled) downstream to the deletion compared to the mother and healthy control. This indicated a possible feedback compensatory mechanism where *IFT74* transcriptional upregulation was being used by the cells to try to overcome the potential defective function of IFT74 secondary to the deletion.

The relative mRNA expression levels of other IFT components were also quantified. Other IFT-B components (*IFT81* and *IFT20*) were expressed at higher levels in the patient compared to the control while they showed

slightly reduced expression in the mother compared to the healthy control. In a similar expression pattern, *IFT140* and *DYNC2LI1* were elevated in the patient compared to the healthy control. (**Figure 6-11**)

Western blotting following protein extraction from nasal brushing biopsies from the patients, mother and a healthy control showed a band of full length IFT74 protein at 69 kDa in the mother's and the healthy control's samples. This band was absent from the patients' samples which had 3 smaller bands of approximately 65, 61 and 57 kDa assumedly representing 3 in-frame truncated IFT74 protein products. Although this is a speculative assumption, we cannot rule out alternate explanations including potential alternative splicing, degradation or post-translational modifications. These 3 truncated protein products were faintly noted in the mother's sample. (**Figure 6-12**)

To understand the translational consequences of the deletion, the IFT74 protein sequence (UniProtKB - Q96LB3) was investigated. There are two protein isoforms produced by alternative splicing: isoform 1 or the canonical isoform (Q96LB3-1) which is formed of 600 amino acid length and has a predicted molecular weight of 69 kDa and isoform 2 (Q96LB3-2) which is 372 amino acid in length with predicted molecular weight of 42 kDa. Both protein isoforms are identical in the first 351 amino acids including the first 40 amino acids corresponding to exon 2. As there are methionine residues at positions 41, 80 and 121, by *in silico* prediction of using them as translation start codons downstream to the deletion, this would yield protein products of an estimated molecular weight of 65 kDa, 61 kDa and 56 kDa. This could explain the presence of three truncated protein products in the patients' samples, however it is difficult to say whether the three patient bands are simply a result of more non-specific protein instability. (**Figure 6-13**)

The IFT74 protein has a basic region extending between amino acids 1 and 90 and a coiled coil domain that extends between amino acids 98 and 482. The exon 2 deletion is expected to affect the basic domain only, if the translation of the mutant protein could start at the downstream AUG-coding methionine residues at position 41 (deleting inframe Δ40 amino acids) or at

position 80 (deleting inframe ∆79 amino acids) of the full length IFT74 protein. If the translation machinery might also use the further downstream inframe AUG-coding methionine residues at position 121 or at an even more downstream positions (∆≥120 amino acids, this would disrupt a part of the coiled coil domain). (**Figure 6-14**)

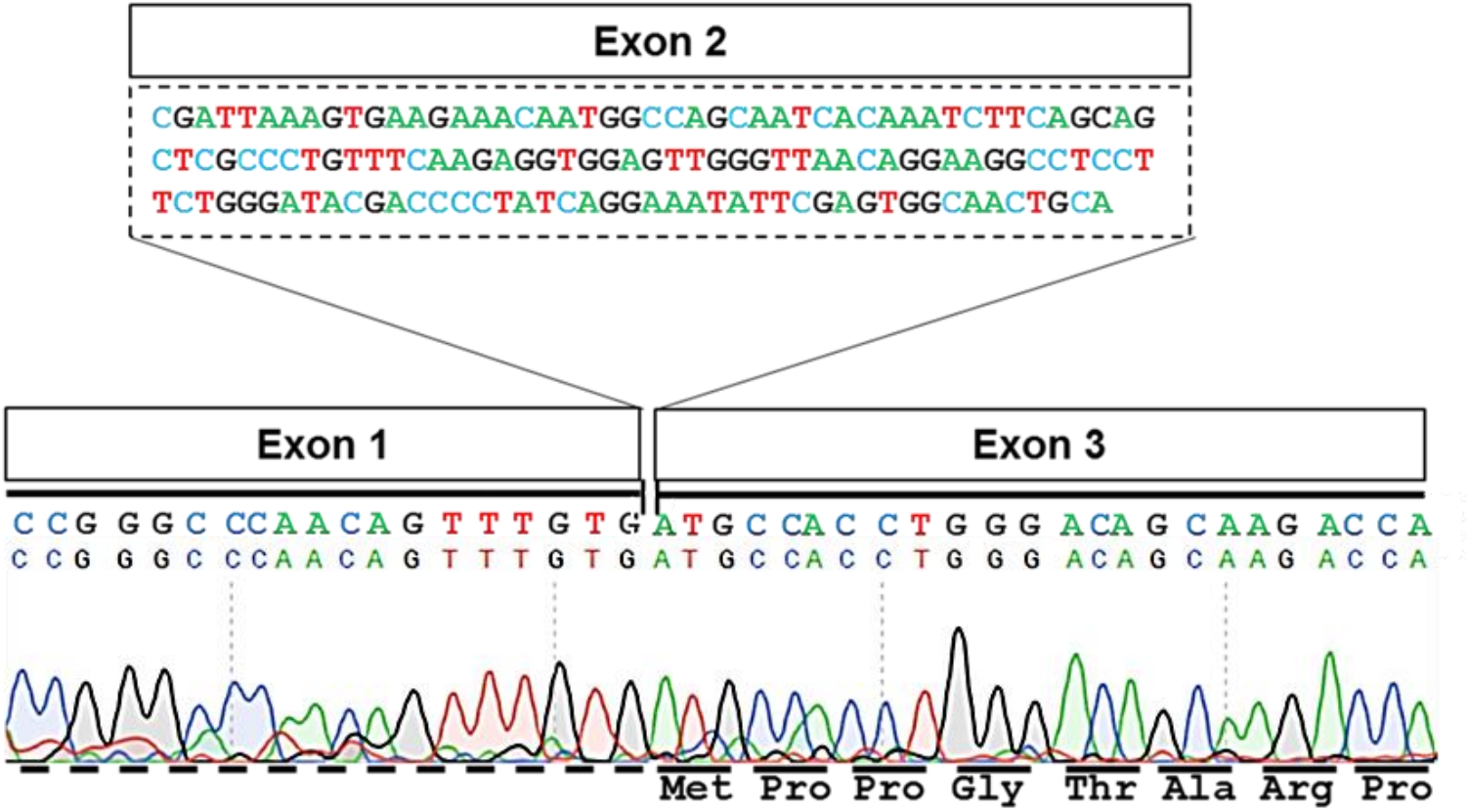


Figure 6-10 The exon 2-spanning genomic deletion only deletes exon 2 from the IFT74 mRNA

Sanger sequencing of the cDNA of the *IFT74* exon 2 deletion patient using primers extending from the 5' UTR and exon 3 of *IFT74* shows a complete deletion of exon 2 with preserved exon 3, resulting in an inframe deletion.

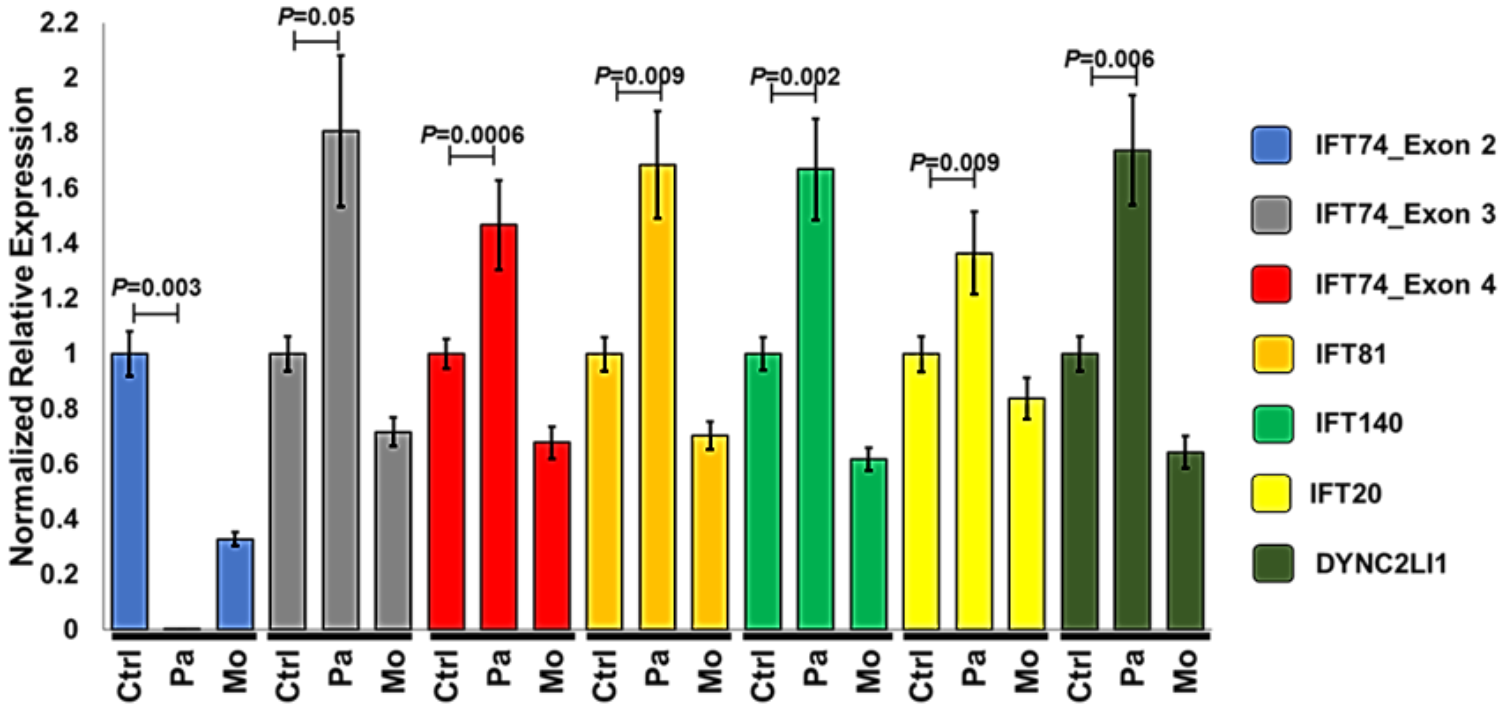


Figure 6-11 Transcriptional upregulation of IFT74 downstream to the deletion and other IFT components

qRT-PCR was used to assess the expression levels of *IFT74* transcripts and other IFT components (IFT81, IFT20, IFT140 and DYNC2LI1) in nasal brushing biopsies obtained from the patient, mother and a healthy control, showing elevated expression levels of all genes quantified and normalized to *GAPDH* quantification in the patient's sample, but no amplification when using primers in *IFT74* exon 2.Ctrl; control, Pa; patient and Mo; mother. Error bars indicate SEM (Standard error of the mean) each PCR reaction was done in technical triplicates.

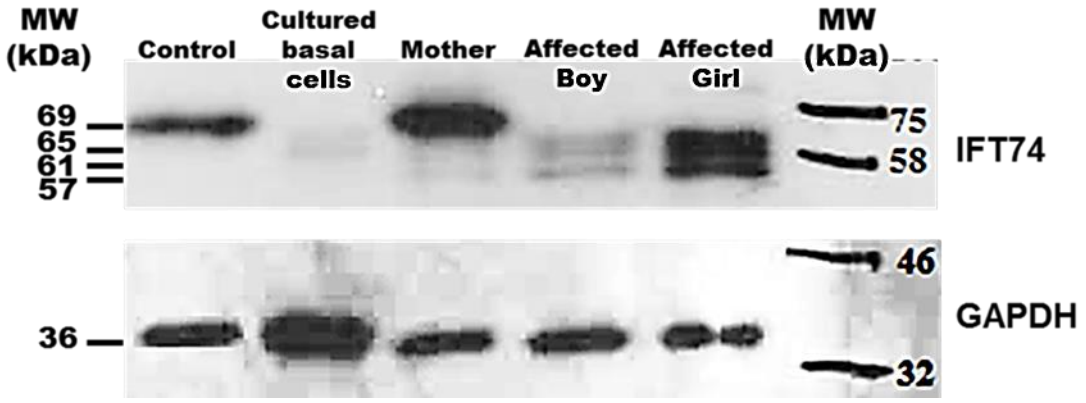


Figure 6-12 Truncated IFT74 protein products present in the family G140 patients but not in the control

Western blotting using specific anti-IFT74 and anti-GAPDH antibodies on the protein lysates of nasal brushing biopsies from the patients, mother and a healthy control in addition to ALI-cultured basal epithelial cells (still non-ciliated) showed a full length wild type IFT74 protein band at 69 kDa in the healthy control and the mother but absent from the patients' samples. Patients have three shorter bands representing truncated protein products at approximately (65, 61 and 57 kDa).

>sp|Q96LB3|IFT74_HUMAN Intraflagellar transport protein 74

1 41
MASNHKSSAARPVSRGGVGLTGRPPSGIRPLSGNIRVATAMPPGTARPGSRGCPIGTGGV
80
LSSQIKVAHRPVTQQGLTGMKTGTKGPQRQILDKSYYLGLLRSKISELTTEVNKLQKGIE
121
MYNQENSVYLSYEKRAETLAVEIKELQGQLADYNMLVDKLNTNTEMEEVMNDYNMLKAQN
DRETQSLDVIFTERQAKEKQIRSVEEEIEQEKQATDDIIKNMSFENQVKYLEMKTTNEKL
LQELDTLQQQLDSQNMKKESLEAEIAHSQVKQEAVLLHEKLYELESHRDQMIAEDKSIGS
PMEEREKLLKQIKDDNQEIASMERQLTDTKEKINQFIEEIRQLDMDLEEHQGEMNQKYKE
LKKREEHMDTFIETFEETKNQELKRKAQIEANIVALLEHCSRNINRIEQISSITNQELKM
MQDDLNFKSTEVQKSQSTAQNLTSDIQRLQLDLQKMELLESKMTEEQHSLKSKIKQMTTD
LEIYNDLPALKSSGEEKIKKLHQERMILSTHRNAFKKIMEKQNIEYEALKTQLQENETHS
QLTNLERKWQHLEQNNFAMKEFIATKSQESDYQPIKKNVTKQIAEYNKTIVDALHSTSGN

Figure 6-13 Amino acid sequence of full length IFT74 protein

The IFT74 canonical protein isoform is comprised of 600 amino acids (UniProtKB database). Exon 2 codes for the first 40 amino acids (deleted amino acid sequence coloured red in bold underlined). Possible inframe downstream methionine residues used as alternative translation starts in *IFT74* exon 2 deletion patients, are highlighted in green.

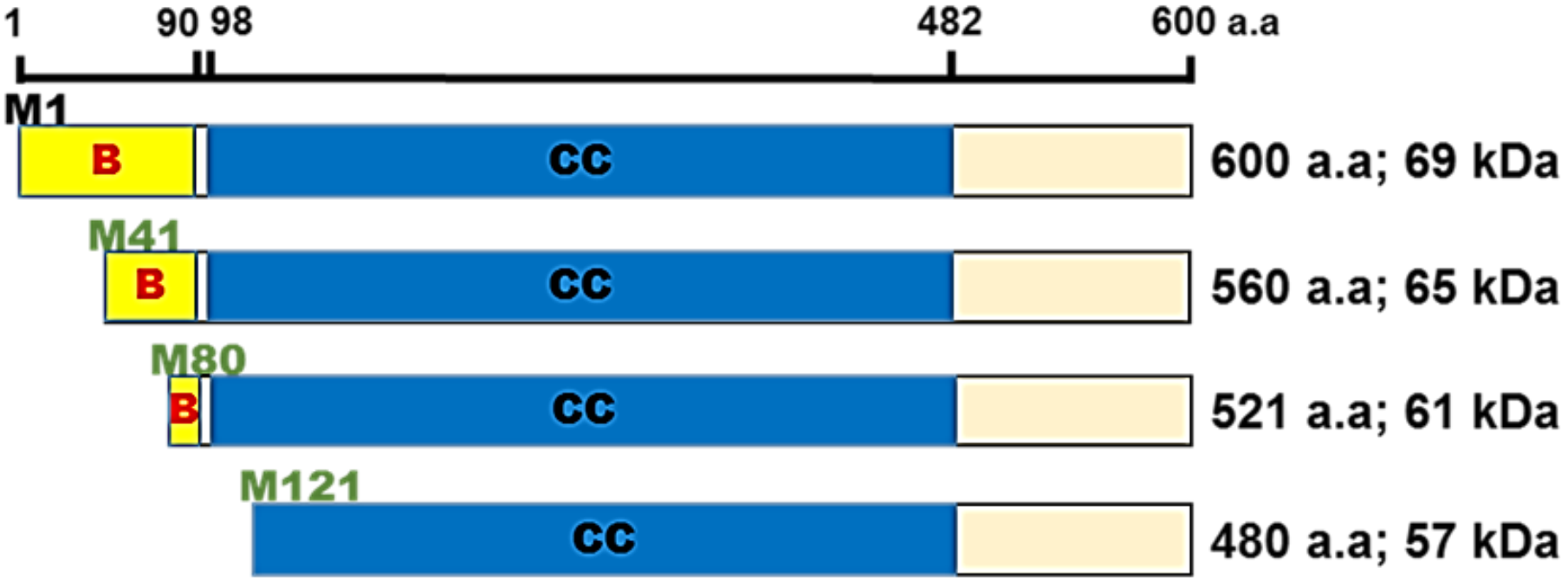


Figure 6-14 IFT74 protein domains and possible consequences of the N-terminus deletion

Full length IFT74 protein has a basic region (1-90 a.a) and a coiled coil domain (98-482 a.a) with a molecular weight of 69 kDa. In case of exon 2 deletion and translation starts at the next downstream inframe methionine residues at positions 41 or 80, this would only affect the basic region. If the translation starts at methionine 121 or other downstream positions, it will disrupt a part of the N-terminal coiled coil domain.

6.2.4 The IFT74 N-terminus is essential for the stability of the ciliary IFT-B complex

The effect of exon 2 deletion on IFT74 interactions was studied using affinity purification coupled with mass spectrometric analysis. This was done during a training placement at Prof. Marius Ueffing's proteomics lab (University of Tubingen, Germany) under the supervision of Dr Karsten Boldt. HEK293 cells were transfected with FLAG-tagged IFT74 constructs on the C-terminus (full-length IFT74, exon 2 deletion, ∆79 a.a, ∆120 a.a) and a negative control (empty FLAG-plasmid) in 6 replicate experiments. The FLAG-tagged IFT74 cDNA constructs were designed to represent the 3 putative truncated protein products found in patient lysates from nasal brushing samples. This was followed by affinity purification of FLAG fusion proteins then mass spectrometry and label free quantification (LFQ) of the eluates.

To study the consequences of IFT74 N-terminus truncations, a two sample t-test was used to analyse the statistical difference between the logarithms (log2) of LFQ values between different groups of experiments. Using the (log2) of LFQ values led to a Gaussian distribution of the data allowing imputation of the missing values within the normal distribution (width = 0.3 and down shift = 1.8) assuming that these missing data are close to the limit of detection.(362) The false discovery rate (FDR) was set at 0.05 (a test result below 0.05 is considered significant). S0 (Artificial within groups variance) was set at 0, as we performed 6 replicates for each construct. The main determinant of the significance statistics is based on *p*-values.(363) (**Table 6-1**)

A comparison of the LFQ values of the pulled down proteins in the full-length IFT74-construct against the empty control experiments was performed to determine the significantly and robustly enriched proteins (i.e. potential IFT74 interactors), for further analysis between the full length and various mutant constructs. From the 41 IFT74 interactors which are all listed in IntAct database (open source database for molecular interaction data), protein components of the IFT-B1 complex (IFT74, IFT81, IFT22, IFT52, IFT27,

IFT46, IFT88, IFT80, TTC26, TTC30A) were found to be highly enriched in the samples where full length of IFT74 was transfected, with some enrichment of components of IFT-B2 sub-complex (IFT172, IFT80, IFT57, IFT20), compared to the empty vector transfection. (**Figure 6-15 & 6-16**)

Further analysis of LFQ values of the enriched IFT-B protein components in the 3 mutant constructs compared to the full length IFT74 constructs revealed lower levels of IFT74 in two of the mutant constructs, exon 2 deletion and ∆79 a.a compared to the full-length IFT74 and ∆120 a.a constructs. This indicates possible instability of the exon 2 deletion and ∆79 a.a mutant constructs and this consequently may explain the low levels of the interactors found in these two experiments compared to the full length one, which therefore may not be secondary to real disturbed interactions of IFT74. Rather, this may be secondary to a disturbed stoichiometry due to low levels of mutant IFT74. This was confirmed by Western blotting of the protein lysate of different experiments and using antibodies against IFT74, FLAG and GAPDH, showing lower amounts of IFT74 and its FLAG tag, in the exon 2 deletion and ∆79 a.a transfected samples. Two bands of IFT74 were also found in these experiments, one of the expected size and another smaller band, indicating possible experimental degradation products. It was also noted that IFT74 was present in higher quantities in the ∆79 a.a-construct transfected samples compared to those transfected with the exon 2 deletion-construct, suggesting that the exon 2 deletion-construct is the least stable one. When LFQ values were compared between experiments of the full length protein and ∆120 a.a-construct disturbing a part of the coiled coil domain which is responsible for IFT74 protein-protein interactions, the levels of IFT74 were found to be stable across all samples of both constructs. However, many IFT proteins were found to be reduced in the mutant samples indicating possible disturbed interactions. (**Figure 6-17**)

The results from proteomics analysis were confirmed by western blotting. IFT-B complex proteins (IFT81, IFT88 and IFT46) could not be pulled down in the IFT74 exon 2 deletion-construct, but they were pulled down with the IFT74 ∆79 a.a-construct transfected samples. On the other hand, these

proteins were extensively reduced to nearly absent in the IFT74 Δ120 a.a-construct transfected samples. (**Figure 6-18**)

Table 6-1 List of IFT74 protein interactors identified by mass spectrometry and statistical significance of WT vs mutant constructs

Protein IDs	Gene names	Mol. weight [kDa]	Unique peptides	EV vs WT (*p* value)	WT vs Exon 2 Del (*p* value)	WT vs Del 79 a.a (p value)	WT vs Del 120 a.a (p value)
IFT-B1 sub-complex proteins							
A0AVF1	***TTC26***	64.177	7	1.72E-07	0.00041583	0.0483214	0.00352772
Q13099	***IFT88***	94.269	7	8.34E-06	2.06773E-06	0.00343335	8.34854E-05
Q86WT1	***TTC30A***	76.135	2	0.003439	0.0186676	0.506557	0.222272
Q9Y366	***IFT52***	49.706	11	3.34E-08	2.06789E-05	0.0995608	0.0112149
Q8WYA0	***IFT81***	79.745	29	2.09E-08	0.00258941	0.241107	0.0548403
Q96LB3	***IFT74***	69.239	84	5.12E-10	4.94658E-08	0.000694129	0.299266
Q9BW83	***IFT27***	20.48	8	4.01E-07	5.14124E-07	0.0484945	0.0290188
Q9H7X7	***IFT22***	20.835	7	7.43E-10	4.43488E-07	0.00960606	0.00300279
Q9NQC8	***IFT46***	34.285	8	1.89E-06	1.03579E-05	0.00913516	0.00508424
IFT-B2 sub-complex proteins							
Q9NWB7	***IFT57***	49.108	5	0.000295	0.0123485	0.0935496	0.11225
Q9P2H3	***IFT80***	88.034	7	0.00045	0.047756	0.0402427	0.0020835
Q9UG01	***IFT172***	197.57	20	2.76E-08	4.88902E-06	0.0138158	5.84285E-07
Q8IY31	***IFT20***	15.28	4	0.000415	0.00708762	0.0379066	0.0196039

Mol. Weight: Molecular Weight, EV: Empty Vector, WT: Wild Type, Del: Deletion, a.a: Amino acids

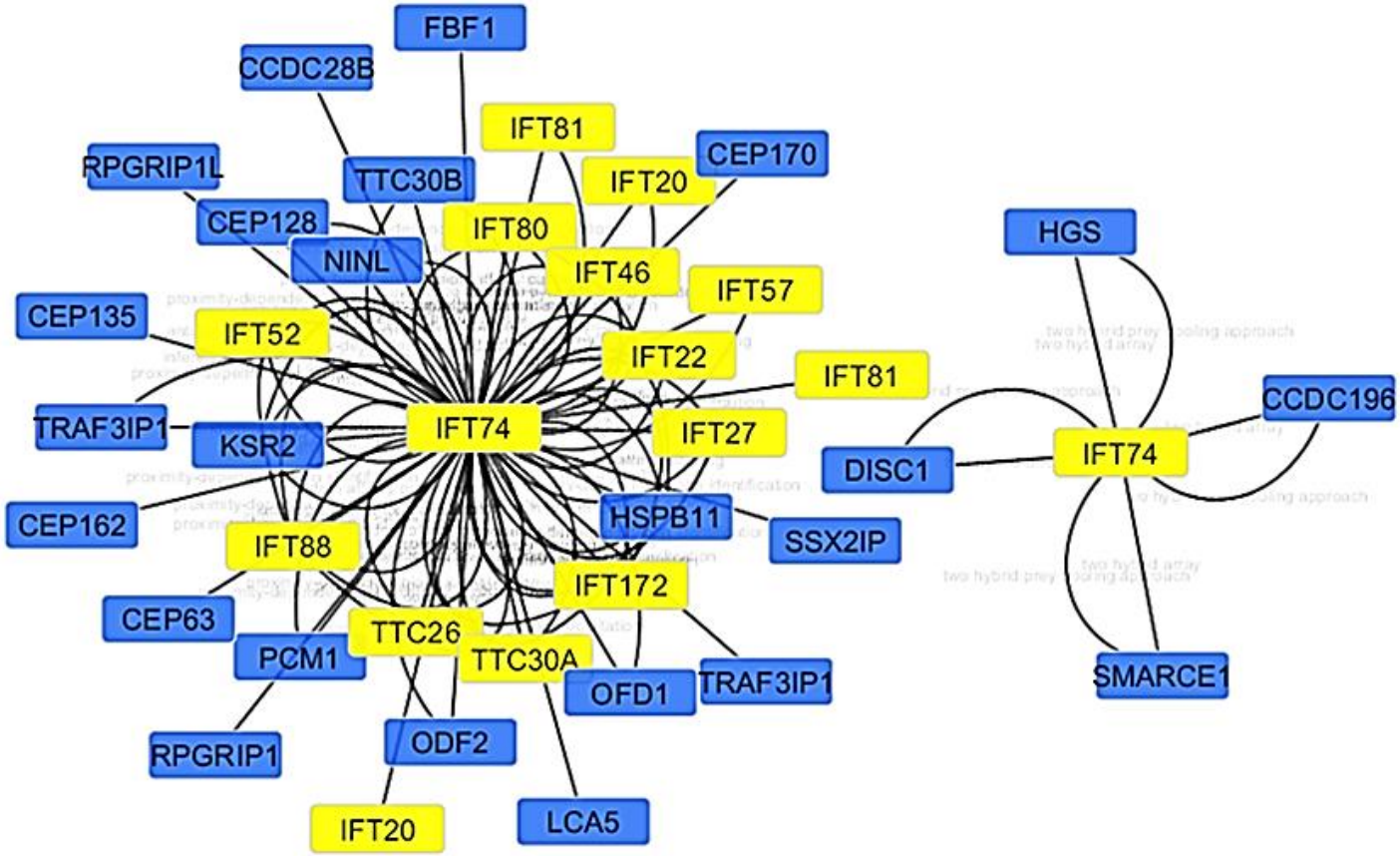


Figure 6-15 IFT74 protein interaction network

Cytoscape network of all 41 IFT74 interactions listed in the IntAct database. Enriched potential interactors identified in the full length IFT74 transfection experiments compared to the empty vector are highlighted in yellow.

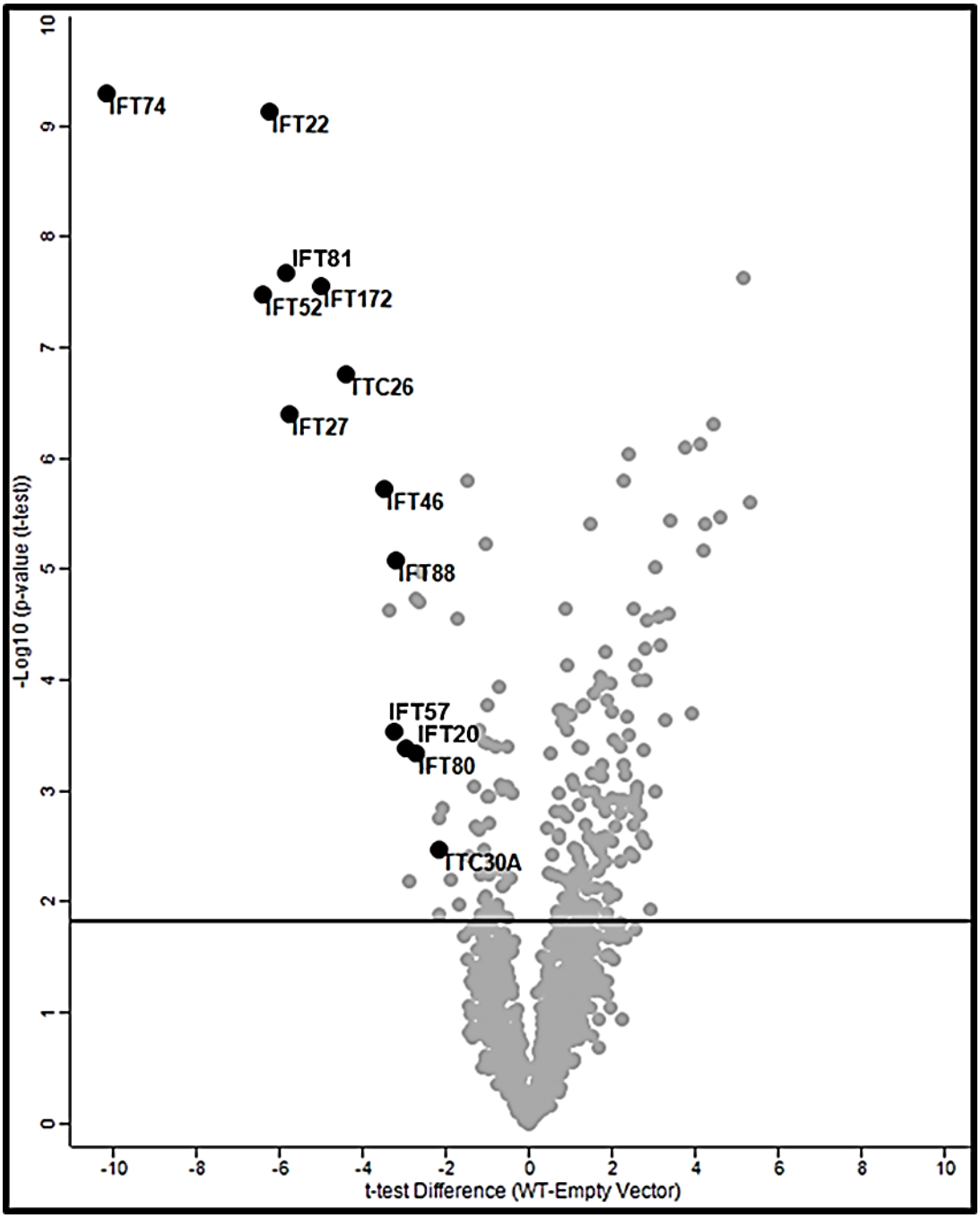


Figure 6-16 Proteins enriched with full length IFT74 transfection compared to empty vector transfection

Protein components of the IFT-B complex that are enriched for interactions with full length IFT74 are highlighted in black. Statistical significance was determined by permutation-based FDR-corrected t-test (Threshold: *P* = 0.05 and *S0* = 0). Horizontal line indicates significance cut off value.

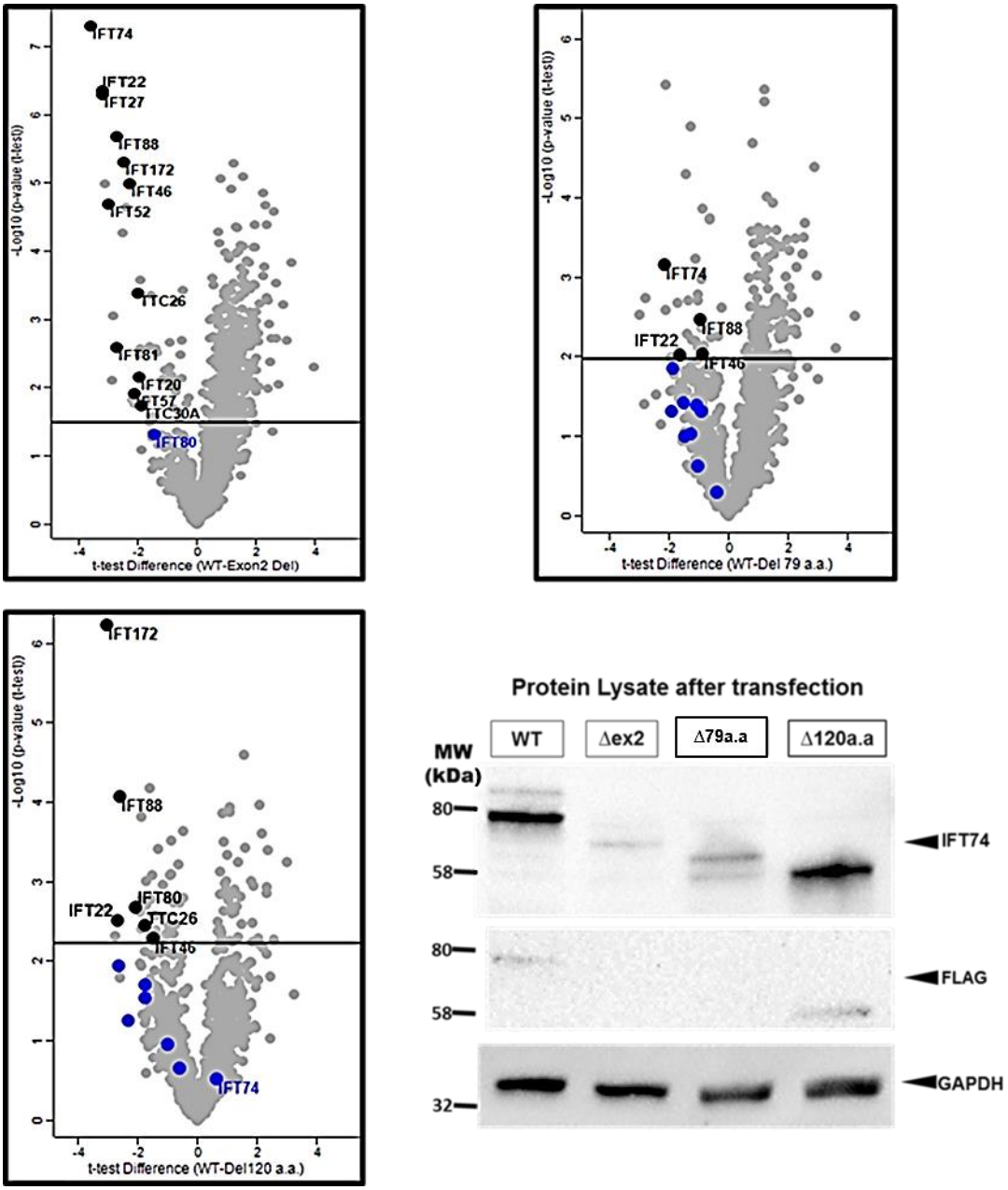


Figure 6-17 Potential disturbed interactions of IFT74 identified by label free quantitative proteomics

IFT74 expression was found to be reduced in samples transfected with the exon 2 deletion and Δ79 a.a mutant constructs, compared to samples transfected with full length IFT74 construct indicating possible instability and degradation. This was confirmed by Western blotting and consequently may explain the low levels of the interactors found in these two experiments (more disruption was noted with the exon 2 deletion than Δ79 a.a mutant construct).The Δ120 a.a mutant construct was found to be more stable but resulted in more disturbed interactions with other IFT-B complex proteins. Horizontal lines indicate significance cut off value. Black dots are proteins significantly changed, while blue dots are proteins not significantly changed across experiments. Statistical significance was determined by permutation-based FDR-corrected t-test (Threshold: P = 0.05 and S0 = 0).

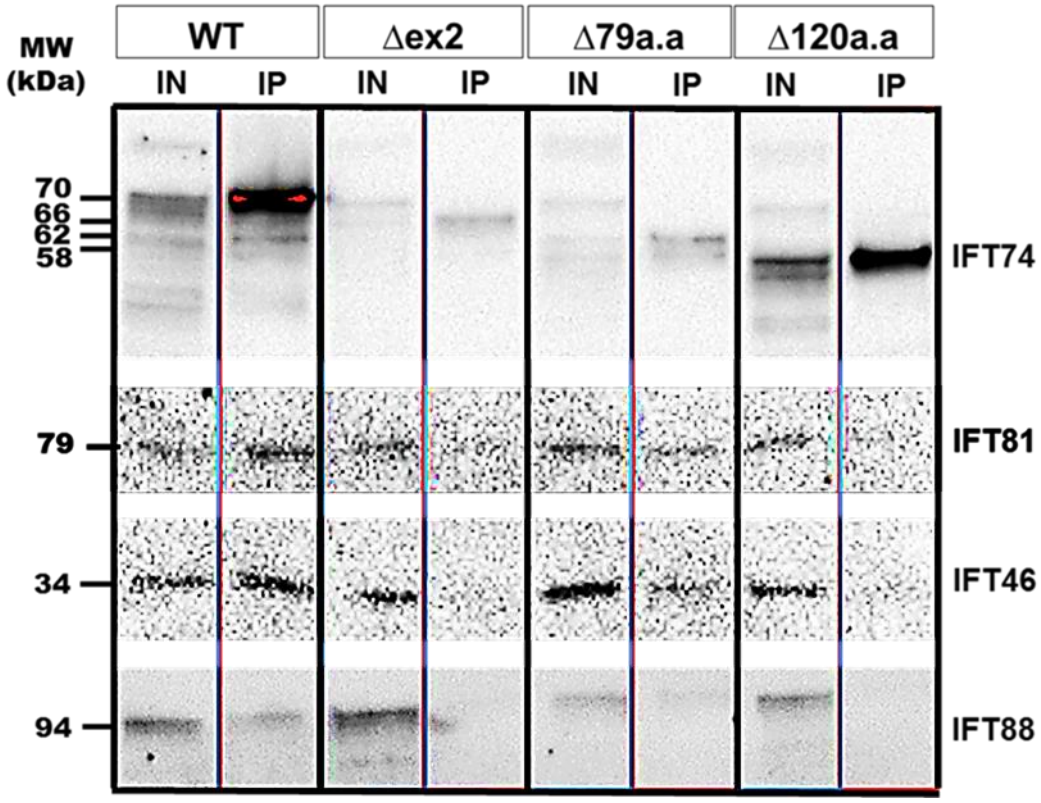


Figure 6-18 Western blotting analysis of IFT74 affinity purification experiments

The quantities of IFT74 pulled down after the exon 2 deletion and ∆79 a.a constructs transfection were much lower than those of the full length and ∆120 a.a constructs. None IFT88 and IFT46 proteins were pulled down with the reduced amount of IFT74 in the exon 2 deletion experiments. Despite the reduced level of ∆79 a.a IFT74, all other IFTs were pulled down in this experiment indicating no disturbed interactions despite less stable IFT74. With the more stable ∆120 a.a construct, no IFT88 and IFT46 interactors were pulled down, indicating real disturbed interactions with this construct. Very faint IFT81 band was noted in pull down experiments of the exon 2 deletion and ∆120 a.a IFT74 protein. Transfection and pull down followed by western blotting experiments were done 6 times with consistent results in all experiments.

6.2.5 *WDR19/IFT144* mutation identified in patients with Sensenbrenner syndrome and respiratory problems

Following on again from NGS gene panel PCD patient screening, a homozygous missense mutation (**c.3722C>T, p.Pro1241Leu**) was identified in the *WDR19* gene in two affected siblings from a consanguineous Pakistani family with a clinical genetic diagnosis of Sensenbrenner syndrome associated with chronic respiratory symptoms consistent with PCD. The mutation was confirmed to be homozygous in the two affected siblings and the parents were found to be heterozygous carriers by Sanger sequencing. The mutation was present in exon 34 and is predicted to change the proline at position 1241 to leucine near the C-terminus of the encoded WDR19 (also called IFT144) protein. (**Figure 6-18**)

This variant (dbSNP: rs576113399) was found in in the ExAc database with a MAF of 0.00079 (95/120756) and in the 1000G control database T=0.001 (7/5008). There are currently 2 submissions of this variant in the ClinVar database accessed in 01/09/2018 (variant ID 348755), where it is interpreted as a variant of uncertain significance linked to Jeune thoracic dysplasia and Sensenbrenner syndrome. This mutation was predicted to be possibly damaging by Polyphen-2 with a score of 0.662 (sensitivity: 0.79; specificity: 0.84). It was predicted as a disease causing variant by Mutation Taster software. Using the SIFT tool, it was predicted to have a damaging effect on the protein. It has a CADD score of 27.5.

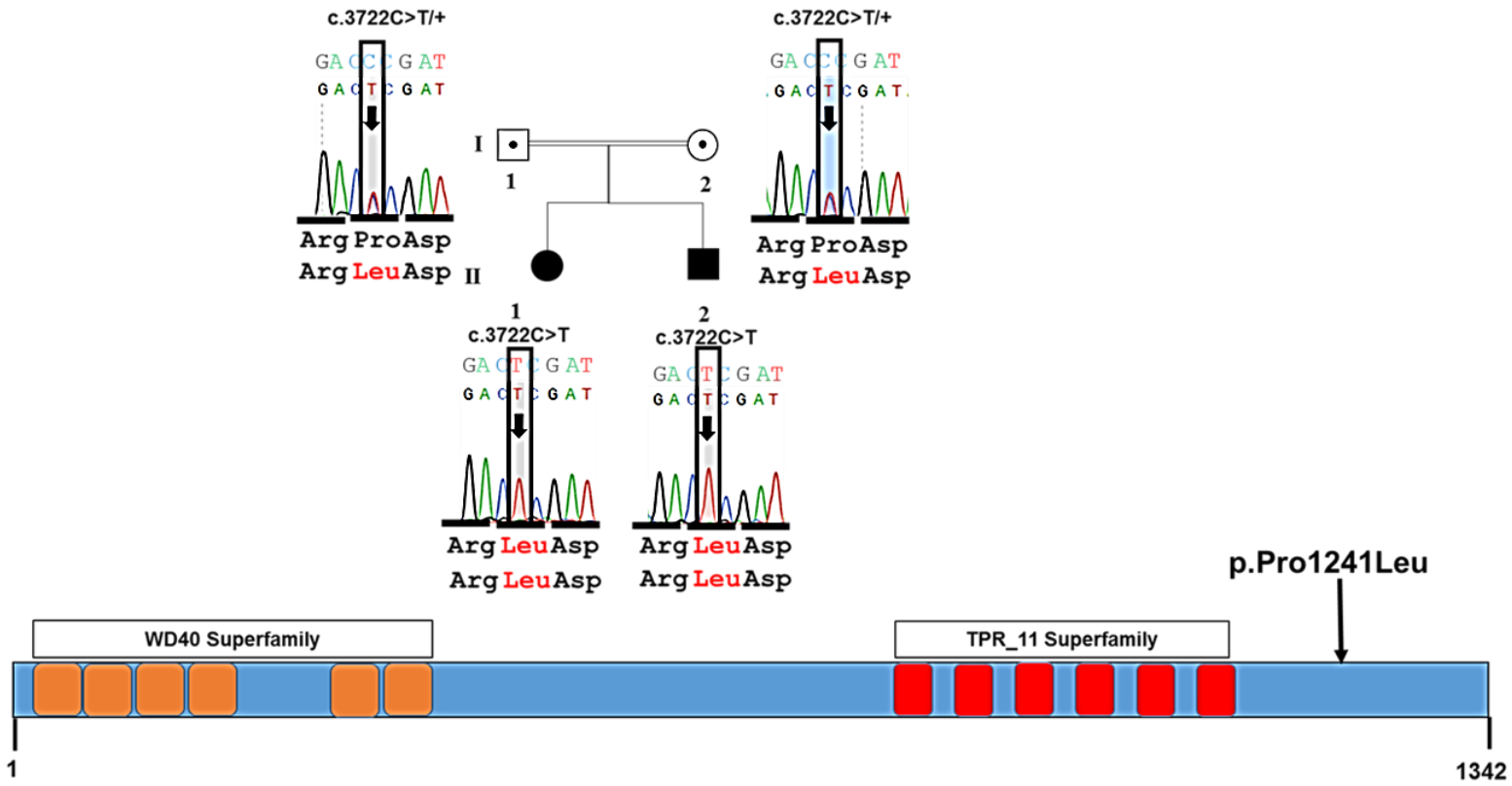


Figure 6-19 Pedigree of affected family (G130) carrying putative WDR19 mutations

Sanger sequencing confirmed the presence of a missense mutation (c.3722C>T, p.Pro1241Leu) in a homozygous state in the two affected siblings, for which both parents are heterozygous carriers. This mutation changes the proline at position 1241 of the predicted protein into a leucine.

6.2.6 Defective motile cilia in patients with *WDR19* mutations

Ultrastructural TEM and HSVM analysis of the *WDR19* patients' respiratory cilia were done at Royal Brompton Hospital, London. Their nasal brushing biopsies under light microscope showed patchy sparse cilia and most of the epithelial cells did not have cilia. Occasional patches of long cilia were also noted. The cilia beating pattern was dyskinetic with a low cilia beating frequency. TEM analysis showed bulbous tips affecting multiple elongated respiratory motile cilia (data not shown).

Immunofluorescence staining of WDR19 in patients' cilia showed up predominantly short sparse cilia with occasional long cilia that displayed reduced levels of WDR19 along the whole length of the patients' cilia compared to healthy controls. (**Figure 6-19**) A recent report showed a synergistic effect of compound mutations in the *WDR19* and *TEKT1* genes on the defective cilia phenotype. TEKT1, one of the tektin family of proteins that are co-assembled with tubulins to form the ciliary and flagellar microtubules, was absent from the motile cilia in patients carrying mutations in both genes.(364) No TEKT1 mutations were found in the patient screened using the gene panel. Staining for TEKT1 was pursued in the current study showing a nearly absent TEKT1 in the cilia of patients with *WDR19* mutations compared to its normal localization along the whole length of the cilia in healthy control.

Localization of the ciliary protein ARL13B, anterograde IFT-B complex component IFT74, and cytoplasmic dynein-2 component DYNC2LI1 were also studied in the *WDR19* patients' respiratory cilia compared to healthy controls. Interestingly, all the studied proteins were extensively reduced to nearly absent in the patients' cilia. (**Figure 6-20 & Figure 6-21**)

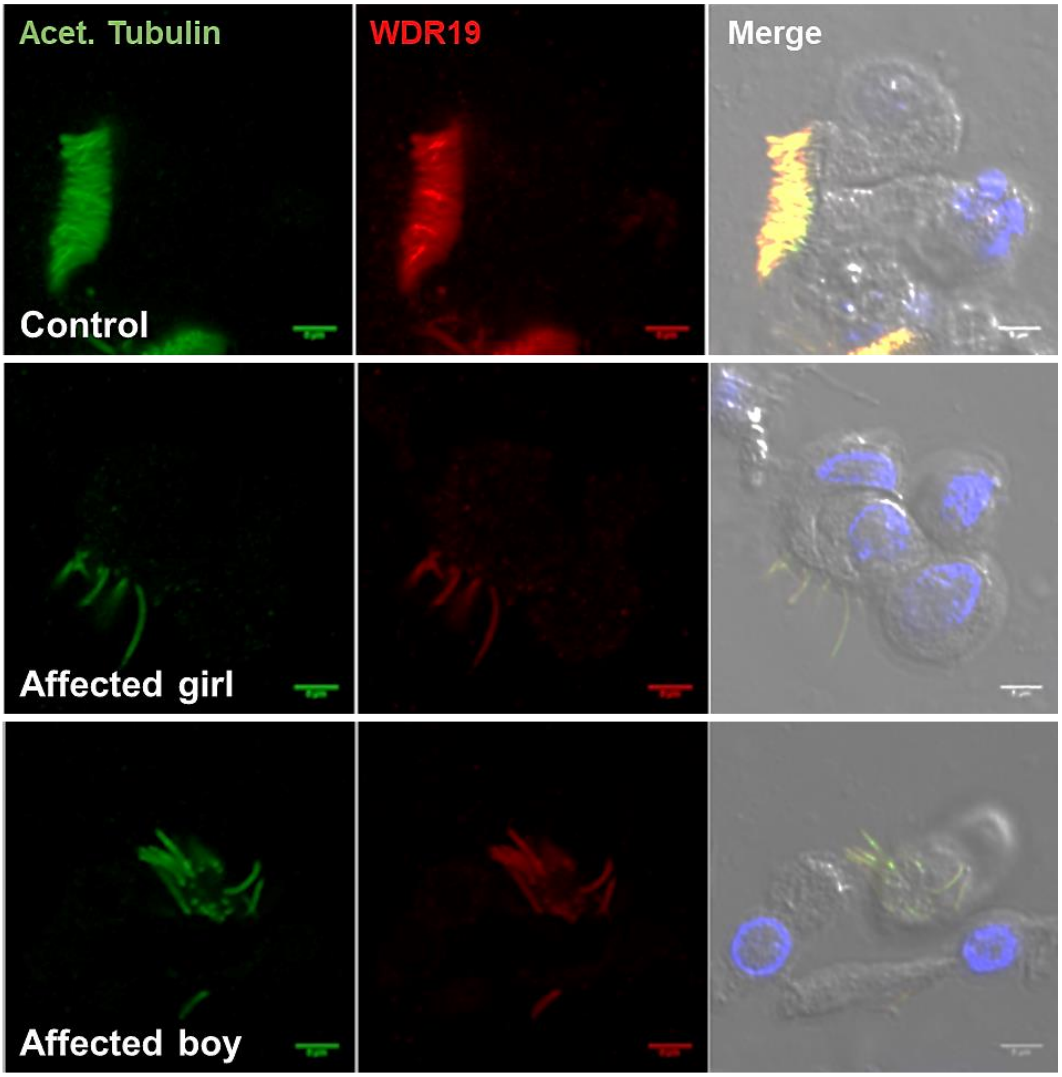


Figure 6-20 Sparse with occasionally long respiratory cilia in patients with WDR19 mutations from family G130

Immunostaining of WDR19 showed few cilia in the patients' cells with occasional presence of long cilia and reduced levels of WDR19 (red). Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 5 µm.

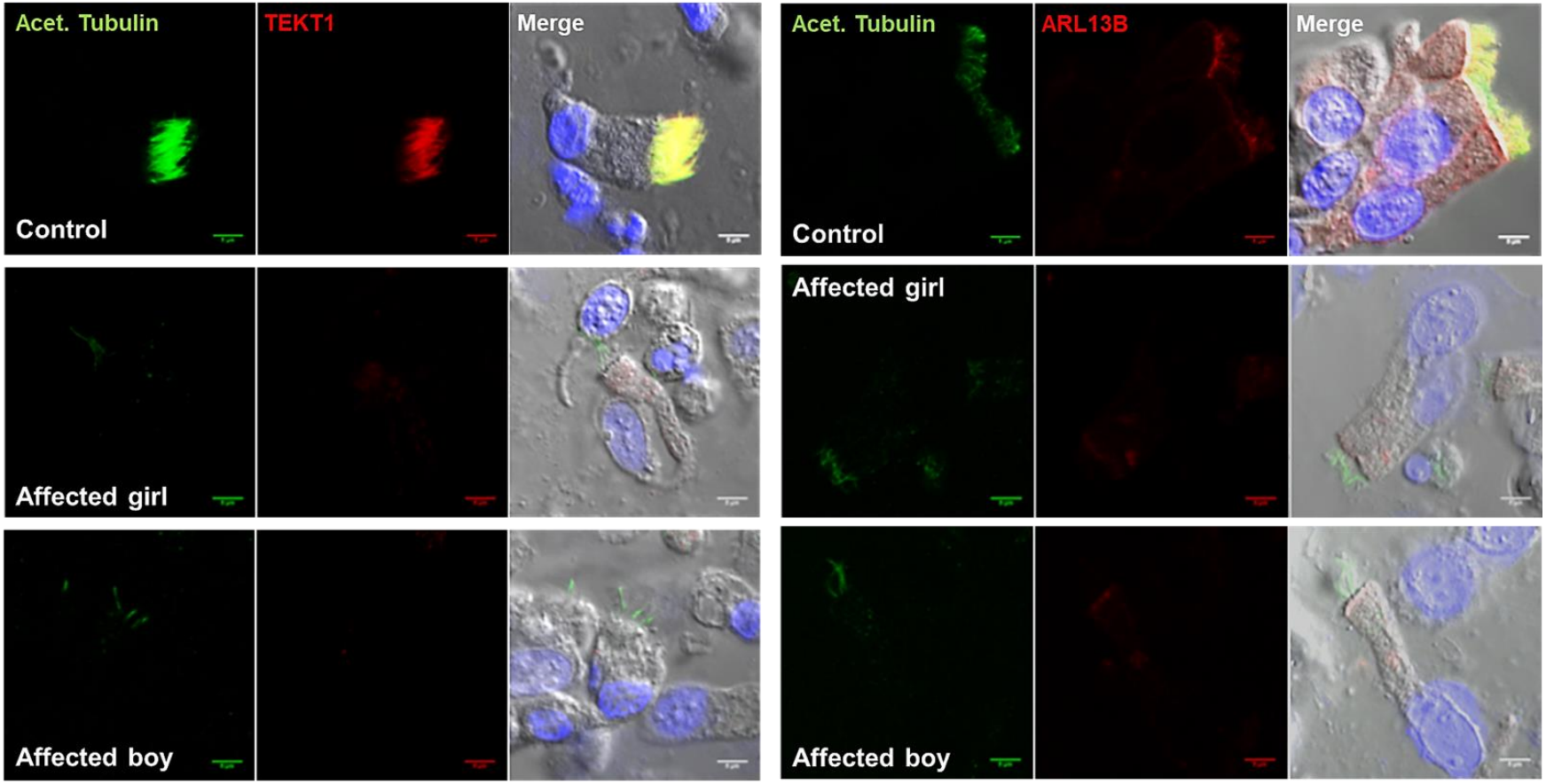


Figure 6-21 Loss of TEKT1 and ARL13B from the motile cilia of patients from family G130 with WDR19 mutations

Immunofluorescence staining using specific anti-TEKT1 and ARL13B antibodies shows loss of TEKT1 (red) in the left panel and ARL13B (red) in the right panel from the cilia of the patients compared to its normal localization along the cilia from the healthy control. Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 5 µm.

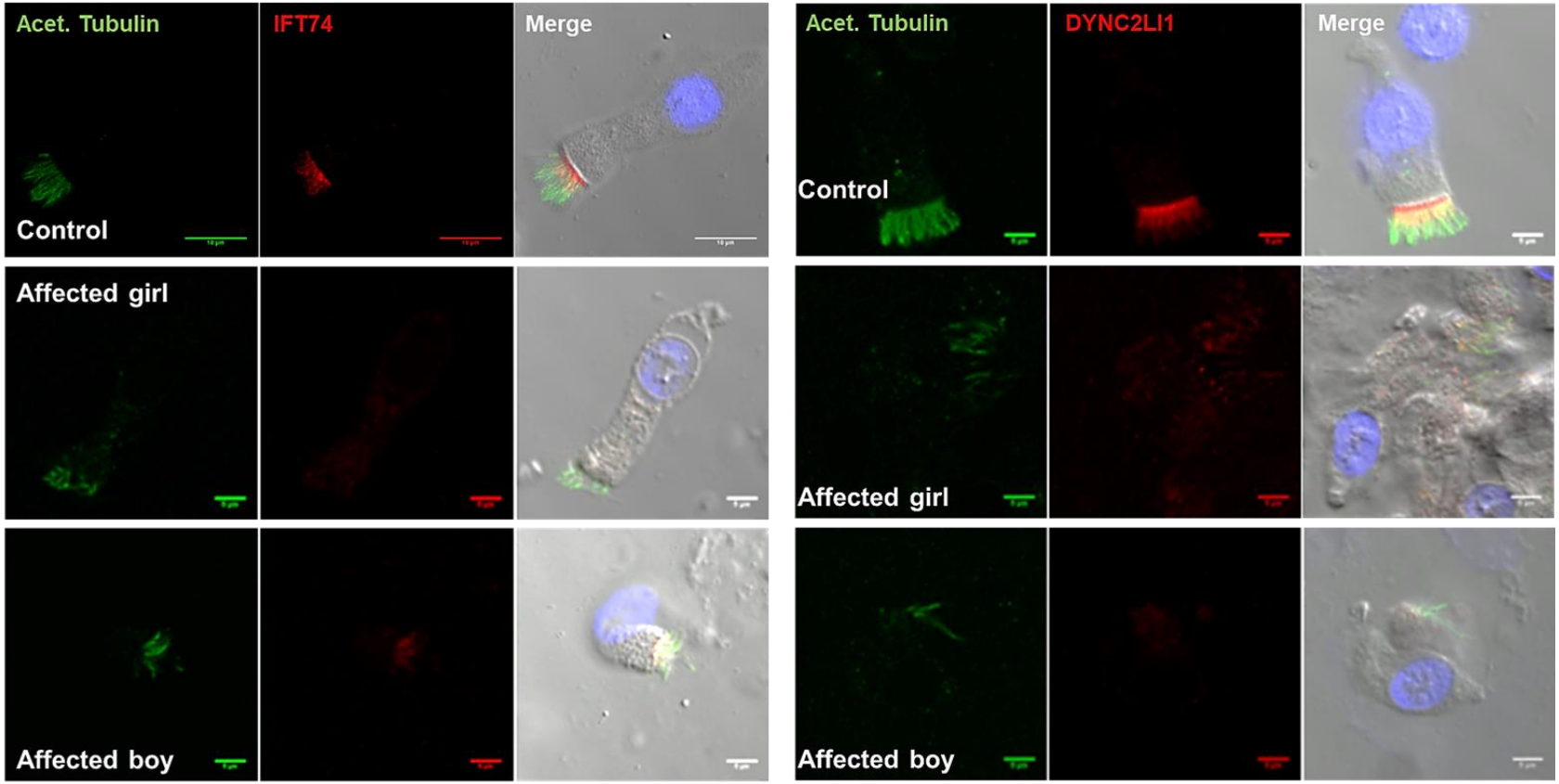


Figure 6-22 Absent IFT74 and DYNC2LI1 in the motile cilia of patients from family G130 with WDR19 mutations

Immunofluorescence staining using specific anti-IFT74 and DYNC2LI1 antibodies shows significantly reduced protein levels/nearly loss of IFT74 (red) and DYNC2LI1 (red) from the cilia of the *WDR19* patients compared to its normal localization along the cilia from the healthy control. Acetylated alpha-tubulin (green) was used as a cilia marker. Scale bar 5 µm.

6.3 Discussion

A homozygous deletion of exon 2 of the *IFT74* gene was identified in a consanguineous family of Arabic origin by targeted sequencing coupled with ExomeDepth analysis to identify CNVs within the sequenced genes. The two affected siblings displayed symptoms indicating combined motile and primary ciliopathy. TEM analysis showed short sparse motile respiratory cilia and this was confirmed by immunofluorescence staining studies.

Symptoms of motile ciliary dyskinesia have been reported before with primary ciliopathy syndromes. *OFD1* mutations known to cause oral-facial-digital syndrome type 1 were previously found in patients with ciliary dyskinesia and situs inversus.(39, 365) Moreover, primary ciliary dyskinesia was found associated with x-linked retinitis pigmentosa in patients with *RPGR* mutations.(36, 37)

Whole genome sequencing identified the *IFT74* genomic deletion breakpoints and a follow-up Sanger sequencing of the deletion region revealed an insertion of 11 bp of unknown origin in the deleted region. RT-PCR confirmed a complete deletion of exon 2 with no effect on exon 1 (non-coding) and exon 3 (coding exon starting normally with AUG). Quantitative real time PCR showed that there was a transcriptional upregulation of *IFT74* downstream to the deletion. Moreover, other IFT genes were shown to be transcriptionally upregulated in the patient's sample compared to a healthy control. This may reflect a possible compensatory feedback mechanism to overcome the defective function of certain genes, a mechanism that has been reported before where a gene paralogue or a network of genes are upregulated in response to defective function of one gene resulting in a milder disease phenotype.(366, 367) Moreover, in *Chlamydomonas* mutant strains carrying an IFT74 N-terminus deletion, there is not a great effect on the IFT complex in general and it was shown that it could still maintain some tubulin transport function compensated through the presence of normal levels of IFT81 (the IFT74-partner protein), both together form a module for tubulin transport.(368, 369) These findings can support the speculation that the *IFT74*

exon 2 deletion is a hypomorphic allele and consequently the clinical picture of the patients is at the milder end of ciliopathy phenotypic spectrum.(26, 43, 333, 336)

Western blotting of protein lysates from nasal brushing biopsies from the patients, mother and a healthy control showed loss of the IFT74 wild type protein band from the patients' samples. There were 3 smaller bands identified in the patients' samples potentially representing inframe truncated IFT74 protein products, since deletion of exon 2 is inframe. These bands were predicted to be produced by using different translation start codons downstream to the deletion secondary to a possible translation leaky scanning mechanism.(370) The use of the first downstream AUG (at a.a 41) would affect the basic region of the IFT74 N-terminus which has been shown to be dispensable for the stability or the formation of the IFT-B complex. It was also previously shown that the IFT74 N-terminus interacts with the acidic tail of β-tubulin to increase the affinity of binding of IFT81 N-terminus to the globular domain of tubulin, for the purpose of IFT-mediated tubulin transport along the ciliary axoneme.(371) However, the use of any further downstream AUGs would affect the coiled coil domain through which IFT74 interacts with other IFT-B core proteins e.g. interactions with IFT81 occurs through the coiled coil domains of both proteins. (341, 371, 372)

The consequences of exon 2 deletion on IFT74 protein interactions were studied by affinity purification assay coupled with mass spectrometry and label free quantification analysis, using plasmid constructs for full length IFT74, exon 2 deletion, ∆79 a.a and ∆120 a.a IFT74 which presumably correspond to the size of the aberrant bands found in the patients' protein lysates. Surprisingly, despite the transfection of equal amounts (and quality) of all plasmids, the quantities of IFT74 pulled down were much less with the exon 2 deletion and ∆79 a.a-constructs compared to the pulled down IFT74 with the full length IFT74 and ∆120 a.a-constructs. This suggested a possible degradation process of the first two constructs (exon 2 deletion, ∆79 a.a) that was also shown by Western blotting. Comparing the quantities of the pulled down proteins between the full length IFT74 and ∆120 a.a constructs showed equal amounts of IFT74 indicating stability of the construct. There was a clear

disruption of IFT74 interactions with many other IFT proteins identified in the ∆120 a.a-construct experiment, which was expected as it affects the coiled coil domain through which IFT74 interacts with its partner.(341, 371, 372) Any protein product arising from the more downstream AUGs will affect the coiled coil region possibly even worse and subsequently will likely disrupt more IFT74 interactions.

The two most common protein modifications that can affect protein homeostasis and possible degradation are N-terminal methionine cleavage and N-terminal acetylation. These two processes can occur either co- or post-translationally.(373) This phenomenon is known as the N- end rule where the protein half-life depends on its N-terminal residues sequence and the modifications of its N-terminus.(374-378) The presence of proline or threonine at position P2' (the third amino acid position "M1, P1', P2'…) will hinder the N-terminal methionine cleavage resulting into protein degradation.(379-381) This is a possible explanation for the relatively greater instability and subsequent degradation of IFT74 exon 2 deletion and ∆79 a.a both in patients and over-expression experiments. As with the inframe exon 2 deletion, the translation potentially would move downstream to start at the next AUG which is located right at the start of exon 3 and methionine at this position 41 of the full length protein is followed by proline in the second and third residues preventing methionine cleavage and rendering this peptide readily available for degradation (**Figure 6-12**). In a similar way the presence of threonine at position 3 (P2') if the translation might have started at the next inframe downstream AUG/methionine which is located at position 80 of the wild type IFT74 protein and is used in the ∆79 a.a construct, this would also lead to inefficient cleavage of methionine (**Figure 6-12**). This is not the case when translation would move further downstream to start with AUG/methionine at position 121.

A missense mutation in the *WDR19 (IFT144)* gene, encoding a component of the retrograde IFT-A complex, was identified in a family with two affected siblings diagnosed with Sensenbrenner syndrome and respiratory manifestations. The mutation affected a proline residue at position 1241 of *WDR19* and was predicted to have a damaging effect on the protein. No other

sequence variants were prioritized in any other screened genes including *TEKT1*. In support of this finding that *WDR19* mutation can contribute to the patients' defective motile cilia phenotype, mutations in *WDR19* were recently reported in a patient with Mainzer-Saldino syndrome associated with respiratory infections and motile cilia dyskinesia. The patient also had mutations in the *TEKT1* gene and it was shown that defective TEKT1 contributes to the motile cilia phenotype.(364) Moreover, morpholino knockdown of *wdr19* in zebrafish embryos yielded defective motile cilia phenotypes (hydrocephalus, shortened curved body axis and small proportion with situs inversus) in addition to extensive loss of cilia in some tissues.(364)

In *Wdr19* mutant mice, both weak mutant allele and a null allele lead to defective ciliogenesis at various degrees. Although the consequences of loss of WDR19 on motile cilia were not investigated, the mutations were shown to affect the ciliary anterograde intraflagellar transport and the localization of the ciliary membrane protein ARL13B in primary cilia in mice.(382) IFT88 was found to accumulate at the tips of *Wdr19* mutant mouse primary cilia, however in the current study IFT74 was found to be extremely reduced in the *WDR19* patients' motile cilia. Moreover, ARL13B and DYNC2LI1 were extensively reduced from the patients' motile cilia. The presence of bulbous tips amongst the patients' motile cilia is similar to the *Chlamydomonas* flagellar phenotype found in mutant strains with defective retrograde IFT.(383)

6.4 Summary

In summary, the presence of motile and primary cilia defects associated with mutations of IFT genes are not as infrequent as previously believed. Mutations in the IFT-B component, *IFT74* led to multiciliogenesis defects with short sparse respiratory cilia associated with skeletal and retinal abnormalities. Mutations in the IFT-A component, *WDR19/IFT144* also affect motile ciliogenesis leading to ciliary paucity with long motile cilia and bulbous tips associated with Sensenbrenner syndrome.

Chapter 7 General discussion and future perspectives

7.1 General discussion

Primary ciliary dyskinesia (PCD) is a rare recessive genetic disease, with a prevalence of about 1:10,000 individuals (53, 54, 282), which increases to as high as 1:2,265 within communities with a high rate of consanguineous marriages like the UK South Asian population.(44) PCD is caused by defective motile cilia present at the left-right organizer of the embryonic node and found along the epithelial lining of the respiratory system, fallopian tubes, and the ependymal lining of the brain ventricles. Motile cilia are also highly related to the sperm flagella.(44)

PCD patients can suffer from neonatal respiratory distress, progressive respiratory problems, bronchiectasis, laterality problems and adult infertility. Chronic wet cough is virtually present in all cases (sensitivity 0.93, specificity 0.15), as it is usually the main reason for referral for further PCD investigations.(78, 384) In a study conducted on 118 paediatric PCD patients from North America, chronic productive cough was reported in almost all cases (99%), however neonatal respiratory distress was only reported in 82% of the cohort. Laterality problems were observed in 55% of the participants.(385) In another study conducted on 168 PCD patients in Belgium, neonatal respiratory problems were reported in only 45% of the participants and laterality defects were found in 41%.(285) In an Australian study conducted on 81 PCD patients, 81% presented with recurrent cough, while only 57% had neonatal respiratory distress and in 46%, laterality problems were found.(64) Generally, PCD symptoms are variable and a typical clinical picture with all reported symptoms does not occur very often.(56-58) This variability of the clinical manifestations leads to confusion over diagnosis and usually a delay in referral for further investigations.

PCD as a rare genetic disease can be the sole disease affecting the patient or it can also co-exist with other rare disorders on very rare occasions. This will necessitate a modified diagnostic and management pathway where further testing and/or therapies might be needed. These rare conditions can occur due to mutations in one gene leading to complex defective motile and

non-motile ciliary manifestations or due to mutations in neighbouring genes. For instance, in patients carrying *RPGR* mutations, retinitis pigmentosa can be associated with PCD phenotypes. Thus, retinal examination is recommended in PCD patients carrying mutations in the *RPGR* gene.(37) Oral–facial–digital type 1 syndrome or Simpson–Golabi–Behmel syndrome type 2 are also associated with the primary ciliary dyskinesia phenotype due to mutations in the *OFD1* gene. These patients should be referred for further genetic consultation.(39) PCD has also been reported in patients with Cri du Chat syndrome due to a large deletion on chromosome 5p associated with a mutation in *DNAH5*.(386-388) All these examples highlight the importance of genetics in understanding the disease phenotype and subsequently will direct to the proper management.

Early diagnosis and management is essential for better respiratory disease prognosis in PCD. Although about 50% of PCD patients have situs abnormalities and neonatal distress, both recognisable conditions, only half of these patients were diagnosed in the first year of life.(389) It was shown that the age at diagnosis might have an impact on the prognosis of respiratory problems in PCD and that lung function can be relatively stable once antibiotic and airway clearance therapy are started.(390) Early diagnosis can help to halt the disease progression and reduce morbidity. It will allow proper management strategies for associated problems like hearing abnormalities and rhino-sinusitis, as treatment options for PCD patients might be different from those offered to the general people.(391)

There is no single standard test to confirm or exclude the diagnosis of PCD, as PCD can result from various defects in ciliogenesis or motile cilia structure, function or organization. A panel of multiple cilia functional tests is usually required.(75) There are many examples where a clinical phenotype consistent with PCD is accompanied with normal or inconclusive diagnostic results. For instance, patients carrying *DNAH11* mutations have classic clinical manifestations and low nasal nitric oxide levels. However, TEM analysis shows normal cilia ultrastructure and only subtle changes can be found in the cilia beating by HSVM.(161) Moreover, patients carrying *RSPH1* mutations show milder and later-onset clinical symptoms. TEM analysis of

RSPH1 patients' cilia shows subtle ultrastructural defects with only a slight change in cilia beating pattern observed by HSVM.(185, 278) In such cases, without genetic testing, the diagnosis can be easily missed even after extensive cilia functional investigations.

The current multi-test clinical diagnostic workup of PCD is expensive and technically demanding requiring well-trained personnel and sophisticated equipment. These resources are not widely available in all centres or countries, for example in the Palestinian and Egyptian clinical centres collaborating in the current study.(76, 77) The PCD diagnostic workflow is modifiable depending on the location as some tests many not be available and also there is variability in the available local expertise in the institution where PCD investigations are performed. Some tests may also not be suitable for certain age groups e.g. the measurement of nasal NO levels, which is not reliable in children less than 6 years old.(57) TEM and HSVM analysis require an invasive procedure to obtain a ciliary biopsy by nasal brushing.(392) On contrary, DNA can be easily extracted from saliva for genetic testing. DNA samples are also very stable and can be transported at room temperature to distant national laboratories or even abroad. This is not easy with other testing methods as the patient must attend to the centre where investigations will be performed.

Although about 70% of PCD cases can be diagnosed by TEM, it remains subjective and depends significantly on the presence of adequate expertise in sample processing and interpretation.(80) As mentioned in the current study with regard to *DNAH9* mutations, those mutation related ultrastructural defects are only localized at the distal portions of the *DNAH9* patients' cilia. Without careful examination of cilia cross sections along the entire length of the cilia, TEM abnormalities could therefore possibly not be identified. Genetic testing in such cases is very valuable to end the confusion over this equivocal ciliary ultrastructural phenotype.

Including since the start of this study in late 2015, mutations have been identified in about 40 PCD genes so far, including the gene discoveries reported here. Mutations in these genes are believed to account for 70-80%

of PCD patients.(124, 131) Since this indicates that additional genes are still to be identified, this reduces the sensitivity of genetic testing for use as a gold standard test for the diagnosis of PCD. With the identification of the whole genetic landscape of PCD possible in the near future, genetics can increasingly be considered as a first-line diagnostic test. Accurate genetic diagnosis will enable affected families to get better advice for proper management and therapy, to access disease-specific family support groups and to accurately determine the risk for recurrence.(28)

Genetic diagnosis can help to predict the disease prognosis and subsequently will help to tailor the management strategies offered to the patients. For instance, PCD-causing mutations in several genes do not lead to laterality abnormalities as the functions of their encoded proteins are not required in the embryonic node e.g. central complex, radial spokes and nexin-links are regarded as dispensable in nodal cilia or not greatly disrupting to their motility when deficient.(43, 80) A lack of laterality defects will also impact the prevalence of congenital heart defects (CHD) among PCD patients: in a recent study conducted on 389 patient from the UK, we reported that laterality defects are a significant risk factor for CHD, which has a high incidence (17.1%) in the UK multi-ethnic PCD cohort.(68) Identification and management of such congenital heart problems at the earliest age will definitely improve the patient prognosis and quality of life.

Most of male PCD patients are thought to be infertile. However, normal fertility status was also reported in PCD-affected males.(393) The incidence of infertility among female PCD patients is not well-defined but there is a widely accepted concept that defective cilia motility of the epithelial lining the fallopian tube can lead to more frequent ectopic pregnancies.(393, 394) In a recent study conducted on a large cohort of adult PCD patients, fertility data was available for 85 patients (36 women, 49 men). Infertility was reported in 75.5% and 61.1% of male and female patients respectively, and in fact, no ectopic pregnancy was reported.(395) Mutations in the PCD genes *CCDC39, CCDC40, DNAAF1* and *LRRC6* were found to be associated with a higher risk of infertility and this is consistent with the previous reports of infertility in PCD patients carrying mutations in these genes.(215, 216, 224, 227, 242, 395)

These findings highlight the importance of genetics in the management of adult PCD patients as most of the infertile patients can benefit from assisted reproductive technologies (ART); fertility counselling should be provided in any standard adult PCD care.(395)

Due to locus and allelic heterogeneity, next generation sequencing (NGS) technology has been able to revolutionize the genetic diagnostics of PCD. The traditional genetic testing of a single candidate gene has been replaced by massively parallel sequencing of coding exons of multiple genes or the whole complement of coding exons (WES) or sequencing of the whole genome (WGS) in a single test.(113) These techniques enable fast and comprehensive genetic analysis and subsequently have improved and expedited the diagnostic process.(56, 78-80, 114)

The use of multi-gene panel sequencing approach including all known PCD genes, especially with the recently developed library preparation techniques and advanced data analysis softwares, is a valuable and efficient tool for genetic diagnosis of PCD which saves considerable time and cost compared to Sanger sequencing approach.(122) Targeted sequencing remains more effective than whole exome sequencing (WES) and whole genome sequencing (WGS) in terms of the achievable level of sequence coverage of the target regions and it is currently linked to reduced time, cost and less of a data burden.(123) The use of targeted sequencing also reduces the possibilities of incidental genetic findings which impose further ethical considerations and difficulties in genetic counselling.

The probands-only panel sequencing approach used in this study yielded a high overall genetic diagnostic output where genetic diagnosis was achieved in 72% of patients with a confirmed or highly suggested PCD diagnosis by other clinical tests. This is higher than the diagnostic output from most of the sequencing studies conducted for genetic screening of PCD patients.(114, 279, 280, 282) This approach is therefore cost effective, with a promising potential to be considered when services for genetic diagnosis of rare diseases are established in developing countries with limited resources.(396, 397)

The genetic results in this study highlighted the impact of ethnicity on the genetic stratification of PCD patients and ensured the importance of including PCD patients from various ethnicities to elucidate the full genetic landscape of PCD. This could be very helpful to prioritize certain genes for genetic testing in certain ethnicities, especially in countries with limited genetic resources such as Palestine and Egypt as shown in this study.(76) For example, *DNAH5* mutations were found to be more frequent in European populations, however they were detected in very few patients from other ethnicities. Employing this type of stratification can lead to establishing an effective early genetic screening program in high risk populations and neonates with symptoms suggestive of PCD, before pursuing further complicated diagnostic testing.(391) Moreover, this can also be beneficial in establishing pre-marital carrier screening programs specially in inbred communities with high rate of consanguineous marriages. These benefits highlight the importance of international collaborations for better understanding of the genetics behind rare diseases.(398)

PCD is generally inherited as an autosomal recessive disease. However in very rare cases, autosomal dominant and X-linked inheritance patterns have been reported for ciliary dyskinesia-like phenotypes.(124, 130) Occasionally, only one mutant allele is identified (7% of the screened cohort in the current study) leading to incomplete genetic diagnosis that may add to the challenge to obtain a definitive diagnosis.(283) These cases may be due to missing more unusual mutations, which are not covered during sequencing such as those present in the deep intronic gene regions or regulatory regions. Identification of these missing causal mutations will require whole genome sequencing as shown in the recent pre-print, by Ellingford et al (2018) who have used WGS to identify a deep intronic variant in *DNAH11* in a PCD patient.(120) RNA sequencing can also be beneficial to identify cases with mono-allelic expression of the mutated allele.(284) There are also other possibilities such as digenic inheritance or the presence of genetic modifiers, which have not been investigated in PCD though these underlying genetic mechanisms were proposed in other ciliopathies.(399)

This study and two other parallel studies conducted within Professor Heymut Omran's laboratory (University Children's Hospital Muenster) led to the identification of two new genes not previously linked to PCD; *C11orf70* and *DNAH9*.(307, 315) These studies not only expand the genetic landscape and improve our understanding of the molecular basis of PCD but will better characterize PCD patients aiming for future improved personalized therapies as well.

The use of model organism studies to understand the underlying disease mechanisms is increasingly growing, especially within the expanding list of cilia related diseases, the ciliopathies.(84) To the best of my knowledge, this is the first study to use *Paramecium* for functional characterization of potential candidate genes in PCD. This added a new model system that is easy to use for the study and validation of genetic variants in PCD.(101, 400)

Identification of PCD-causing mutations in the previously uncharacterized gene, *C11orf70* has improved the current understanding of dynein assembly processes in humans. Functional characterization of this gene revealed a significant role of its encoded protein in the assembly and IFT-dependent transport of motor dyneins to the ciliary compartment.(307, 315) These findings have extended the list of genes that are linked to PCD and are implicated in the human dynein assembly process. Mutation screening of *C11orf70* should be included in any genetic testing of PCD patients with combined loss of IDAs and ODAs.

PCD-causal mutations that have been identified in *DNAH9,* expanded the mutation spectrum of PCD patients with ODA loss. This finding highlighted the importance of screening genes for which their encoded proteins are segmentally located along the ciliary axoneme and thus more likely to cause more subtle motility defects, as is also the case for DNAH11 which mirrors the segmental expression of DNAH9. Such kind of genetic mutations affecting only sub-compartments of the motile cilium should be considered when investigating for the 20-30% of PCD patients where no genetic diagnosis was determined.

Dysfunction of the respiratory cilia associated with non-motile ciliopathies is poorly studied, despite the increased incidence of bronchiectasis in patients with polycystic kidney disease.(401) Identification of motile cilia defects associated with primary ciliopathy syndromes adds a new aspect to the phenotyping of these patients. Motile cilia defects are mostly overlooked in skeletal ciliopathy patients with IFT mutations and respiratory symptoms, since these manifestation are usually attributed secondarily to chest dysplasia and lung restriction.(302, 343) Linking ciliary dyskinesia phenotypes to mutations in IFT genes here, has added a new subgroup of PCD genes for which motile cilia defects are associated with primary ciliopathy phenotypes.

To decipher the whole morbid genome of PCD, the assumption must be that new PCD genes must have an essential role in motile cilia biology and function. These potential candidate genes may likely include genes in which mutations confer motility defects at the milder end of the spectrum explaining why they have been missed to-date, or else they harbour high impact mutations but technically harder to identify. This may be the case for example for genes encoding additional structural proteins such as components of the inner dynein arms. Isolated IDA defects are difficult to be identified by TEM (402) with a prediction that their dysfunction could create a subtle defect in cilia motility and milder manifestations, hence no mutations in such genes have yet been linked to PCD.

Genotype-phenotype correlation for PCD is still not clearly understood. Although the majority of affected patients tend to have static cilia, some cases show a degree of retention of cilia motility that is apparently not sufficient for mucociliary clearance. This partial nature of some mutations in PCD make them potentially more promising targets for therapeutic rescue approaches. This includes for instance, the putative hypomorphic allele p.Val16Gly in *ZMYND10* where significant cilia motility is retained (248) and the striking phenotype associated with p.His154Pro mutation in *CCDC103* where cilia can often appear normal or lacking the IDAs only.(254, 255)

Gene discovery studies in PCD have more recently shifted attention from genes essential for the structural integrity of the cilia, to genes involved in the assembly, transport and docking of these ciliary components. It is possible that there could be points of entry in these pathways, as well as in the multiciliogenesis pathways governed by *CCNO* and *MCIDAS*, that are more potentially 'druggable' than the structural integrity associated ciliary processes. Although at this point, our understanding is somewhat limited in this area and therefore it warrants further molecular investigation to assess these possibilities.

7.2 Future perspectives

A majority of the UK patients with no definitive genetic diagnosis arising from this study, have been submitted to the Genomics England 100,000 Genomes Project for Whole Genome Sequencing. The current study has effectively therefore become a preliminary panel-based genetic screening analysis for the UK-based PCD patients, part of ongoing work that has accelerated since the 100,000 Genomes Project commenced in 2012. The next step arising from this project will be WGS re-analysis of the group of patients, where novel mutation types such as deep intronic mutations or other potential candidate genes may be expected to be identified.

Introducing novel approaches for PCD genomics diagnostics and research will advance our understanding and the current knowledge about the molecular basis of the disease. RNA-sequencing is a rapidly emerging technology that can overcome the limitation of genetic analysis conducted at the DNA level, through identifying global aberrant gene expression profiles mostly due to mutations in promoters or intronic regions. Also, it can detect cases with allele-specific expression where only a heterozygous single coding variant may be identified in DNA-sequencing. Moreover, it can identify defective splicing, recognized as a major cause of Mendelian disorders, which is currently poorly studied in PCD.

The full spectrum of transcript isoforms and gene expression profiles in multiciliated cells are not very well annotated so far. The use of long read RNA-sequencing, which is a promising strategy for confident calling of all gene transcripts, will help to investigate the motile cilia transcriptome. This will provide a gene expression catalogue for various time points during multiciliogenesis, which will provide a valuable resource for PCD and respiratory research. Deciphering the whole transcriptional profile during multiciliogenesis will help accurate interpretation of variants, especially in cases where a variant may only affect a certain transcript or lead to abnormal or ectopic splicing.

Another area requiring attention is our understanding of the cytoplasmic assembly of motor dyneins, which is still poorly characterized despite identification of more than 10 genes linked to defective dynein assembly and associated loss of dynein arms from the ciliary axoneme causing PCD. The complex process of building these multi-protein complexes involves a protein network essential for protein folding/stabilization, as well as assembly and transport factors. Systematic analysis including structural, functional and molecular studies, is essential to provide a better understanding of the biology underlying this complex process as well as to further characterize defective cilia resulting from genetic mutations affecting this pathway. Transcriptome analysis and protein-protein interaction studies will be useful to determine the defective consequences of the dynein assembly process in disease models of the distinct gene defects observed in patients. Importantly, this may help to highlight the molecular partners within the dynein arm pre-assembly program that could be points of entry for targeted therapeutic rescue approaches.

Genotype-phenotype correlation in PCD is also not clearly understood. Although the majority of the affected patients tend to have static cilia. With better imaging methods and ever improving knowledge, we now find increasing numbers of cases that show a degree of retention of cilia motility. This hypomorphic state imposes difficulties in the traditional diagnosis of patients with inconclusive cilia-functional testing results, whilst expand the motile ciliopathy disease spectrum beyond PCD. In particular, further molecular investigation is warranted at the single cell level in order to understand this heterogeneity of genetically identical cells. This would include analysis of cell-specific changes in RNA/protein expression from single cells among a heterogeneous subpopulation e.g. where in the same patient, some cells have static cilia and others still retain cilia motility. However, the partial nature of such cilia defects in PCD make them a potentially promising target for novel therapeutic rescue approaches.

Multi-omics approaches and integrating genomics, transcriptomics and proteomics data with deep phenotyping of the patients will assist in genotype-phenotype correlations for rare diseases in general. Particularly in PCD, multi-

omics studies can improve the molecular understanding of the disease variability and pre-symptomatic disease detection, especially in the increasingly emerging cases with milder disease phenotypes. PCD patients, despite carrying the same genetic mutations, can manifest disease symptoms quite differently. By linking the multi-omics results with the phenotypic data, variable PCD disease features can be placed into the context of their underlying biology and this approach would be significantly promising for the development of genetically relevant personalized therapeutic PCD interventions. Additionally, it would lead to better understanding of how specific gene mutations can cause different aspects and various degrees of severity of PCD. Multi-omics studies can also identify novel biomarkers for improving the understanding of PCD disease severity and staging.

Upper and lower respiratory infections are very common in PCD and consequently can lead to serious complications such as bronchiectasis and lung failure. Investigating host-pathogen interactions and the cellular transcriptional response to infection at a single cell resolution is also another important prospect with translational impact in PCD research. The transcriptomic profile of each single cell during infection may provide meaningful insights at both biological and clinical level and identify novel disease biomarkers.

Investigating the role of motile cilia in the pathogenesis of respiratory diseases other than PCD is another important area. This includes studying the molecular and structural basis underlying the potential secondary cilia defects associated with respiratory infections and other frequent respiratory diseases like cystic fibrosis, asthma and chronic obstructive pulmonary diseases. Inflammation and defective mucociliary clearance represent the main pathophysiological mechanisms underlying these diseases. Studying the interactions between motile cilia and their environment (e.g. composition of mucus or inflammatory mediators) will be beneficial to understand the disease mechanism and the contribution of motile cilia defects in such more common diseases.

Underlying the advances in understanding the genetic landscape of PCD, emerging therapeutic strategies have an increasing potential to be effective in developing a cure for PCD. Gene transfer and genome editing have been used in trials to rescue cilia motility e.g. the repair of *DNAH11*-defective cultured epithelial cells by site-specific recombination using transcription activator-like effector nucleases (TALENs).(403) Other gene editing strategies such as CRISPR/Cas9 and base editing technologies should be pursued not only as therapeutic trials but also to develop better *in vivo* and *in vitro* disease models.

Other therapeutic tools that might be applied in PCD with great promise of favourable outcomes include read-through therapy for premature stop gain mutations, anti-sense oligonucleotides (AONs) for splicing mutations and RNA replacement therapy can also be applied. With the growing understanding of the underlying biology, a subset of PCD can now be considered more as a protein misfolding disease where pharmacological chaperones and proteases regulators have the potential to be effective to stabilize protein structures and provide another potential treatment for PCD. A better understanding of the genetics and molecular mechanisms of PCD offer great hope for therapeutic breakthroughs and even cure for this progressive disease for which little treatment is currently available.

References

1. Fliegauf M, Benzing T, Omran H. When cilia go bad: cilia defects and ciliopathies (vol 8, pg 880, 2007). Nat Rev Mol Cell Bio. 2008;9(1):88-.
2. Mitchison HM, Valente EM. Motile and non-motile cilia in human pathology: from function to phenotypes. J Pathol. 2016.
3. Gherman A, Davis EE, Katsanis N. The ciliary proteome database: an integrated community resource for the genetic and functional dissection of cilia. Nat Genet. 2006;38(9):961-2.
4. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. Nat Rev Mol Cell Bio. 2017;18(9):533-47.
5. Carvalho-Santos Z, Azimzadeh J, Pereira-Leal JB, Bettencourt-Dias M. Evolution: Tracing the origins of centrioles, cilia, and flagella. J Cell Biol. 2011;194(2):165-75.
6. Fisch C, Dupuis-Williams P. Ultrastructure of cilia and flagella - back to the future! Biol Cell. 2011;103(6):249-70.
7. Deane JA, Cole DG, Seeley ES, Diener DR, Rosenbaum JL. Localization of intraflagellar transport protein IFT52 identifies basal body transitional fibers as the docking site for IFT particles. Curr Biol. 2001;11(20):1586-90.
8. Jana SC, Machado P, Rocha J, Mendonca S, Werner S, Bettencourt-Dias M. Architectural variation at the base of the cilia is required for cilia structure and function diversity. Mol Biol Cell. 2014;25.
9. Reiter JF, Blacque OE, Leroux MR. The base of the cilium: roles for transition fibres and the transition zone in ciliary formation, maintenance and compartmentalization. Embo Reports. 2012;13(7):608-18.
10. Garcia-Gonzalo FR, Corbit KC, Sirerol-Piquer MS, Ramaswami G, Otto EA, Noriega TR, et al. A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. Nat Genet. 2011;43(8):776-U88.
11. Fliegauf M, Olbrich H, Horvath J, Wildhaber JH, Zariwala MA, Kennedy M, et al. Mislocalization of DNAH5 and DNAH9 in respiratory cells from patients with primary ciliary dyskinesia. Am J Respir Crit Care Med. 2005;171(12):1343-9.
12. Pennekamp P, Menchen T, Dworniczak B, Hamada H. Situs inversus and ciliary abnormalities: 20 years later, what is the connection? Cilia. 2015;4(1):1.
13. Satir P, Christensen ST. Structure and function of mammalian cilia. Histochem Cell Biol. 2008;129(6):687-93.
14. Milla CE. The evolving spectrum of ciliopathies and respiratory disease. Curr Opin Pediatr. 2016;28(3):339-47.
15. Fowkes ME, Mitchell DR. The role of preassembled cytoplasmic complexes in assembly of flagellar dynein subunits. Mol Biol Cell. 1998;9(9):2337-47.
16. Huizar R, Lee C, Boulgakov A, Horani A, Tu F, Drew K, et al. A phase separated organelle at the root of motile ciliopathy. bioRxiv. 2017.

17. Mali GR, Yeyati PL, Mizuno S, Dodd DO, Tennant PA, Keighren MA, et al. ZMYND10 functions in a chaperone relay during axonemal dynein assembly. Elife. 2018;7.
18. Rosenbaum JL, Witman GB. Intraflagellar transport. Nat Rev Mol Cell Bio. 2002;3(11):813-25.
19. Avidor-Reiss T, Maer AM, Koundakjian E, Polyanovsky A, Keil T, Subramaniam S, et al. Decoding cilia function: Defining specialized genes required for compartmentalized cilia biogenesis. Cell. 2004;117(4):527-39.
20. Mourao A, Christensen ST, Lorentzen E. The intraflagellar transport machinery in ciliary signaling. Current Opinion in Structural Biology. 2016;41:98-108.
21. Sung CH, Leroux MR. The roles of evolutionarily conserved functional modules in cilia-related trafficking. Nature Cell Biology. 2013;15(12):1387-97.
22. Ibanez-Tallon I, Heintz N, Omran H. To beat or not to beat: roles of cilia in development and disease. Hum Mol Genet. 2003;12:R27-R35.
23. Badano JL, Mitsuma N, Beales PL, Katsanis N. The ciliopathies: An emerging class of human genetic disorders. Annual Review of Genomics and Human Genetics. 2006;7:125-48.
24. Satir P, Christensen ST. Overview of structure and function of mammalian cilia. Annual Review of Physiology. 2007;69:377-400.
25. Marshall WF. The cell biological basis of ciliary disease. J Cell Biol. 2008;180(1):17-21.
26. Shaheen R, Szymanska K, Basu B, Patel N, Ewida N, Faqeih E, et al. Characterizing the morbid genome of ciliopathies. Genome Biol. 2016;17(1):242.
27. Quinlan RJ, Tobin JL, Beales PL. Modeling ciliopathies: Primary cilia in development and disease. Curr Top Dev Biol. 2008;84:249-310.
28. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. Nat Rev Genet. 2018;19(5):325.
29. Fry AM, Leaper MJ, Bayliss R. The primary cilium: guardian of organ development and homeostasis. Organogenesis. 2014;10(1):62-8.
30. Praetorius HA, Spring KR. A physiological view of the primary cilium. Annual Review of Physiology. 2005;67:515-29.
31. Huangfu DW, Liu AM, Rakeman AS, Murcia NS, Niswander L, Anderson KV. Hedgehog signalling in the mouse requires intraflagellar transport proteins. Nature. 2003;426(6962):83-7.
32. Schneider L, Cammer M, Lehman J, Nielsen SK, Guerra CF, Veland IR, et al. Directional Cell Migration and Chemotaxis in Wound Healing Response to PDGF-AA are Coordinated by the Primary Cilium in Fibroblasts. Cellular Physiology and Biochemistry. 2010;25(2-3):279-92.
33. Schneider L, Clement CA, Teilmann SC, Pazour GJ, Hoffmann EK, Satir P, et al. PDGFR alpha alpha signaling is regulated through the primary cilium in fibroblasts. Curr Biol. 2005;15(20):1861-6.
34. Ross AJ, May-Simera H, Eichers ER, Kai M, Hill J, Jagger DJ, et al. Disruption of Bardet-Biedl syndrome ciliary proteins perturbs planar cell polarity in vertebrates (vol 37, pg 1135, 2005). Nat Genet. 2005;37(12):1381-
.
35. Ware SM, Aygun MG, Hildebrandt F. Spectrum of clinical diseases caused by disorders of primary cilia. Proc Am Thorac Soc. 2011;8(5):444-50.

36. Bukowy-Bieryllo Z, Zietkiewicz E, Loges NT, Wittmer M, Geremek M, Olbrich H, et al. RPGR mutations might cause reduced orientation of respiratory cilia. Pediatr Pulm. 2013;48(4):352-63.
37. Moore A, Escudier E, Roger G, Tamalet A, Pelosse B, Marlin S, 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):326-33.
38. Li Y, Garrod AS, Madan-Khetarpal S, Sreedher G, McGuire M, Yagi H, et al. Respiratory Motile Cilia Dysfunction in a Patient with Cranioectodermal Dysplasia. American Journal of Medical Genetics Part A. 2015;167(9):2188-96.
39. Budny B, Chen W, Omran H, Fliegauf M, Tzschach A, Wisniewska M, et al. A novel X-linked recessive mental retardation syndrome comprising macrocephaly and ciliary dysfunction is allelic to oral-facial-digital type I syndrome. Hum Genet. 2006;120(2):171-8.
40. Elgeti J, Gompper G. Emergence of metachronal waves in cilia arrays. P Natl Acad Sci USA. 2013;110(12):4470-5.
41. Marshall WF, Kintner C. Cilia orientation and the fluid mechanics of development. Current Opinion in Cell Biology. 2008;20(1):48-52.
42. Nonaka S, Tanaka Y, Okada Y, Takeda S, Harada A, Kanai Y, et al. Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. Cell. 1998;95(6):829-37.
43. Mitchison HM, Valente EM. Motile and non-motile cilia in human pathology: from function to phenotypes. J Pathol. 2017;241(2):294-309.
44. Lucas JS, Burgess A, Mitchison HM, Moya E, Williamson M, Hogg C. Diagnosis and management of primary ciliary dyskinesia. Arch Dis Child. 2014;99(9):850-6.
45. Afzelius BA. Genetical and ultrastructural aspects of the immotile-cilia syndrome. Am J Hum Genet. 1981;33(6):852-64.
46. Kartagener M, Stucki P. Bronchiectasis with Situs Inversus. Archives of Pediatrics. 1962;79(6):193-+.
47. Gorham GW, Merselis JG, Jr. Kartagener's triad: a family study. Bull Johns Hopkins Hosp. 1959;104(1):11-6.
48. Arge E. Transposition of the Viscera and Sterility in Men. Lancet. 1960;1(Feb20):412-4.
49. Afzelius BA. A human syndrome caused by immotile cilia. Science. 1976;193(4250):317-9.
50. Eliasson R, Mossberg B, Camner P, Afzelius BA. The immotile-cilia syndrome. A congenital ciliary abnormality as an etiologic factor in chronic airway infections and male sterility. N Engl J Med. 1977;297(1):1-6.
51. Greenstone M, Cole PJ. Primary Ciliary Dyskinesia. Archives of Disease in Childhood. 1984;59(8):704-6.
52. Rossman CM, Forrest JB, Lee RMKW, Newhouse MT. The Dyskinetic Cilia Syndrome - Ciliary Motility in Immotile Cilia Syndrome. Chest. 1980;78(4):580-2.
53. Boon M, Jorissen M, Proesmans M, De Boeck K. Primary ciliary dyskinesia, an orphan disease. European Journal of Pediatrics. 2013;172(2):151-62.

54. 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;36(6):1248-58.
55. Onoufriadis A, Paff T, Antony D, Shoemark A, Micha D, Kuyt B, et al. Splice-site mutations in the axonemal outer dynein arm docking complex gene CCDC114 cause primary ciliary dyskinesia. Am J Hum Genet. 2013;92(1):88-98.
56. Daniels ML, Noone PG. Genetics, diagnosis, and future treatment strategies for primary ciliary dyskinesia. Expert Opin Orphan Drugs. 2015;3(1):31-44.
57. Shapiro AJ, Zariwala MA, Ferkol T, Davis SD, Sagel SD, Dell SD, et al. Diagnosis, monitoring, and treatment of primary ciliary dyskinesia: PCD foundation consensus recommendations based on state of the art review. Pediatr Pulmonol. 2016;51(2):115-32.
58. 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. 2016.
59. Mullowney T, Manson D, Kim R, Stephens D, Shah V, Dell S. Primary Ciliary Dyskinesia and Neonatal Respiratory Distress. Pediatrics. 2014;134(6):1160-6.
60. Mall MA. Role of cilia, mucus, and airway surface liquid in mucociliary dysfunction: Lessons from mouse models. Journal of Aerosol Medicine and Pulmonary Drug Delivery. 2008;21(1):13-24.
61. Lucas JS, Paff T, Goggin P, Haarman E. Diagnostic Methods in Primary Ciliary Dyskinesia. Paediatr Respir Rev. 2016;18:8-17.
62. Chapelin C, Coste A, Reinert P, Boucherat M, Millepied MC, Poron F, et al. Incidence of primary ciliary dyskinesia in children with recurrent respiratory diseases. Annals of Otology Rhinology and Laryngology. 1997;106(10):854-8.
63. Kennedy MP, Noone PG, Leigh MW, Zariwala MA, Minnix SL, Knowles MR, et al. High-resolution CT of patients with primary ciliary dyskinesia. American Journal of Roentgenology. 2007;188(5):1232-8.
64. Hosie PH, Fitzgerald DA, Jaffe A, Birman CS, Rutland J, Morgan LC. Presentation of primary ciliary dyskinesia in children: 30 years' experience. J Paediatr Child Health. 2015;51(7):722-6.
65. Shapiro AJ, Davis SD, Ferkol T, Dell SD, Rosenfeld M, Olivier KN, et al. Laterality Defects Other Than Situs Inversus Totalis in Primary Ciliary Dyskinesia Insights Into Situs Ambiguus and Heterotaxy. Chest. 2014;146(5):1176-86.
66. Harrison MJ, Shapiro AJ, Kennedy MP. Congenital Heart Disease and Primary Ciliary Dyskinesia. Paediatric Respiratory Reviews. 2016;18:25-32.
67. Kennedy MP, Omran H, Leigh MW, Dell S, Morgan L, Molina PL, et al. Congenital heart disease and other heterotaxic defects in a large cohort of patients with primary ciliary dyskinesia. Circulation. 2007;115(22):2814-21.
68. Hogg C, Best S, Shoemark A, Rubbo B, Patel M, Fassad M, et al. Genetic risk factors for laterality defects and congenital heart disease (CHD) in patients with primary ciliary dyskinesia (PCD). Eur Respir J. 2017;50.
69. Takeuchi K, Kitano M, Ishinaga H, Kobayashi M, Ogawa S, Nakatani K, et al. Recent advances in primary ciliary dyskinesia. Auris Nasus Larynx. 2016;43(3):229-36.

70. Amirav I, Wallmeier J, Loges NT, Menchen T, Pennekamp P, Mussaffi H, et al. Systematic Analysis of CCNO Variants in a Defined Population: Implications for Clinical Phenotype and Differential Diagnosis. Hum Mutat. 2016;37(4):396-405.
71. Ibanez-Tallon I, Pagenstecher A, Fliegauf M, Olbrich H, Kispert A, Ketelsen UP, et al. Dysfunction of axonemal dynein heavy chain Mdnah5 inhibits ependymal flow and reveals a novel mechanism for hydrocephalus formation. Hum Mol Genet. 2004;13(18):2133-41.
72. Faubel R, Westendorf C, Bodenschatz E, Eichele G. Cilia-based flow network in the brain ventricles. Science. 2016;353(6295):176-8.
73. Wallmeier J, Al-Mutairi DA, Chen CT, 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;46(6):646-51.
74. 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;5:4418.
75. Lucas JS, Leigh MW. Diagnosis of primary ciliary dyskinesia: searching for a gold standard. Eur Respir J. 2014;44(6):1418-22.
76. Rumman N, Jackson C, Collins S, Goggin P, Coles J, Lucas JS. Diagnosis of primary ciliary dyskinesia: potential options for resource-limited countries. European Respiratory Review. 2017;26(143).
77. Kinsella E, Dora N, Mellis D, Lettice L, Deveney P, Hill R, et al. Use of a Conditional Ubr5 Mutant Allele to Investigate the Role of an N-End Rule Ubiquitin-Protein Ligase in Hedgehog Signalling and Embryonic Limb Development. Plos One. 2016;11(6).
78. 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;49(1).
79. 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;197(12):e24-e39.
80. Knowles MR, Daniels LA, Davis SD, Zariwala MA, Leigh MW. Primary ciliary dyskinesia. Recent advances in diagnostics, genetics, and characterization of clinical disease. Am J Respir Crit Care Med. 2013;188(8):913-22.
81. Horani A, Brody SL, Ferko TW. Picking up speed: advances in the genetics of primary ciliary dyskinesia. Pediatric Research. 2014;75(1):158-64.
82. Snijders D, Bertozzi I, Barbato A. Nasal NO, high-speed video microscopy, electron microscopy, and genetics: a primary ciliary dyskinesia puzzle to complete. Pediatric Research. 2014;76(3):321-.
83. Ostrowski LE, Dutcher SK, Lo CW. Cilia and models for studying structure and function. Proc Am Thorac Soc. 2011;8(5):423-9.
84. Vincensini L, Blisnick T, Bastin P. 1001 model organisms to study cilia and flagella. Biol Cell. 2011;103(3):109-30.
85. Hom EFY, Witman GB, Harris EH, Dutcher SK, Kamiya R, Mitchell DR, et al. A Unified Taxonomy for Ciliary Dyneins. Cytoskeleton. 2011;68(10):555-65.

86. Omran H, Haffner K, Volkel A, Kuehr J, Ketelsen UP, Ross UH, et al. Homozygosity mapping of a gene locus for primary ciliary dyskinesia on chromosome 5p and identification of the heavy dynein chain DNAH5 as a candidate gene. Am J Respir Cell Mol Biol. 2000;23(5):696-702.
87. Pennarun G, Escudier E, Chapelin C, Bridoux AM, Cacheux V, Roger G, et al. Loss-of-function mutations in a human gene related to Chlamydomonas reinhardtii dynein IC78 result in primary ciliary dyskinesia. Am J Hum Genet. 1999;65(6):1508-19.
88. Nicastro D, Schwartz C, Pierson J, Gaudette R, Porter ME, McIntosh JR. The molecular architecture of axonemes revealed by cryoelectron tomography. Science. 2006;313(5789):944-8.
89. Oda T, Yanagisawa H, Kamiya R, Kikkawa M. A molecular ruler determines the repeat length in eukaryotic cilia and flagella. Science. 2014;346(6211):857-60.
90. Oda T, Hirokawa N, Kikkawa M. Three-dimensional structures of the flagellar dynein-microtubule complex by cryoelectron microscopy. J Cell Biol. 2007;177(2):243-52.
91. Reiten I, Uslu FE, Fore S, Pelgrims R, Ringers C, Verdugo CD, et al. Motile-Cilia-Mediated Flow Improves Sensitivity and Temporal Resolution of Olfactory Computations. Curr Biol. 2017;27(2):166-74.
92. Essner JJ, Amack JD, Nyholm MK, Harris EB, Yost HJ. Kupffer's vesicle is a ciliated organ of asymmetry in the zebrafish embryo that initiates left-right development of the brain, heart and gut. Development. 2005;132(6):1247-60.
93. Yu XW, Lau D, Ng CP, Roy S. Cilia-driven fluid flow as an epigenetic cue for otolith biomineralization on sensory hair cells of the inner ear. Development. 2011;138(3):487-94.
94. Sullivan-Brown J, Schottenfeld J, Okabe N, Hostetter CL, Serluca FC, Thiberge SY, et al. Zebrafish mutations affecting cilia motility share similar cystic phenotypes and suggest a mechanism of cyst formation that differs from pkd2 morphants. Dev Biol. 2008;314(2):261-75.
95. Serluca FC, Xu B, Okabe N, Baker K, Lin SY, Sullivan-Brown J, et al. Mutations in zebrafish leucine-rich repeat-containing six-like affect cilia motility and result in pronephric cysts, but have variable effects on left-right patterning. Development. 2009;136(10):1621-31.
96. Gao C, Wang G, Amack JD, Mitchell DR. Oda16/Wdr69 is essential for axonemal dynein assembly and ciliary motility during zebrafish embryogenesis. Dev Dyn. 2010;239(8):2190-7.
97. Chandrasekar G, Vesterlund L, Hultenby K, Tapia-Paez I, Kere J. The zebrafish orthologue of the dyslexia candidate gene DYX1C1 is essential for cilia growth and function. Plos One. 2013;8(5):e63123.
98. Austin-Tse C, Halbritter J, Zariwala MA, Gilberti RM, Gee HY, Hellman N, et al. Zebrafish Ciliopathy Screen Plus Human Mutational Analysis Identifies C21orf59 and CCDC65 Defects as Causing Primary Ciliary Dyskinesia. Am J Hum Genet. 2013;93(4):672-86.
99. Blum M, De Robertis EM, Wallingford JB, Niehrs C. Morpholinos: Antisense and Sensibility. Dev Cell. 2015;35(2):145-9.
100. Ostrowski LE, Yin WN, Rogers TD, Busalacchi KB, Chua M, O'Neal WK, et al. Conditional Deletion of Dnaic1 in a Murine Model of Primary Ciliary

Dyskinesia Causes Chronic Rhinosinusitis. Am J Resp Cell Mol. 2010;43(1):55-63.
101. Beisson J, Betermier M, Bre MH, Cohen J, Duharcourt S, Duret L, et al. Paramecium tetraurelia: the renaissance of an early unicellular model. Cold Spring Harb Protoc. 2010;2010(1):pdb.emo140.
102. Beisson J, Betermier M, Bre MH, Cohen J, Duharcourt S, Duret L, et al. Silencing specific Paramecium tetraurelia genes by feeding double-stranded RNA. Cold Spring Harb Protoc. 2010;2010(1):pdb.prot5363.
103. Jorissen M, Van der Schueren B, Tyberghein J, Van der Berghe H, Cassiman JJ. Ciliogenesis and coordinated ciliary beating in human nasal epithelial cells cultured in vitro. Acta Otorhinolaryngol Belg. 1989;43(1):67-73.
104. Smith CM, Djakow J, Free RC, Djakow P, Lonnen R, Williams G, et al. ciliaFA: a research tool for automated, high-throughput measurement of ciliary beat frequency using freely available software. Cilia. 2012;1:14.
105. Hirst RA, Jackson CL, Coles JL, Williams G, Rutman A, Goggin PM, et al. Culture of Primary Ciliary Dyskinesia Epithelial Cells at Air-Liquid Interface Can Alter Ciliary Phenotype but Remains a Robust and Informative Diagnostic Aid. Plos One. 2014;9(2).
106. Bukowy Z, Zietkiewicz E, Witt M. In vitro culturing of ciliary respiratory cells-a model for studies of genetic diseases. J Appl Genet. 2011;52(1):39-51.
107. Pifferi M, Montemurro F, Cangiotti AM, Ragazzo V, Di Cicco M, Vinci B, et al. Simplified cell culture method for the diagnosis of atypical primary ciliary dyskinesia. Thorax. 2009;64(12):1077-81.
108. Ross AJ, Dailey LA, Brighton LE, Devlin RB. Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. Am J Respir Cell Mol Biol. 2007;37(2):169-85.
109. Pazour GJ, Agrin N, Leszyk J, Witman GB. Proteomic analysis of a eukaryotic cilium. J Cell Biol. 2005;170(1):103-13.
110. Blackburn K, Bustamante-Marin X, Yin W, Goshe MB, Ostrowski LE. Quantitative Proteomic Analysis of Human Airway Cilia Identifies Previously Uncharacterized Proteins of High Abundance. J Proteome Res. 2017;16(4):1579-92.
111. Mianne J, Ahmed E, Bourguignon C, Fieldes M, Vachier I, Bourdin A, et al. Induced Pluripotent Stem Cells for Primary Ciliary Dyskinesia Modeling and Personalized Medicine. Am J Respir Cell Mol Biol. 2018.
112. Omran H, Kobayashi D, Olbrich H, Tsukahara T, Loges NT, Hagiwara H, et al. Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins. Nature. 2008;456(7222):611-6.
113. Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. Nat Rev Genet. 2013;14(6):415-26.
114. Kim RH, D AH, Cutz E, Knowles MR, Nelligan KA, Nykamp K, et al. The role of molecular genetic analysis in the diagnosis of primary ciliary dyskinesia. Ann Am Thorac Soc. 2014;11(3):351-9.
115. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. Eur J Hum Genet. 2012;20(5):490-7.
116. Ng SB, Nickerson DA, Bamshad MJ, Shendure J. Massively parallel sequencing and rare disease. Hum Mol Genet. 2010;19:R119-R24.
117. 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. Genetics in medicine : official journal of the American College of Medical Genetics. 2015;17(5):405-24.
118. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. Nat Rev Genet. 2018;19(5):253-68.
119. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013;14(10):681-91.
120. Ellingford JM, Beaman G, Webb K, O'Callaghan C, Hirst RA, Black GC, et al. Whole genome sequencing enables definitive diagnosis of Cystic Fibrosis and Primary Ciliary Dyskinesia. bioRxiv. 2018.
121. Haga SB. Update: looking beyond the 100,000 Genome Project. Per Med. 2017;14(2):85-7.
122. Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: promises and challenges. Cancer Lett. 2013;340(2):284-95.
123. Grada A, Weinbrecht K. Next-generation sequencing: methodology and application. J Invest Dermatol. 2013;133(8):e11.
124. Horani A, Ferkol TW, Dutcher SK, Brody SL. Genetics and biology of primary ciliary dyskinesia. Paediatr Respir Rev. 2016;18:18-24.
125. Berg JS, Evans JP, Leigh MW, Omran H, Bizon C, Mane K, et al. Next generation massively parallel sequencing of targeted exomes to identify genetic mutations in primary ciliary dyskinesia: implications for application to clinical testing. Genetics in medicine : official journal of the American College of Medical Genetics. 2011;13(3):218-29.
126. Wooderchak-Donahue WL, O'Fallon B, Furtado LV, Durtschi JD, Plant P, Ridge PG, et al. A direct comparison of next generation sequencing enrichment methods using an aortopathy gene panel- clinical diagnostics perspective. Bmc Med Genomics. 2012;5.
127. 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 Resp Crit Care. 2018;197(12):E24-E39.
128. Olcese C, Patel MP, Shoemark A, Kiviluoto S, Legendre M, Williams HJ, et al. X-linked primary ciliary dyskinesia due to mutations in the cytoplasmic axonemal dynein assembly factor PIH1D3. Nat Commun. 2017;8:14279.
129. Paff T, Loges NT, Aprea I, Wu K, Bakey Z, Haarman EG, et al. Mutations in PIH1D3 Cause X-Linked Primary Ciliary Dyskinesia with Outer and Inner Dynein Arm Defects. Am J Hum Genet. 2017;100(1):160-8.
130. Narayan D, Krishnan SN, Upender M, Ravikumar TS, Mahoney MJ, Dolan TF, Jr., et al. Unusual inheritance of primary ciliary dyskinesia (Kartagener's syndrome). J Med Genet. 1994;31(6):493-6.
131. Horani A, Ferkol TW. Advances in the Genetics of Primary Ciliary Dyskinesia: Clinical Implications. Chest. 2018;154(3):645-52.
132. Mitchell DR. The evolution of eukaryotic cilia and flagella as motile and sensory organelles. Eukaryotic Membranes and Cytoskeleton: Origins and Evolution. 2007;607:130-40.

133. Leigh M, Daniels A, Noone PG. Genetics, diagnosis and future treatment strategies for primary ciliary dyskinesia. Expert Opin Orphan D. 2015;3(1):31-44.
134. Chapelin C, Duriez B, Magnino F, Goossens M, Escudier E, Amselem S. Isolation of several human axonemal dynein heavy chain genes: genomic structure of the catalytic site, phylogenetic analysis and chromosomal assignment. Febs Letters. 1997;412(2):325-30.
135. Maiti AK, Mattei MG, Jorissen M, Volz A, Zeigler A, Bouvagnet P. Identification, tissue specific expression, and chromosomal localisation of several human dynein heavy chain genes. Eur J Hum Genet. 2000;8(12):923-32.
136. Zariwala M, Noone PG, Sannuti A, Minnix S, Zhou Z, Leigh MW, et al. Germline mutations in an intermediate chain dynein cause primary ciliary dyskinesia. Am J Respir Cell Mol Biol. 2001;25(5):577-83.
137. Guichard C, Harricane MC, Lafitte JJ, Godard P, Zaegel M, Tack V, et al. Axonemal dynein intermediate-chain gene (DNAI1) mutations result in situs inversus and primary ciliary dyskinesia (Kartagener syndrome). Am J Hum Genet. 2001;68(4):1030-5.
138. Zariwala MA, Leigh MW, Ceppa F, Kennedy MP, Noone PG, Carson JL, 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):858-66.
139. Zietkiewicz E, Nitka B, Voelkel K, Skrzypczak U, Bukowy Z, Rutkiewicz E, et al. Population specificity of the DNAI1 gene mutation spectrum in primary ciliary dyskinesia (PCD). Respir Res. 2010;11:174.
140. Rupp G, OToole E, Gardner LC, Mitchell BF, Porter ME. The sup-pf-2 mutations of Chlamydomonas alter the activity of the outer dynein arms by modification of the gamma-dynein heavy chain. J Cell Biol. 1996;135(6):1853-65.
141. Olbrich H, Haffner K, Kispert A, Volkel A, Volz A, Sasmaz G, et al. Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right and asymmetry. Nat Genet. 2002;30(2):143-4.
142. Ibanez-Tallon I, Gorokhova S, Heintz N. Loss of function of axonemal dynein Mdnah5 causes primary ciliary dyskinesia and hydrocephalus. Hum Mol Genet. 2002;11(6):715-21.
143. 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;174(2):120-6.
144. 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;46(4):281-6.
145. Mitchell DR, Kang Y. Identification of Oda6 as a Chlamydomonas Dynein Mutant by Rescue with the Wild-Type Gene. J Cell Biol. 1991;113(4):835-42.
146. Pennarun G, Chapelin C, Escudier E, Bridoux AM, Dastot F, Cacheux V, et al. The human dynein intermediate chain 2 gene (DNA12): cloning, mapping, expression pattern, and evaluation as a candidate for primary ciliary dyskinesia. Hum Genet. 2000;107(6):642-9.
147. Yang ZJ, Wu J. Mouse dynein axonemal intermediate chain 2: Cloning and expression. DNA Cell Biol. 2008;27(9):479-88.

148. Kobayashi D, Iijima N, Hagiwara H, Kamura K, Takeda H, Yokoyama T. Characterization of the medaka (Oryzias latipes) primary ciliary dyskinesia mutant, jaodori: Redundant and distinct roles of dynein axonemal intermediate chain 2 (dnai2) in motile cilia. Dev Biol. 2010;347(1):62-70.
149. Nagao Y, Cheng JL, Kamura K, Seki R, Maeda A, Nihei D, et al. Dynein axonemal intermediate chain 2 is required for formation of the left-right body axis and kidney in medaka. Dev Biol. 2010;347(1):53-61.
150. Loges NT, Olbrich H, Fenske L, Mussaffi H, Horvath J, Fliegauf M, et al. DNAI2 Mutations Cause Primary Ciliary Dyskinesia with Defects in the Outer Dynein Arm. Am J Hum Genet. 2008;83(5):547-58.
151. Mazor M, Alkrinawi S, Chalifa-Caspi V, Manor E, Sheffield VC, Aviram M, et al. Primary Ciliary Dyskinesia Caused by Homozygous Mutation in DNAL1, Encoding Dynein Light Chain 1. Am J Hum Genet. 2011;88(5):599-607.
152. Ichikawaa M, Saito K, Yanagisawa H, Yagi T, Kamiya R, Yamaguchi S, et al. Axonemal dynein light chain-1 locates at the microtubule-binding domain of the gamma heavy chain. Mol Biol Cell. 2015;26(23):4236-47.
153. Horvath J, Fliegauf M, Olbrich H, Kispert A, King SM, Mitchison H, et al. Identification and analysis of axonemal dynein light chain 1 in primary ciliary dyskinesia patients. Am J Resp Cell Mol. 2005;33(1):41-7.
154. Gutierrez-Roelens I, Sluysmans T, Jorissen M, Amyere M, Vikkula M. Localization of candidate regions for a novel gene for Kartagener syndrome. Eur J Hum Genet. 2006;14(7):809-15.
155. Duriez B, Duquesnoy P, Escudier E, Bridoux AM, Escalier D, Rayet I, et al. A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia (vol 104, pg 3336, 2007). P Natl Acad Sci USA. 2007;104(15):6490-.
156. Sadek CM, Damdimopoulos AE, Pelto-Huikko M, Gustafsson JA, Spyrou G, Miranda-Vizuete A. Sptrx-2, a fusion protein composed of one thioredoxin and three tandemly repeated NDP-kinase domains is expressed in human testis germ cells. Genes Cells. 2001;6(12):1077-90.
157. Ogawa K, Takai H, Ogiwara A, Yokota E, Shimizu T, Inaba K, et al. Is outer arm dynein intermediate chain 1 multifunctional? Mol Biol Cell. 1996;7(12):1895-907.
158. Bartoloni L, Blouin JL, Pan Y, Gehrig C, Maiti AK, Scamuffa N, et al. Mutations in the DNAH11 (axonemal heavy chain dynein type 11) gene cause one form of situs inversus totalis and most likely primary ciliary dyskinesia. Proc Natl Acad Sci U S A. 2002;99(16):10282-6.
159. Schwabe GC, Hoffmann K, Loges NT, Birker D, Rossier C, de Santi MM, et al. Primary ciliary dyskinesia associated with normal axoneme ultrastructure is caused by DNAH11 mutations. Hum Mutat. 2008;29(2):289-98.
160. Lucas JS, Adam EC, Goggin PM, Jackson CL, Powles-Glover N, Patel SH, et al. Static respiratory cilia associated with mutations in Dnahc11/DNAH11: a mouse model of PCD. Hum Mutat. 2012;33(3):495-503.
161. 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;67(5):433-41.
162. 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. American Journal of Respiratory Cell and Molecular Biology. 2016;55(2):213-24.
163. Shoemark A, Burgoyne T, Kwan R, Dixon M, Patel MP, Rogers AV, et al. Primary ciliary dyskinesia with normal ultrastructure: three-dimensional tomography detects absence of DNAH11. Eur Respir J. 2018;51(2).
164. Takada S, Wilkerson CG, Wakabayashi K, Kamiya R, Witman GB. The outer dynein arm-docking complex: Composition and characterization of a subunit (Oda1) necessary for outer arm assembly. Mol Biol Cell. 2002;13(3):1015-29.
165. Knowles MR, Leigh MW, Ostrowski LE, Huang L, Carson JL, Hazucha MJ, et al. Exome sequencing identifies mutations in CCDC114 as a cause of primary ciliary dyskinesia. Am J Hum Genet. 2013;92(1):99-106.
166. Wakabayashi K, Takada S, Witman GB, Kamiya R. Transport and arrangement of the outer-dynein-arm docking complex in the flagella of Chlamydomonas mutants that lack outer dynein arms. Cell Motil Cytoskel. 2001;48(4):277-86.
167. Hjeij R, Lindstrand A, Francis R, Zariwala MA, Liu XQ, Li Y, et al. ARMC4 Mutations Cause Primary Ciliary Dyskinesia with Randomization of Left/Right Body Asymmetry. Am J Hum Genet. 2013;93(2):357-67.
168. McClintock TS, Glasser CE, Bose SC, Bergman DA. Tissue expression patterns identify mouse cilia genes. Physiol Genomics. 2008;32(2):198-206.
169. Lonergan KM, Chari R, deLeeuw RJ, Shadeo A, Chi B, Tsao MS, et al. Identification of novel lung genes in bronchial epithelium by serial analysis of gene expression. Am J Resp Cell Mol. 2006;35(6):651-61.
170. Onoufriadis A, Shoemark A, Munye MM, James CT, Schmidts M, Patel M, et al. Combined exome and whole-genome sequencing identifies mutations in ARMC4 as a cause of primary ciliary dyskinesia with defects in the outer dynein arm. J Med Genet. 2014;51(1):61-7.
171. Hjeij R, Onoufriadis A, Watson CM, Slagle CE, Klena NT, Dougherty GW, et al. CCDC151 mutations cause primary ciliary dyskinesia by disruption of the outer dynein arm docking complex formation. Am J Hum Genet. 2014;95(3):257-74.
172. Jerber J, Baas D, Soulavie F, Chhin B, Cortier E, Vesque C, et al. The coiled-coil domain containing protein CCDC151 is required for the function of IFT-dependent motile cilia in animals. Hum Mol Genet. 2014;23(3):563-77.
173. Wallmeier J, Shiratori H, Dougherty GW, Edelbusch C, Hjeij R, Loges NT, et al. TTC25 Deficiency Results in Defects of the Outer Dynein Arm Docking Machinery and Primary Ciliary Dyskinesia with Left-Right Body Asymmetry Randomization. Am J Hum Genet. 2016;99(2):460-9.
174. Hayes JM, Kim SK, Abitua PB, Park TJ, Herrington ER, Kitayama A, et al. Identification of novel ciliogenesis factors using a new in vivo model for mucociliary epithelial development. Dev Biol. 2007;312(1):115-30.
175. Pigino G, Bui KH, Maheshwari A, Lupetti P, Diener D, Ishikawa T. Cryoelectron tomography of radial spokes in cilia and flagella. J Cell Biol. 2011;195(4):673-87.
176. Castleman VH, Romio L, Chodhari R, Hirst RA, de Castro SC, Parker KA, et al. 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.

177. Diener DP, Yang PF, Geimer S, Cole DG, Sale WS, Rosenbaum JL. Sequential Assembly of Flagellar Radial Spokes. Cytoskeleton. 2011;68(7):389-400.
178. Zietkiewicz E, Bukowy-Bieryllo Z, Voelkel K, Klimek B, Dmenska H, Pogorzelski A, et al. Mutations in Radial Spoke Head Genes and Ultrastructural Cilia Defects in East-European Cohort of Primary Ciliary Dyskinesia Patients. Plos One. 2012;7(3).
179. Daniels MLA, Leigh MW, Davis SD, Armstrong MC, Carson JL, Hazucha M, et al. Founder Mutation in RSPH4A Identified in Patients of Hispanic Descent with Primary Ciliary Dyskinesia. Hum Mutat. 2013;34(10):1352-6.
180. Alsaadi MM, Gaunt TR, Boustred CR, Guthrie PA, Liu X, Lenzi L, et al. From a single whole exome read to notions of clinical screening: primary ciliary dyskinesia and RSPH9 p.Lys268del in the Arabian Peninsula. Ann Hum Genet. 2012;76(3):211-20.
181. Reish O, Slatkin M, Chapman-Shimshoni D, Elizur A, Chioza B, Castleman V, et al. Founder mutation(s) in the RSPH9 gene leading to primary ciliary dyskinesia in two inbred Bedouin families. Ann Hum Genet. 2010;74(2):117-25.
182. Huang B, Piperno G, Ramanis Z, Luck DJL. Radial Spokes of Chlamydomonas Flagella - Genetic-Analysis of Assembly and Function. J Cell Biol. 1981;88(1):80-8.
183. Sedykh I, TeSlaa JJ, Tatarsky RL, Keller AN, Toops KA, Lakkaraju A, et al. Novel roles for the radial spoke head protein 9 in neural and neurosensory cilia. Sci Rep-Uk. 2016;6.
184. Kott E, Legendre M, Copin B, Papon JF, Dastot-Le Moal F, Montantin G, 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):561-70.
185. Knowles MR, Ostrowski LE, Leigh MW, Sears PR, Davis SD, Wolf WE, et al. Mutations in RSPH1 Cause Primary Ciliary Dyskinesia with a Unique Clinical and Ciliary Phenotype. Am J Resp Crit Care. 2014;189(6):707-17.
186. Onoufriadis A, Shoemark A, Schmidts M, Patel M, Jimenez G, Liu H, et al. Targeted NGS gene panel identifies mutations in RSPH1 causing primary ciliary dyskinesia and a common mechanism for ciliary central pair agenesis due to radial spoke defects. Hum Mol Genet. 2014;23(13):3362-74.
187. Jeanson L, Copin B, Papon JF, Dastot-Le Moal F, Duquesnoy P, Montantin G, et al. RSPH3 Mutations Cause Primary Ciliary Dyskinesia with Central-Complex Defects and a Near Absence of Radial Spokes. Am J Hum Genet. 2015;97(1):153-62.
188. Sivadas P, Dienes JM, St Maurice M, Meek WD, Yang PF. A flagellar A-kinase anchoring protein with two amphipathic helices forms a structural scaffold in the radial spoke complex. J Cell Biol. 2012;199(4):639-51.
189. El Khouri E, Thomas L, Jeanson L, Bequignon E, Vallette B, Duquesnoy P, et al. Mutations in DNAJB13, Encoding an HSP40 Family Member, Cause Primary Ciliary Dyskinesia and Male Infertility. Am J Hum Genet. 2016;99(2):489-500.
190. Guan J, Yuan L. A heat-shock protein 40, DNAJB13, is an axoneme-associated component in mouse spermatozoa. Mol Reprod Dev. 2008;75(9):1379-86.

191. Guan J, Ekwurtzel E, Kvist U, Hultenby K, Yuan L. DNAJB13 is a Radial Spoke Protein of Mouse '9+2' Axoneme. Reprod Domest Anim. 2010;45(6):992-6.
192. Loreng TD, Smith EF. The Central Apparatus of Cilia and Eukaryotic Flagella. Cold Spring Harb Perspect Biol. 2017;9(2).
193. Davy BE, Robinson ML. Congenital hydrocephalus in hy3 mice is caused by a frameshift mutation in Hydin, a large novel gene. Hum Mol Genet. 2003;12(10):1163-70.
194. Lechtreck KF, Delmotte P, Robinson ML, Sandersoll MJ, Witman GB. Mutations in Hydin impair ciliary motility in mice. J Cell Biol. 2008;180(3):633-43.
195. Lechtreck KF, Witman GB. Chlamydomonas reinhardtii hydin is a central pair protein required for flagellar motility. J Cell Biol. 2007;176(4):473-82.
196. Olbrich H, Schmidts M, Werner C, Onoufriadis A, Loges NT, Raidt J, et al. Recessive HYDIN Mutations Cause Primary Ciliary Dyskinesia without Randomization of Left-Right Body Asymmetry. Am J Hum Genet. 2012;91(4):672-84.
197. Edelbusch C, Cindric S, Dougherty GW, Loges NT, Olbrich H, Rivlin J, et al. Mutation of serine/threonine protein kinase 36 (STK36) causes primary ciliary dyskinesia with a central pair defect. Hum Mutat. 2017;38(8):964-9.
198. Vogel P, Read RW, Hansen GM, Payne BJ, Small D, Sands AT, et al. Congenital Hydrocephalus in Genetically Engineered Mice. Vet Pathol. 2012;49(1):166-81.
199. Nozawa YI, Yao E, Lin CW, Yang JH, Wilson CW, Gacayan R, et al. Fused (Stk36) is a Ciliary Protein Required for Central Pair Assembly and Motile Cilia Orientation in the Mammalian Oviduct. Dev Dynam. 2013;242(11):1307-19.
200. Bower R, Tritschler D, VanderWaal K, Perrone CA, Mueller J, Fox L, et al. The N-DRC forms a conserved biochemical complex that maintains outer doublet alignment and limits microtubule sliding in motile axonemes. Mol Biol Cell. 2013;24(8):1134-52.
201. Gardner LC, Otoole E, Perrone CA, Giddings T, Porter ME. Components of a Dynein-Regulatory-Complex Are Located at the Junction between the Radial Spokes and the Dynein Arms in Chlamydomonas-Flagella. J Cell Biol. 1994;127(5):1311-25.
202. Piperno G, Mead K, Ledizet M, Moscatelli A. Mutations in the Dynein Regulatory Complex Alter the Atp-Insensitive Binding-Sites for Inner Arm Dyneins in Chlamydomonas Axonemes. J Cell Biol. 1994;125(5):1109-17.
203. Wirschell M, Olbrich H, Werner C, Tritschler D, Bower R, Sale WS, et al. The nexin-dynein regulatory complex subunit DRC1 is essential for motile cilia function in algae and humans. Nat Genet. 2013;45(3):262-8.
204. Horani A, Brody SL, Ferkol TW, Shoseyov D, Wasserman MG, Ta-shma A, et al. CCDC65 Mutation Causes Primary Ciliary Dyskinesia with Normal Ultrastructure and Hyperkinetic Cilia. Plos One. 2013;8(8).
205. Lin JF, Tritschler D, Song KK, Barber CF, Cobb JS, Porter ME, et al. Building Blocks of the Nexin-Dynein Regulatory Complex in Chlamydomonas Flagella. J Biol Chem. 2011;286(33):29175-91.
206. Kato T, Kagami O, Yagi T, Kamiya R. Isolation of 2 Species of Chlamydomonas-Reinhardtii Flagellar Mutants, Ida5 and Ida6, That Lack a

Newly Identified Heavy-Chain of the Inner Dynein Arm. Cell Struct Funct. 1993;18(6):371-7.
207. Bower R, Tritschler D, Mills KV, Heuser T, Nicastro D, Porter ME. DRC2/CCDC65 is a central hub for assembly of the nexin-dynein regulatory complex and other regulators of ciliary and flagellar motility. Mol Biol Cell. 2018;29(2):137-53.
208. Olbrich H, Cremers C, Loges NT, Werner C, Nielsen KG, Marthin JK, et al. Loss-of-Function GAS8 Mutations Cause Primary Ciliary Dyskinesia and Disrupt the Nexin-Dynein Regulatory Complex. Am J Hum Genet. 2015;97(4):546-54.
209. Jeanson L, Thomas L, Copin B, Coste A, Sermet-Gaudelus I, Dastot-Le Moal F, et al. Mutations in GAS8, a Gene Encoding a Nexin-Dynein Regulatory Complex Subunit, Cause Primary Ciliary Dyskinesia with Axonemal Disorganization. Hum Mutat. 2016;37(8):776-85.
210. Yeh SD, Chen YJ, Chang ACY, Ray R, She BR, Lee WS, et al. Isolation and properties of Gas8, a growth arrest-specific gene regulated during male gametogenesis to produce a protein associated with the sperm motility apparatus. J Biol Chem. 2002;277(8):6311-7.
211. Merveille AC, Davis EE, Becker-Heck A, Legendre M, Amirav I, Bataille G, 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. Nat Genet. 2011;43(1):72-U98.
212. Abdelhamed Z, Vuong SM, Hill L, Shula C, Timms A, Beier D, et al. A mutation in Ccdc39 causes neonatal hydrocephalus with abnormal motile cilia development in mice. Development. 2018;145(1).
213. Lin HW, Zhang ZY, Guo SY, Chen F, Kessler JM, Wang YM, et al. A NIMA-Related Kinase Suppresses the Flagellar Instability Associated with the Loss of Multiple Axonemal Structures. Plos Genet. 2015;11(9).
214. Becker-Heck A, Zohn IE, Okabe N, Pollock A, Lenhart KB, Sullivan-Brown J, et al. The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis formation. Nat Genet. 2011;43(1):79-U105.
215. Blanchon S, Legendre M, Copin B, Duquesnoy P, Montantin G, Kott E, et al. Delineation of CCDC39/CCDC40 mutation spectrum and associated phenotypes in primary ciliary dyskinesia. J Med Genet. 2012;49(6):410-6.
216. Antony D, Becker-Heck A, Zariwala MA, Schmidts M, Onoufriadis A, Forouhan M, 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):462-72.
217. Kobayashi D, Takeda H. Ciliary motility: the components and cytoplasmic preassembly mechanisms of the axonemal dyneins. Differentiation. 2012;83(2):S23-9.
218. Kollmar M. Fine-Tuning Motile Cilia and Flagella: Evolution of the Dynein Motor Proteins from Plants to Humans at High Resolution. Mol Biol Evol. 2016;33(12):3249-67.
219. King SM, Patel-King RS. The oligomeric outer dynein arm assembly factor CCDC103 is tightly integrated within the ciliary axoneme and exhibits periodic binding to microtubules. J Biol Chem. 2015;290(12):7388-401.

220. Hojo M, Takashima S, Kobayashi D, Sumeragi A, Shimada A, Tsukahara T, et al. Right-elevated expression of charon is regulated by fluid flow in medaka Kupffer's vesicle. Dev Growth Differ. 2007;49(5):395-405.
221. Huang B, Piperno G, Luck DJ. Paralyzed flagella mutants of Chlamydomonas reinhardtii. Defective for axonemal doublet microtubule arms. J Biol Chem. 1979;254(8):3091-9.
222. Stephens RE. Synthesis and turnover of embryonic sea urchin ciliary proteins during selective inhibition of tubulin synthesis and assembly. Mol Biol Cell. 1997;8(11):2187-98.
223. Matsuo M, Shimada A, Koshida S, Saga Y, Takeda H. The establishment of rotational polarity in the airway and ependymal cilia: analysis with a novel cilium motility mutant mouse. Am J Physiol Lung Cell Mol Physiol. 2013;304(11):L736-45.
224. Loges NT, Olbrich H, Becker-Heck A, Haffner K, Heer A, Reinhard C, et al. Deletions and point mutations of LRRC50 cause primary ciliary dyskinesia due to dynein arm defects. Am J Hum Genet. 2009;85(6):883-9.
225. van Rooijen E, Giles RH, Voest EE, van Rooijen C, Schulte-Merker S, van Eeden FJ. LRRC50, a conserved ciliary protein implicated in polycystic kidney disease. J Am Soc Nephrol. 2008;19(6):1128-38.
226. Freshour J, Yokoyama R, Mitchell DR. Chlamydomonas flagellar outer row dynein assembly protein ODA7 interacts with both outer row and I1 inner row dyneins. J Biol Chem. 2007;282(8):5404-12.
227. Duquesnoy P, Escudier E, Vincensini L, Freshour J, Bridoux AM, Coste A, et al. Loss-of-function mutations in the human ortholog of Chlamydomonas reinhardtii ODA7 disrupt dynein arm assembly and cause primary ciliary dyskinesia. Am J Hum Genet. 2009;85(6):890-6.
228. Ha S, Lindsay AM, Timms AE, Beier DR. Mutations in Dnaaf1 and Lrrc48 Cause Hydrocephalus, Laterality Defects, and Sinusitis in Mice. G3 (Bethesda). 2016;6(8):2479-87.
229. Miao C, Jiang Q, Li H, Zhang Q, Bai B, Bao Y, et al. Mutations in the Motile Cilia Gene DNAAF1 Are Associated with Neural Tube Defects in Humans. G3 (Bethesda). 2016;6(10):3307-16.
230. Basten SG, Davis EE, Gillis AJ, van Rooijen E, Stoop H, Babala N, et al. Mutations in LRRC50 predispose zebrafish and humans to seminomas. Plos Genet. 2013;9(4):e1003384.
231. Hartill VL, van de Hoek G, Patel MP, Little R, Watson CM, Berry IR, et al. DNAAF1 links heart laterality with the AAA+ ATPase RUVBL1 and ciliary intraflagellar transport. Hum Mol Genet. 2017.
232. Mitchison HM, Schmidts M, Loges NT, Freshour J, Dritsoula A, Hirst RA, et al. Mutations in axonemal dynein assembly factor DNAAF3 cause primary ciliary dyskinesia. Nat Genet. 2012;44(4):381-9, s1-2.
233. 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;45(9):995-1003.
234. Chen Y, Zhao M, Wang S, Chen J, Wang Y, Cao Q, et al. A novel role for DYX1C1, a chaperone protein for both Hsp70 and Hsp90, in breast cancer. J Cancer Res Clin Oncol. 2009;135(9):1265-76.
235. Vaughan CK. Hsp90 picks PIKKs via R2TP and Tel2. Structure. 2014;22(6):799-800.

236. Tammimies K, Bieder A, Lauter G, Sugiaman-Trapman D, Torchet R, Hokkanen ME, et al. Ciliary dyslexia candidate genes DYX1C1 and DCDC2 are regulated by Regulatory Factor X (RFX) transcription factors through X-box promoter motifs. Faseb j. 2016;30(10):3578-87.
237. Dong F, Shinohara K, Botilde Y, Nabeshima R, Asai Y, Fukumoto A, et al. Pih1d3 is required for cytoplasmic preassembly of axonemal dynein in mouse sperm. J Cell Biol. 2014;204(2):203-13.
238. Horani A, Druley TE, Zariwala MA, Patel AC, Levinson BT, Van Arendonk LG, et al. Whole-exome capture and sequencing identifies HEATR2 mutation as a cause of primary ciliary dyskinesia. Am J Hum Genet. 2012;91(4):685-93.
239. Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, et al. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res. 2012;40(Database issue):D284-9.
240. Diggle CP, Moore DJ, Mali G, zur Lage P, Ait-Lounis A, Schmidts M, et al. HEATR2 plays a conserved role in assembly of the ciliary motile apparatus. Plos Genet. 2014;10(9):e1004577.
241. Horani A, Ustione A, Huang T, Firth AL, Pan J, Gunsten SP, et al. Establishment of the early cilia preassembly protein complex during motile ciliogenesis. P Natl Acad Sci USA. 2018;115(6):E1221-E8.
242. Kott E, Duquesnoy P, Copin B, Legendre M, Dastot-Le Moal F, Montantin G, et al. Loss-of-function mutations in LRRC6, a gene essential for proper axonemal assembly of inner and outer dynein arms, cause primary ciliary dyskinesia. Am J Hum Genet. 2012;91(5):958-64.
243. Newton FG, zur Lage PI, Karak S, Moore DJ, Gopfert MC, Jarman AP. Forkhead transcription factor Fd3F cooperates with Rfx to regulate a gene expression program for mechanosensory cilia specialization. Dev Cell. 2012;22(6):1221-33.
244. Horani A, Ferkol TW, Shoseyov D, Wasserman MG, Oren YS, Kerem B, et al. LRRC6 mutation causes primary ciliary dyskinesia with dynein arm defects. Plos One. 2013;8(3):e59436.
245. Kavlie RG, Kernan MJ, Eberl DF. Hearing in Drosophila requires TilB, a conserved protein associated with ciliary motility. Genetics. 2010;185(1):177-88.
246. Knowles MR, Ostrowski LE, Loges NT, Hurd T, Leigh MW, Huang L, et al. Mutations in SPAG1 cause primary ciliary dyskinesia associated with defective outer and inner dynein arms. Am J Hum Genet. 2013;93(4):711-20.
247. Neesse A, Gangeswaran R, Luettges J, Feakins R, Weeks ME, Lemoine NR, et al. Sperm-associated antigen 1 is expressed early in pancreatic tumorigenesis and promotes motility of cancer cells. Oncogene. 2007;26(11):1533-45.
248. Moore DJ, Onoufriadis A, Shoemark A, Simpson MA, zur Lage PI, de Castro SC, et al. Mutations in ZMYND10, a gene essential for proper axonemal assembly of inner and outer dynein arms in humans and flies, cause primary ciliary dyskinesia. Am J Hum Genet. 2013;93(2):346-56.
249. Kurkowiak M, Zietkiewicz E, Greber A, Voelkel K, Wojda A, Pogorzelski A, et al. ZMYND10--Mutation Analysis in Slavic Patients with Primary Ciliary Dyskinesia. Plos One. 2016;11(1):e0148067.

250. Cho KJ, Noh SH, Han SM, Choi WI, Kim HY, Yu S, et al. ZMYND10 stabilizes intermediate chain proteins in the cytoplasmic pre-assembly of dynein arms. Plos Genet. 2018;14(3).
251. Dubruille R, Laurencon A, Vandaele C, Shishido E, Coulon-Bublex M, Swoboda P, et al. Drosophila regulatory factor X is necessary for ciliated sensory neuron differentiation. Development. 2002;129(23):5487-98.
252. Zariwala MA, Gee HY, Kurkowiak M, Al-Mutairi DA, Leigh MW, Hurd TW, et al. ZMYND10 is mutated in primary ciliary dyskinesia and interacts with LRRC6. Am J Hum Genet. 2013;93(2):336-45.
253. Jaffe KM, Grimes DT, Schottenfeld-Roames J, Werner ME, Ku TS, Kim SK, et al. c21orf59/kurly Controls Both Cilia Motility and Polarization. Cell Rep. 2016;14(8):1841-9.
254. Panizzi JR, Becker-Heck A, Castleman VH, Al-Mutairi DA, Liu Y, Loges NT, et al. CCDC103 mutations cause primary ciliary dyskinesia by disrupting assembly of ciliary dynein arms. Nat Genet. 2012;44(6):714-9.
255. Shoemark A, Moya E, Hirst RA, Patel MP, Robson EA, Hayward J, 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):157-66.
256. You Y, Huang T, Richer EJ, Schmidt JE, Zabner J, Borok Z, et al. Role of f-box factor foxj1 in differentiation of ciliated airway epithelial cells. Am J Physiol Lung Cell Mol Physiol. 2004;286(4):L650-7.
257. Rock JR, Gao X, Xue Y, Randell SH, Kong YY, Hogan BL. Notch-dependent differentiation of adult airway basal stem cells. Cell Stem Cell. 2011;8(6):639-48.
258. Kyrousi C, Arbi M, Pilz GA, Pefani DE, Lalioti ME, Ninkovic J, et al. Mcidas and GemC1 are key regulators for the generation of multiciliated ependymal cells in the adult neurogenic niche. Development. 2015;142(21):3661-74.
259. Funk MC, Bera AN, Menchen T, Kuales G, Thriene K, Lienkamp SS, et al. Cyclin O (Ccno) functions during deuterosome-mediated centriole amplification of multiciliated cells. Embo J. 2015;34(8):1078-89.
260. Trump N, McTague A, Brittain H, Papandreou A, Meyer E, Ngoh A, et al. Improving diagnosis and broadening the phenotypes in early-onset seizure and severe developmental delay disorders through gene panel analysis. J Med Genet. 2016;53(5):310-7.
261. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics. 2012;28(21):2747-54.
262. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.
263. Sheridan E, Wright J, Small N, Corry PC, Oddie S, Whibley C, et al. Risk factors for congenital anomaly in a multiethnic birth cohort: an analysis of the Born in Bradford study. Lancet. 2013;382(9901):1350-9.
264. Wright J, Small N, Raynor P, Tuffnell D, Bhopal R, Cameron N, et al. Cohort Profile: The Born in Bradford multi-ethnic family cohort study. International Journal of Epidemiology. 2013;42(4):978-91.

265. Koshy R, Ranawat A, Scaria V. al mena: a comprehensive resource of human genetic variants integrating genomes and exomes from Arab, Middle Eastern and North African populations. J Hum Genet. 2017;62(10):889-94.
266. Shoemark A, Dixon M, Corrin B, Dewar A. Twenty-year review of quantitative transmission electron microscopy for the diagnosis of primary ciliary dyskinesia. J Clin Pathol. 2012;65(3):267-71.
267. Chilvers MA, Rutman A, O'Callaghan C. Functional analysis of cilia and ciliated epithelial ultrastructure in healthy children and young adults. Thorax. 2003;58(4):333-8.
268. Munye MM, Shoemark A, Hirst RA, Delhove JM, Sharp TV, McKay TR, et al. BMI-1 extends proliferative potential of human bronchial epithelial cells while retaining their mucociliary differentiation capacity. Am J Physiol Lung Cell Mol Physiol. 2017;312(2):L258-l67.
269. Hirst RA, Rutman A, Williams G, O'Callaghan C. Ciliated air-liquid cultures as an aid to diagnostic testing of primary ciliary dyskinesia. Chest. 2010;138(6):1441-7.
270. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods. 2001;25(4):402-8.
271. Boldt K, van Reeuwijk J, Gloeckner CJ, Ueffing M, Roepman R. Tandem Affinity Purification of Ciliopathy-Associated Protein Complexes. Cilia: Structure and Motility. 2009;91:143-60.
272. Boldt K, van Reeuwijk J, Lu Q, Koutroumpas K, Nguyen TMT, Texier Y, et al. An organelle-specific protein landscape identifies novel diseases and molecular mechanisms. Nature Communications. 2016;7.
273. Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, et al. Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature. 2006;444(7116):171-8.
274. Beisson J, Betermier M, Bre MH, Cohen J, Duharcourt S, Duret L, et al. Mass culture of Paramecium tetraurelia. Cold Spring Harb Protoc. 2010;2010(1):pdb.prot5362.
275. Skouri F, Cohen J. Genetic approach to regulated exocytosis using functional complementation in Paramecium: Identification of the ND7 gene required for membrane fusion. Mol Biol Cell. 1997;8(6):1063-71.
276. Caspar SM, Dubacher N, Kopps AM, Meienberg J, Henggeler C, Matyas G. Clinical sequencing: From raw data to diagnosis with lifetime value. Clin Genet. 2018;93(3):508-19.
277. Hehir-Kwa JY, Claustres M, Hastings RJ, van Ravenswaaij-Arts C, Christenhusz G, Genuardi M, et al. Towards a European consensus for reporting incidental findings during clinical NGS testing. Eur J Hum Genet. 2015;23(12):1601-6.
278. 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;44(6):1579-88.
279. 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;5(8):1775-81.
280. Djakow J, Kramna L, Dusatkova L, Uhlik J, Pursiheimo JP, Svobodova T, et al. An effective combination of sanger and next generation sequencing

in diagnostics of primary ciliary dyskinesia. Pediatr Pulmonol. 2016;51(5):498-509.
281. Fedick AM, Jalas C, Treff NR, Knowles MR, Zariwala MA. Carrier frequencies of eleven mutations in eight genes associated with primary ciliary dyskinesia in the Ashkenazi Jewish population. Molecular Genetics & Genomic Medicine. 2015;3(2):137-42.
282. Boaretto F, Snijders D, Salvoro C, Spalletta A, Mostacciuolo ML, Collura M, et al. Diagnosis of Primary Ciliary Dyskinesia by a Targeted Next-Generation Sequencing Panel: Molecular and Clinical Findings in Italian Patients. J Mol Diagn. 2016;18(6):912-22.
283. Paff T, Kooi IE, Moutaouakil Y, Riesebos E, Sistermans EA, Daniels H, et al. Diagnostic yield of a targeted gene panel in primary ciliary dyskinesia patients. Hum Mutat. 2018;39(5):653-65.
284. Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat Commun. 2017;8:15824.
285. Boon M, Smits A, Cuppens H, Jaspers M, Proesmans M, Dupont LJ, et al. Primary ciliary dyskinesia: critical evaluation of clinical symptoms and diagnosis in patients with normal and abnormal ultrastructure. Orphanet J Rare Dis. 2014;9:11.
286. Werner C, Kouis P. Should transmission electron microscopy and ultrastructural cilia evaluation remain part of the diagnostic work-up for primary ciliary dyskinesia? Ultrastruct Pathol. 2017;41(6):386-9.
287. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511(7509):344-7.
288. Glessner JT, Bick AG, Ito K, Homsy J, Rodriguez-Murillo L, Fromer M, et al. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. Circ Res. 2014;115(10):884-96.
289. Ellingford JM, Campbell C, Barton S, Bhaskar S, Gupta S, Taylor RL, et al. Validation of copy number variation analysis for next-generation sequencing diagnostics. Eur J Hum Genet. 2017;25(6):719-24.
290. Hjeij R, Lindstrand A, Francis R, Zariwala MA, Liu X, Li Y, et al. ARMC4 mutations cause primary ciliary dyskinesia with randomization of left/right body asymmetry. Am J Hum Genet. 2013;93(2):357-67.
291. Funkhouser WK, 3rd, Niethammer M, Carson JL, Burns KA, Knowles MR, Leigh MW, et al. A new tool improves diagnostic test performance for transmission em evaluation of axonemal dynein arms. Ultrastruct Pathol. 2014;38(4):248-55.
292. Lin J, Okada K, Raytchev M, Smith MC, Nicastro D. Structural mechanism of the dynein power stroke. Nat Cell Biol. 2014;16(5):479-85.
293. Kamiya R, Yagi T. Functional diversity of axonemal dyneins as assessed by in vitro and in vivo motility assays of Chlamydomonas mutants. Zoolog Sci. 2014;31(10):633-44.
294. Desai PB, Freshour JR, Mitchell DR. Chlamydomonas axonemal dynein assembly locus ODA8 encodes a conserved flagellar protein needed for cytoplasmic maturation of outer dynein arm complexes. Cytoskeleton (Hoboken). 2015;72(1):16-28.

295. Kozminski KG, Beech PL, Rosenbaum JL. The Chlamydomonas kinesin-like protein FLA10 is involved in motility associated with the flagellar membrane. J Cell Biol. 1995;131(6 Pt 1):1517-27.
296. Ahmed NT, Gao C, Lucker BF, Cole DG, Mitchell DR. ODA16 aids axonemal outer row dynein assembly through an interaction with the intraflagellar transport machinery. J Cell Biol. 2008;183(2):313-22.
297. Inaba Y, Shinohara K, Botilde Y, Nabeshima R, Takaoka K, Ajima R, et al. Transport of the outer dynein arm complex to cilia requires a cytoplasmic protein Lrrc6. Genes Cells. 2016;21(7):728-39.
298. Olcese C, Patel MP, Shoemark A, Kiviluoto S, Legendre M, Williams HJ, et al. X-linked primary ciliary dyskinesia due to mutations in the cytoplasmic axonemal dynein assembly factor PIH1D3. Nat Commun. 2017;8:14279.
299. Paff T, Loges NT, Aprea I, Wu K, Bakey Z, Haarman EG, et al. Mutations in PIH1D3 Cause X-Linked Primary Ciliary Dyskinesia with Outer and Inner Dynein Arm Defects. Am J Hum Genet. 2017;100(1):160-8.
300. Kee HL, Dishinger JF, Blasius TL, Liu CJ, Margolis B, Verhey KJ. A size-exclusion permeability barrier and nucleoporins characterize a ciliary pore complex that regulates transport into cilia. Nature Cell Biology. 2012;14(4):431-+.
301. Lechtreck KF. IFT-Cargo Interactions and Protein Transport in Cilia. Trends in Biochemical Sciences. 2015;40(12):765-78.
302. Kozminski KG, Johnson KA, Forscher P, Rosenbaum JL. A Motility in the Eukaryotic Flagellum Unrelated to Flagellar Beating. P Natl Acad Sci USA. 1993;90(12):5519-23.
303. Hou Y, Witman GB. The N-terminus of IFT46 mediates intraflagellar transport of outer arm dynein and its cargo-adaptor ODA16. Mol Biol Cell. 2017;28(18):2420-33.
304. Taschner M, Mourao A, Awasthi M, Basquin J, Lorentzen E. Structural basis of outer dynein arm intraflagellar transport by the transport adaptor protein ODA16 and the intraflagellar transport protein IFT46. J Biol Chem. 2017;292(18):7462-73.
305. Leigh MW, Hazucha MJ, Chawla KK, Baker BR, Shapiro AJ, Brown DE, et al. Standardizing nasal nitric oxide measurement as a test for primary ciliary dyskinesia. Annals of the American Thoracic Society. 2013;10(6):574-81.
306. Consortium GT, Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5.
307. Hoben IM, Hjeij R, Olbrich H, Dougherty GW, Nothe-Menchen T, Aprea I, et al. Mutations in C11orf70 Cause Primary Ciliary Dyskinesia with Randomization of Left/Right Body Asymmetry Due to Defects of Outer and Inner Dynein Arms. Am J Hum Genet. 2018;102(5):973-84.
308. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Molecular & cellular proteomics : MCP. 2014;13(2):397-406.
309. Geremek M, Bruinenberg M, Zietkiewicz E, Pogorzelski A, Witt M, Wijmenga C. Gene expression studies in cells from primary ciliary dyskinesia patients identify 208 potential ciliary genes. Hum Genet. 2011;129(3):283-93.

310. Arnaiz O, Malinowska A, Klotz C, Sperling L, Dadlez M, Koll F, et al. Cildb: a knowledgebase for centrosomes and cilia. Database (Oxford). 2009;2009:bap022.
311. Arnaiz O, Gout JF, Betermier M, Bouhouche K, Cohen J, Duret L, et al. Gene expression in a paleopolyploid: a transcriptome resource for the ciliate Paramecium tetraurelia. BMC Genomics. 2010;11:547.
312. Stolc V, Samanta MP, Tongprasit W, Marshall WF. Genome-wide transcriptional analysis of flagellar regeneration in Chlamydomonas reinhardtii identifies orthologs of ciliary disease genes. Proc Natl Acad Sci U S A. 2005;102(10):3703-7.
313. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, et al. The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science. 2007;318(5848):245-50.
314. Albee AJ, Kwan AL, Lin H, Granas D, Stormo GD, Dutcher SK. Identification of cilia genes that affect cell-cycle progression using whole-genome transcriptome analysis in Chlamydomonas reinhardtti. G3 (Bethesda). 2013;3(6):979-91.
315. Fassad MR, Shoemark A, le Borgne P, Koll F, Patel M, Dixon M, et al. C11orf70 Mutations Disrupting the Intraflagellar Transport-Dependent Assembly of Multiple Axonemal Dyneins Cause Primary Ciliary Dyskinesia. Am J Hum Genet. 2018;102(5):956-72.
316. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. Nat Rev Mol Cell Biol. 2017;18(9):533-47.
317. Holzbaur ELF, Vallee RB. Dyneins - Molecular-Structure and Cellular Function. Annual Review of Cell Biology. 1994;10:339-72.
318. King SM. Axonemal Dynein Arms. Cold Spring Harb Perspect Biol. 2016;8(11).
319. Pazour GJ, Agrin N, Walker BL, Witman GB. Identification of predicted human outer dynein arm genes: candidates for primary ciliary dyskinesia genes. J Med Genet. 2006;43(1):62-73.
320. Burgoyne T, Dixon M, Luther P, Hogg C, Shoemark A. Generation of a three-dimensional ultrastructural model of human respiratory cilia. Am J Respir Cell Mol Biol. 2012;47(6):800-6.
321. Viswanadha R, Sale WS, Porter ME. Ciliary Motility: Regulation of Axonemal Dynein Motors. Cold Spring Harb Perspect Biol. 2017;9(8).
322. Fliegauf M, Olbrich H, Horvath J, Wildhaber JH, Zariwala MA, Kennedy M, et al. Mislocalization of DNAH5 and DNAH9 in respiratory cells from patients with primary ciliary dyskinesia. American Journal of Respiratory and Critical Care Medicine. 2005;171(12):1343-9.
323. Bartoloni L, Blouin JL, Maiti AK, Sainsbury A, Rossier C, Gehrig C, et al. Axonemal beta heavy chain dynein DNAH9: cDNA sequence, genomic structure, and investigation of its role in primary ciliary dyskinesia. Genomics. 2001;72(1):21-33.
324. Olbrich H, Haffner K, Kispert A, Volkel A, Volz A, Sasmaz G, et al. Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. Nat Genet. 2002;30(2):143-4.
325. DiBella LM, King SM. Dynein motors of the Chlamydomonas flagellum. International Review of Cytology - a Survey of Cell Biology, Vol 210. 2001;210:227-68.

326. Pifferi M, Michelucci A, Conidi ME, Cangiotti AM, Simi P, Macchia P, et al. New DNAH11 mutations in primary ciliary dyskinesia with normal axonemal ultrastructure. Eur Respir J. 2010;35(6):1413-6.
327. Yang CZ, Asthagiri AR, Iyer RR, Lu J, Xu DS, Ksendzovsky A, et al. Missense mutations in the NF2 gene result in the quantitative loss of merlin protein and minimally affect protein intrinsic function. P Natl Acad Sci USA. 2011;108(12):4980-5.
328. Yang CZ, Matro JC, Huntoon KM, Ye DY, Huynh TT, Fliedner SMJ, et al. Missense mutations in the human SDHB gene increase protein degradation without altering intrinsic enzymatic function. Faseb Journal. 2012;26(11):4506-16.
329. Baralle D, Baralle M. Splicing in action: assessing disease causing sequence changes. J Med Genet. 2005;42(10):737-48.
330. Lykke-Andersen S, Jensen TH. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat Rev Mol Cell Bio. 2015;16(11):665-77.
331. Beckwith SM, Asai DJ. Ciliary Dynein of Paramecium-Tetraurelia - Photolytic Maps of the 3 Heavy-Chains. Cell Motil Cytoskel. 1993;24(1):29-38.
332. Kandl KA, Forney JD, Asai DJ. The Dynein Genes of Paramecium-Tetraurelia - the Structure and Expression of the Ciliary Beta-Chains and Cytoplasmic Heavy-Chains. Mol Biol Cell. 1995;6(11):1549-62.
333. Braun DA, Hildebrandt F. Ciliopathies. Cold Spring Harbor Perspectives in Biology. 2017;9(3).
334. Noda K, Kitami M, Kitami K, Kaku M, Komatsu Y. Canonical and noncanonical intraflagellar transport regulates craniofacial skeletal development. P Natl Acad Sci USA. 2016;113(19):E2589-E97.
335. Oh EC, Katsanis N. Cilia in vertebrate development and disease. Development. 2012;139(3):443-8.
336. Hildebrandt F, Benzing T, Katsanis N. Mechanisms of Disease: Ciliopathies. New England Journal of Medicine. 2011;364(16):1533-43.
337. Taschner M, Lorentzen E. The Intraflagellar Transport Machinery. Cold Spring Harbor Perspectives in Biology. 2016;8(10).
338. Scholey JM. Intraflagellar transport motors in cilia: moving along the cell's antenna. J Cell Biol. 2008;180(1):23-9.
339. Swiderski RE, Nakano Y, Mullins RF, Seo S, Banfi B. A mutation in the mouse ttc26 gene leads to impaired hedgehog signaling. Plos Genet. 2014;10(10):e1004689.
340. Ishikawa H, Ide T, Yagi T, Jiang X, Hirono M, Sasaki H, et al. TTC26/DYF13 is an intraflagellar transport protein required for transport of motility-related proteins into flagella. Elife. 2014;3.
341. Taschner M, Bhogaraju S, Vetter M, Morawetz M, Lorentzen E. Biochemical Mapping of Interactions within the Intraflagellar Transport (IFT) B Core Complex IFT52 BINDS DIRECTLY TO FOUR OTHER IFT-B SUBUNITS. J Biol Chem. 2011;286(30):26344-52.
342. Taschner M, Weber K, Mourao A, Vetter M, Awasthi M, Stiegler M, et al. Intraflagellar transport proteins 172, 80, 57, 54, 38, and 20 form a stable tubulin-binding IFT-B2 complex. Embo J. 2016;35(7):773-90.
343. Ishikawa H, Marshall WF. Ciliogenesis: building the cell's antenna. Nat Rev Mol Cell Bio. 2011;12(4):222-34.

344. Pan XY, Ou GS, Civelekoglu-Scholey G, Blacque OE, Endres NF, Tao L, et al. Mechanism of transport of IFT particles in C-elegans cilia by the concerted action of kinesin-II and OSM-3 motors. J Cell Biol. 2006;174(7):1035-45.
345. Halbritter J, Bizet AA, Schmidts M, Porath JD, Braun DA, Gee HY, et al. Defects in the IFT-B Component IFT172 Cause Jeune and Mainzer-Saldino Syndromes in Humans. Am J Hum Genet. 2013;93(5):915-25.
346. Beales PL, Bland E, Tobin JL, Bacchelli C, Tuysuz B, Hill J, et al. IFT80, which encodes a conserved intraflagellar transport protein, is mutated in Jeune asphyxiating thoracic dystrophy. Nat Genet. 2007;39(6):727-9.
347. Zhang WJ, Taylor SP, Nevarez L, Lachman RS, Nickerson DA, Bamshad M, et al. IFT52 mutations destabilize anterograde complex assembly, disrupt ciliogenesis and result in short rib polydactyly syndrome. Hum Mol Genet. 2016;25(18):4012-20.
348. Duran I, Taylor SP, Zhang WJ, Martin J, Forlenza KN, Spiro RP, et al. Destabilization of the IFT-B cilia core complex due to mutations in IFT81 causes a Spectrum of Short-Rib Polydactyly Syndrome. Sci Rep-Uk. 2016;6.
349. Aldahmesh MA, Li YY, Alhashem A, Anazi S, Alkuraya H, Hashem M, et al. IFT27, encoding a small GTPase component of IFT particles, is mutated in a consanguineous family with Bardet-Biedl syndrome. Hum Mol Genet. 2014;23(12):3307-15.
350. Lindstrand A, Frangakis S, Carvalho CMB, Richardson EB, McFadden KA, Willer JR, et al. Copy-Number Variation Contributes to the Mutational Load of Bardet-Biedl Syndrome. Am J Hum Genet. 2016;99(2):318-36.
351. Arts HH, Bongers EMHF, Mans DA, van Beersum SEC, Oud MM, Bolat E, et al. C14ORF179 encoding IFT43 is mutated in Sensenbrenner syndrome. J Med Genet. 2011;48(6):390-5.
352. Bredrup C, Saunier S, Oud MM, Fiskerstrand T, Hoischen A, Brackman D, et al. Ciliopathies with Skeletal Anomalies and Renal Insufficiency due to Mutations in the IFT-A Gene WDR19. Am J Hum Genet. 2011;89(5):634-43.
353. Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, et al. Exome Sequencing Identifies WDR35 Variants Involved in Sensenbrenner Syndrome. Am J Hum Genet. 2010;87(3):418-23.
354. Mill P, Lockhart PJ, Fitzpatrick E, Mountford HS, Hall EA, Reijns MAM, et al. Human and Mouse Mutations in WDR35 Cause Short-Rib Polydactyly Syndromes Due to Abnormal Ciliogenesis. Am J Hum Genet. 2011;88(4):508-15.
355. Walczak-Sztulpa J, Eggenschwiler J, Osborn D, Brown DA, Emma F, Klingenberg C, et al. Cranioectodermal Dysplasia, Sensenbrenner Syndrome, Is a Ciliopathy Caused by Mutations in the IFT122 Gene. Am J Hum Genet. 2010;86(6):949-56.
356. Perrault I, Saunier S, Hanein S, Filhol E, Bizet AA, Collins F, et al. Mainzer-Saldino Syndrome Is a Ciliopathy Caused by IFT140 Mutations. Am J Hum Genet. 2012;90(5):864-70.
357. Davis EE, Zhang Q, Liu Q, Diplas BH, Davey LM, Hartley J, et al. TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum (vol 43, pg 189, 2011). Nat Genet. 2011;43(5):499-.
358. Cortellino S, Wang C, Wang B, Bassi MR, Caretti E, Champeval D, et al. Defective ciliogenesis, embryonic lethality and severe impairment of the Sonic Hedgehog pathway caused by inactivation of the mouse complex A

intraflagellar transport gene Ift122/Wdr10, partially overlapping with the DNA repair gene Med1/Mbd4. Dev Biol. 2009;325(1):225-37.
359. Zhang Y, Liu H, Li W, Zhang ZG, Zhang SY, Teves ME, et al. Intraflagellar transporter protein 140 (IFT140), a component of IFT-A complex, is essential for male fertility and spermiogenesis in mice. Cytoskeleton. 2018;75(2):70-84.
360. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Research. 2014;42(D1):D986-D92.
361. Vladar EK, Bayly RD, Sangoram AM, Scott MP, Axelrod JD. Microtubules Enable the Planar Cell Polarity of Airway Cilia. Curr Biol. 2012;22(23):2203-12.
362. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nature Methods. 2016;13(9):731-40.
363. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response (vol 98, pg 5116, 2001). P Natl Acad Sci USA. 2001;98(18):10515-.
364. Ryan R, Failler M, Reilly ML, Garfa-Traore M, Delous M, Filhol E, et al. Functional characterization of tektin-1 in motile cilia and evidence for TEKT1 as a new candidate gene for motile ciliopathies. Hum Mol Genet. 2018;27(2):266-82.
365. Thauvin-Robinet C, Thomas S, Sinico M, Aral B, Burglen L, Gigot N, et al. OFD1 mutations in males: phenotypic spectrum and ciliary basal body docking impairment. Clin Genet. 2013;84(1):86-90.
366. Rossi A, Kontarakis Z, Gerri C, Nolte H, Holper S, Kruger M, et al. Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature. 2015;524(7564):230-+.
367. Sztal TE, McKaige EA, Williams C, Ruparelia AA, Bryson-Richardson RJ. Genetic compensation triggered by actin mutation prevents the muscle damage caused by loss of actin protein. Plos Genet. 2018;14(2).
368. Brown JM, Cochran DA, Craige B, Kubo T, Witman GB. Assembly of IFT Trains at the Ciliary Base Depends on IFT74. Curr Biol. 2015;25(12):1583-93.
369. Kubo T, Brown JM, Bellve K, Craige B, Craft JM, Fogarty K, et al. Together, the IFT81 and IFT74 N-termini form the main module for intraflagellar transport of tubulin. Journal of Cell Science. 2016;129(10):2106-19.
370. Kozak M. Pushing the limits of the scanning mechanism for initiation of translation. Gene. 2002;299(1-2):1-34.
371. Bhogaraju S, Cajanek L, Fort C, Blisnick T, Weber K, Taschner M, et al. Molecular Basis of Tubulin Transport Within the Cilium by IFT74 and IFT81. Science. 2013;341(6149):1009-12.
372. Lucker BF, Behal RH, Qin HM, Siron LC, Taggart WD, Rosenbaum JL, et al. Characterization of the intraflagellar transport complex B core - Direct interaction of the IFT81 AND IFT74/72 subunits. J Biol Chem. 2005;280(30):27688-96.
373. Polevoda B, Sherman F. N alpha-terminal acetylation of eukaryotic proteins. J Biol Chem. 2000;275(47):36479-82.

374. Varshavsky A, Bachmair A, Finley D. The N-End Rule of Selective Protein-Turnover and Its Implications. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences. 1987;317(1187):471-.
375. Gonda DK, Bachmair A, Wunning I, Tobias JW, Lane WS, Varshavsky A. Universality and Structure of the N-End Rule. J Biol Chem. 1989;264(28):16700-12.
376. Varshavsky A. The N-End Rule. Cell. 1992;69(5):725-35.
377. Eldeeb M, Fahlman R. The-N-End Rule: The Beginning Determines the End. Protein and Peptide Letters. 2016;23(4):343-8.
378. Nguyen KT, Mun SH, Lee CS, Hwang CS. Control of protein degradation by N-terminal acetylation and the N-end rule pathway. Experimental and Molecular Medicine. 2018;50.
379. Brower CS, Varshavsky A. Ablation of Arginylation in the Mouse N-End Rule Pathway: Loss of Fat, Higher Metabolic Rate, Damaged Spermatogenesis, and Neurological Perturbations. Plos One. 2009;4(11).
380. Xiao Q, Zhang FR, Nacev BA, Liu JO, Pei DH. Protein N-Terminal Processing: Substrate Specificity of Escherichia coli and Human Methionine Aminopeptidases. Biochemistry. 2010;49(26):5588-99.
381. Frottin F, Martinez A, Peynot P, Mitra S, Holz RC, Giglione C, et al. The proteomics of N-terminal methionine cleavage. Molecular & Cellular Proteomics. 2006;5(12):2336-49.
382. Liem KF, Jr., Ashe A, He M, Satir P, Moran J, Beier D, et al. The IFT-A complex regulates Shh signaling through cilia structure and membrane protein trafficking. J Cell Biol. 2012;197(6):789-800.
383. Iomini C, Li L, Esparza JM, Dutcher SK. Retrograde intraflagellar transport mutants identify complex A proteins with multiple genetic interactions in Chlamydomonas reinhardtii. Genetics. 2009;183(3):885-96.
384. Behan L, Dimitrov BD, Kuehni CE, Hogg C, Carroll M, Evans HJ, et al. PICADAR: a diagnostic predictive tool for primary ciliary dyskinesia. Eur Respir J. 2016;47(4):1103-12.
385. Davis SD, Ferkol TW, Rosenfeld M, Lee HS, Dell SD, Sagel SD, et al. Clinical Features of Childhood Primary Ciliary Dyskinesia by Genotype and Ultrastructural Phenotype. Am J Resp Crit Care. 2015;191(3):316-24.
386. Shapiro AJ, Weck KE, Chao KC, Rosenfeld M, Nygren AOH, Knowles MR, et al. Cri du Chat Syndrome and Primary Ciliary Dyskinesia: A Common Genetic Cause on Chromosome 5p. J Pediatr-Us. 2014;165(4):858-61.
387. Guimbellot JS, Shapiro AJ, Weck KE, Chao KC, Zariwala MA, Rosenfeld M, et al. Primary Ciliary Dyskinesia In Children With Cri-Du-Chat Syndrome: A Genetic Connection. Am J Resp Crit Care. 2014;189.
388. Donnellan C, Leigh M, Zariwala M, Knowles M, Wolf W, Campbell D, et al. Cri-Du-Chat Syndrome And Primary Ciliary Dyskinesia: A Genetic And Clinical Connection. Am J Resp Crit Care. 2016;193.
389. Coren ME, Meeks M, Morrison I, Buchdahl RM, Bush A. Primary ciliary dyskinesia: age at diagnosis and symptom history. Acta Paediatr. 2002;91(6):667-9.
390. Ellerman A, Bisgaard H. Longitudinal study of lung function in a cohort of primary ciliary dyskinesia. Eur Respir J. 1997;10(10):2376-9.
391. Collins SA, Walker WT, Lucas JS. Genetic Testing in the Diagnosis of Primary Ciliary Dyskinesia: State-of-the-Art and Future Perspectives. J Clin Med. 2014;3(2):491-503.

392. Olin JT, Burns K, Carson JL, Metjian H, Atkinson JJ, Davis SD, et al. Diagnostic Yield of Nasal Scrape Biopsies in Primary Ciliary Dyskinesia: A Multicenter Experience. Pediatr Pulm. 2011;46(5):483-8.
393. Goutaki M, Meier AB, Halbeisen FS, Lucas JS, Dell SD, Maurer E, et al. Clinical manifestations in primary ciliary dyskinesia: systematic review and meta-analysis. Eur Respir J. 2016;48(4):1081-95.
394. Blyth M, Wellesley D. Ectopic pregnancy in primary ciliary dyskinesia. J Obstet Gynaecol. 2008;28(3):358-.
395. Vanaken GJ, Bassinet L, Boon M, Mani R, Honore I, Papon JF, et al. Infertility in an adult cohort with primary ciliary dyskinesia: phenotype-gene association (vol 50, 1700314, 2017). Eur Respir J. 2017;50(6).
396. Hu X, Li N, Xu Y, Li G, Yu T, Yao RE, et al. Proband-only medical exome sequencing as a cost-effective first-tier genetic diagnostic test for patients without prior molecular tests and clinical diagnosis in a developing country: the China experience. Genetics in medicine : official journal of the American College of Medical Genetics. 2017.
397. Abdulwahab F, Grp SM, Abouelhoda M, Abouthuraya R, Abumansour I, Ahmed SO, et al. Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases (vol 16, 134, 2015). Genome Biol. 2015;16.
398. Zariwala MA, Omran H, Ferkol TW. The emerging genetics of primary ciliary dyskinesia. Proc Am Thorac Soc. 2011;8(5):430-3.
399. Deltas C. Digenic inheritance and genetic modifiers. Clin Genet. 2018;93(3):429-38.
400. Galvani A, Sperling L. RNA interference by feeding in Paramecium. Trends Genet. 2002;18(1):11-2.
401. Driscoll JA, Bhalla S, Liapis H, Ibricevic A, Brody SL. Autosomal dominant polycystic kidney disease is associated with an increased prevalence of radiographic bronchiectasis. Chest. 2008;133(5):1181-8.
402. Escudier E, Couprie M, Duriez B, Roudot-Thoraval F, Millepied MC, Pruliere-Escabasse V, et al. Computer-assisted analysis helps detect inner dynein arm abnormalities. Am J Respir Crit Care Med. 2002;166(9):1257-62.
403. Lai M, Pifferi M, Bush A, Piras M, Michelucci A, Di Cicco M, et al. Gene editing of DNAH11 restores normal cilia motility in primary ciliary dyskinesia. J Med Genet. 2016;53(4):242-9.

Appendix

Table A-0-1 List of genes included in 651 gene motile ciliome panel version

Input	HGNC Approved symbol	Description	Locus
ARMC4	***ARMC4***	armadillo repeat containing 4	10p12.1
C21ORF59	***CFAP298***	cilia and flagella associated protein 298	21q22.11
CCDC103	***CCDC103***	coiled-coil domain containing 103	17q21.31
CCDC114	***CCDC114***	coiled-coil domain containing 114	19q13.33
CCDC151	***CCDC151***	coiled-coil domain containing 151	19p13.2
CCDC39	***CCDC39***	coiled-coil domain containing 39	3q26.33
CCDC40	***CCDC40***	coiled-coil domain containing 40	17q25.3
CCDC65	***CCDC65***	coiled-coil domain containing 65	12q13.12
CCNO	***CCNO***	cyclin O	5q11.2
DNAAF3	***DNAAF3***	dynein axonemal assembly factor 3	19q13.42
DNAH11	***DNAH11***	dynein axonemal heavy chain 11	7p15.3
DNAH5	***DNAH5***	dynein axonemal heavy chain 5	5p15.2
DNAI1	***DNAI1***	dynein axonemal intermediate chain 1	9p13.3
DNAI2	***DNAI2***	dynein axonemal intermediate chain 2	17q25.1
DNAL1	***DNAL1***	dynein axonemal light chain 1	14q24.3
DYX1C1	***DNAAF4***	dynein axonemal assembly factor 4	15q21.3
GAS8	***GAS8***	growth arrest specific 8	16q24.3
HYDIN	***HYDIN***	HYDIN, axonemal central pair apparatus protein	16q22.2
LRRC6	***LRRC6***	leucine rich repeat containing 6	8q24.22
MCIDAS	***MCIDAS***	multiciliate differentiation and DNA synthesis associated cell cycle protein	5q11.2
NME8	***NME8***	NME/NM23 family member 8	7p14.1
RPGR	***RPGR***	retinitis pigmentosa GTPase regulator	Xp11.4
RSPH1	***RSPH1***	radial spoke head component 1	21q22.3
RSPH3	***RSPH3***	radial spoke head 3 homolog	6q25.3
RSPH4A	***RSPH4A***	radial spoke head component 4A	6q22.1
RSPH9	***RSPH9***	radial spoke head 9 homolog	6p21.1
SPAG1	***SPAG1***	sperm associated antigen 1	8q22.2
ZMYND10	***ZMYND10***	zinc finger MYND-type containing 10	3p21.31
DNAAF1	***DNAAF1***	dynein axonemal assembly factor 1	16q24.1
KTU	***DNAAF2***	dynein axonemal assembly factor 2	14q21.3
HEATR2	***DNAAF5***	dynein axonemal assembly factor 5	7p22.3
OFD1	***OFD1***	OFD1, centriole and centriolar satellite protein	Xp22.2
DRC1	***DRC1***	dynein regulatory complex subunit 1	2p23.3

C11orf70	***CFAP300***	cilia and flagella associated protein 300	11q22.1
DNAH9	***DNAH9***	dynein axonemal heavy chain 9	17p12
IFT74	***IFT74***	intraflagellar transport 74	9p21.2
WDR19	***WDR19***	WD repeat domain 19	4p14
DNAH2	***DNAH2***	dynein axonemal heavy chain 2	17p13.1
DNAH3	***DNAH3***	dynein axonemal heavy chain 3	16p12.3
CFAP46	***CFAP46***	cilia and flagella associated protein 46	10q26.3
DNAH14	***DNAH14***	dynein axonemal heavy chain 14	1q42.12
DNAH10	***DNAH10***	dynein axonemal heavy chain 10	12q24.31
DNAH12	***DNAH12***	dynein axonemal heavy chain 12	3p14.3
DNAH17	***DNAH17***	dynein axonemal heavy chain 17	17q25.3
DNAH6	***DNAH6***	dynein axonemal heavy chain 6	2p11.2
DNAH8	***DNAH8***	dynein axonemal heavy chain 8	6p21.2
LRRC48	***DRC3***	dynein regulatory complex subunit 3	17p11.2
LRRC56	***LRRC56***	leucine rich repeat containing 56	11p15.5
ROPN1L	***ROPN1L***	rhophilin associated tail protein 1 like	5p15.2
RSPH10B2	***RSPH10B2***	radial spoke head 10 homolog B2	7p22.1
RSPH6A	***RSPH6A***	radial spoke head 6 homolog A	19q13.3
RUVBL2	***RUVBL2***	RuvB like AAA ATPase 2	19q13.33
SPAG17	***SPAG17***	sperm associated antigen 17	1p12
SPEF2	***SPEF2***	sperm flagellar 2	5p13.2
WDR66	***WDR66***	WD repeat domain 66	12q24.31
ZBBX	***ZBBX***	zinc finger B-box domain containing	3q26.1
CFAP44	***CFAP44***	cilia and flagella associated protein 44	3q13.2
PIH1D3	***PIH1D3***	PIH1 domain containing 3	Xq22.3
RUVBL1	***RUVBL1***	RuvB like AAA ATPase 1	3q21.3
ACVR2B	***ACVR2B***	activin A receptor type 2B	3p22.2
CFC1	***CFC1***	cripto, FRL-1, cryptic family 1	2q21.1
CRELD1	***CRELD1***	cysteine rich with EGF like domains 1	3p25.3
GALNT11	***GALNT11***	polypeptide N-acetylgalactosaminyltransferase 11	7q36.1
GDF1	***GDF1***	growth differentiation factor 1	19p13.11
NEK2	***NEK2***	NIMA related kinase 2	1q32.3
NODAL	***NODAL***	nodal growth differentiation factor	10q22.1
NUP188	***NUP188***	nucleoporin 188	9q34.11
ROCK2	***ROCK2***	Rho associated coiled-coil containing protein kinase 2	2p25.1
ZIC3	***ZIC3***	Zic family member 3	Xq26.3
CFAP53	***CFAP53***	cilia and flagella associated protein 53	18q21.1
CCDC105	***CCDC105***	coiled-coil domain containing 105	19p13.12
CCDC24	***CCDC24***	coiled-coil domain containing 24	1p34.1
CCDC78	***CCDC78***	coiled-coil domain containing 78	16p13.3
WDR38	***WDR38***	WD repeat domain 38	9q33.3
CFAP43	***CFAP43***	cilia and flagella associated protein 43	10q25.1
CFAP45	***CFAP45***	cilia and flagella associated protein 45	1q23.2

CFAP57	***CFAP57***	cilia and flagella associated protein 57	1p34.2
CFAP58	***CFAP58***	cilia and flagella associated protein 58	10q25.1
CFAP221	***CFAP221***	cilia and flagella associated protein 221	2q14.2
PPP1R42	***PPP1R42***	protein phosphatase 1 regulatory subunit 42	8q13.1
RSPH14	***RSPH14***	radial spoke head 14 homolog	22q11.22-q11.23
A1CF	***A1CF***	APOBEC1 complementation factor	10q11.23
ABCG2	***ABCG2***	ATP binding cassette subfamily G member 2 (Junior blood group)	4q22.1
ACTN3	***ACTN3***	actinin alpha 3 (gene/pseudogene)	11q13.2
ACTR3	***ACTR3***	ARP3 actin related protein 3 homolog	2q14.1
ADAMTS12	***ADAMTS12***	ADAM metallopeptidase with thrombospondin type 1 motif 12	5p13.3-p13.2
ADAMTSL5	***ADAMTSL5***	ADAMTS like 5	19p13.3
ADD3	***ADD3***	adducin 3	10q25.1-q25.2
ADH4	***ADH4***	alcohol dehydrogenase 4 (class II), pi polypeptide	4q23
AFTPH	***AFTPH***	aftiphilin	2p14
AGBL2	***AGBL2***	ATP/GTP binding protein like 2	11p11.2
AGFG2	***AGFG2***	ArfGAP with FG repeats 2	7q22.1
AGPAT9	***GPAT3***	glycerol-3-phosphate acyltransferase 3	4q21.23
AHSA1	***AHSA1***	activator of HSP90 ATPase activity 1	14q24.3
AK8	***AK8***	adenylate kinase 8	9q34.13
AMFR	***AMFR***	autocrine motility factor receptor	16q13
AMY2B	***AMY2B***	amylase, alpha 2B (pancreatic)	1p21.1
ANGPTL5	***ANGPTL5***	angiopoietin like 5	11q22.1
ANKAR	***ANKAR***	ankyrin and armadillo repeat containing	2q32.2
ANKMY1	***ANKMY1***	ankyrin repeat and MYND domain containing 1	2q37.3
ANKRD44	***ANKRD44***	ankyrin repeat domain 44	2q33.1
ANKEF1	***ANKEF1***	ankyrin repeat and EF-hand domain containing 1	20p12.2
ANXA2	***ANXA2***	annexin A2	15q22.2
APBB2	***APBB2***	amyloid beta precursor protein binding family B member 2	4p14-p13
APOB	***APOB***	apolipoprotein B	2p24.1
APOE	***APOE***	apolipoprotein E	19q13.32
APOLD1	***APOLD1***	apolipoprotein L domain containing 1	12p13.1
ARAP1	***ARAP1***	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1	11q13.4
ARF4	***ARF4***	ADP ribosylation factor 4	3p14.3
ARHGEF18	***ARHGEF18***	Rho/Rac guanine nucleotide exchange factor 18	19p13.2
ARL15	***ARL15***	ADP ribosylation factor like GTPase 15	5q11.2

ARL6IP1	***ARL6IP1***	ADP ribosylation factor like GTPase 6 interacting protein 1	16p12.3
ARL9	***ARL9***	ADP ribosylation factor like GTPase 9	4q12
ARMC2	***ARMC2***	armadillo repeat containing 2	6q21
ASB14	***ASB14***	ankyrin repeat and SOCS box containing 14	3p14.3
ASB5	***ASB5***	ankyrin repeat and SOCS box containing 5	4q34.2
ASPG	***ASPG***	asparaginase	14q32.33
ATF5	***ATF5***	activating transcription factor 5	19q13.33
ATXN1	***ATXN1***	ataxin 1	6p22.3
AVPR1A	***AVPR1A***	arginine vasopressin receptor 1A	12q14.2
B4GALNT3	***B4GALNT3***	beta-1,4-N-acetyl-galactosaminyltransferase 3	12p13.33
BAIAP2	***BAIAP2***	BAI1 associated protein 2	17q25.3
BARHL2	***BARHL2***	BarH like homeobox 2	1p22.2
BCHE	***BCHE***	butyrylcholinesterase	3q26.1
BMP7	***BMP7***	bone morphogenetic protein 7	20q13.31
BTBD16	***BTBD16***	BTB domain containing 16	10q26.13
TEX36	***TEX36***	testis expressed 36	10q26.13
TBATA	***TBATA***	thymus, brain and testes associated	10q22.1
FAM213A	***PRXL2A***	peroxiredoxin like 2A	10q23.1
C11ORF1	***C11orf1***	chromosome 11 open reading frame 1	11q23.1
C11ORF63	***JHY***	junctional cadherin complex regulator	11q24.1
C11ORF65	***C11orf65***	chromosome 11 open reading frame 65	11q22.3
NOXRED1	***NOXRED1***	NADP dependent oxidoreductase domain containing 1	14q24.3
BBOF1	***BBOF1***	basal body orientation factor 1	14q24.3
CFAP20	***CFAP20***	cilia and flagella associated protein 20	16q21
C17ORF105	***CFAP97D1***	CFAP97 domain containing 1	17q21.31
TPGS1	***TPGS1***	tubulin polyglutamylase complex subunit 1	19p13.3
DZANK1	***DZANK1***	double zinc ribbon and ankyrin repeat domains 1	20p11.23
CFAP61	***CFAP61***	cilia and flagella associated protein 61	20p11.23
C20ORF96	***C20orf96***	chromosome 20 open reading frame 96	20p13
C21ORF33	***GATD3A***	glutamine amidotransferase like class 1 domain containing 3A	21q22.3
C21ORF58	***C21orf58***	chromosome 21 open reading frame 58	21q22.3
CATIP	***CATIP***	ciliogenesis associated TTC17 interacting protein	2q35
CCDC173	***CCDC173***	coiled-coil domain containing 173	2q31.1
C3	***C3***	complement C3	19p13.3
C3ORF67	***C3orf67***	chromosome 3 open reading frame 67	3p14.2
C4ORF22	***CFAP299***	cilia and flagella associated protein 299	4q21.21

C5ORF49	***C5orf49***	chromosome 5 open reading frame 49	5p15.31
ADGB	***ADGB***	androglobin	6q24.3
C6ORF165	***CFAP206***	cilia and flagella associated protein 206	6q15
TBC1D32	***TBC1D32***	TBC1 domain family member 32	6q22.31
TMEM244	***TMEM244***	transmembrane protein 244	6q22.33
C6ORF58	***C6orf58***	chromosome 6 open reading frame 58	6q22.33
CCDC170	***CCDC170***	coiled-coil domain containing 170	6q25.1
FAM221A	***FAM221A***	family with sequence similarity 221 member A	7p15.3
C7ORF62	***TEX47***	testis expressed 47	7q21.13
C8ORF74	***C8orf74***	chromosome 8 open reading frame 74	8p23.1
C9ORF116	***C9orf116***	chromosome 9 open reading frame 116	9q34.3
C9ORF117	***CFAP157***	cilia and flagella associated protein 157	9q34.11
C9ORF171	***CFAP77***	cilia and flagella associated protein 77	9q34.13
CCDC180	***CCDC180***	coiled-coil domain containing 180	9q22.33
C9ORF9	***SPACA9***	sperm acrosome associated 9	9q34.13
CA14	***CA14***	carbonic anhydrase 14	1q21.2
CABP5	***CABP5***	calcium binding protein 5	19q13.33
CACNA1D	***CACNA1D***	calcium voltage-gated channel subunit alpha1 D	3p21.1
CAMK4	***CAMK4***	calcium/calmodulin dependent protein kinase IV	5q22.1
CAPN5	***CAPN5***	calpain 5	11q13.5
CAPN8	***CAPN8***	calpain 8	1q41
CAPS2	***CAPS2***	calcyphosine 2	12q21.1-q21.2
CAPSL	***CAPSL***	calcyphosine like	5p13.2
CASP8	***CASP8***	caspase 8	2q33.1
CCDC64B	***BICDL2***	BICD family like cargo adaptor 2	16p13.3
CCNA1	***CCNA1***	cyclin A1	13q13.3
CCR6	***CCR6***	C-C motif chemokine receptor 6	6q27
CCR7	***CCR7***	C-C motif chemokine receptor 7	17q21.2
CD9	***CD9***	CD9 molecule	12p13.31
CDC14A	***CDC14A***	cell division cycle 14A	1p21.2
CDH18	***CDH18***	cadherin 18	5p14.3
CEP19	***CEP19***	centrosomal protein 19	3q29
CEP44	***CEP44***	centrosomal protein 44	4q34.1
CEP68	***CEP68***	centrosomal protein 68	2p14
CERKL	***CERKL***	ceramide kinase like	2q31.3
CES1	***CES1***	carboxylesterase 1	16q12.2
CFD	***CFD***	complement factor D	19p13.3
CHD4	***CHD4***	chromodomain helicase DNA binding protein 4	12p13.31
CLEC4E	***CLEC4E***	C-type lectin domain family 4 member E	12p13.31

CLIC6	***CLIC6***	chloride intracellular channel 6	21q22.12
CLRN1	***CLRN1***	clarin 1	3q25.1
CNGA4	***CNGA4***	cyclic nucleotide gated channel alpha 4	11p15.4
CNKSR2	***CNKSR2***	connector enhancer of kinase suppressor of Ras 2	Xp22.12
CNTN6	***CNTN6***	contactin 6	3p26.3
COL11A2	***COL11A2***	collagen type XI alpha 2 chain	6p21.32
COL7A1	***COL7A1***	collagen type VII alpha 1 chain	3p21.31
COMT	***COMT***	catechol-O-methyltransferase	22q11.21
CP	***CP***	ceruloplasmin	3q24-q25.1
CPB1	***CPB1***	carboxypeptidase B1	3q24
CPNE2	***CPNE2***	copine 2	16q13
CRIP1	***CRIP1***	cysteine rich protein 1	14q32.33
CRMP1	***CRMP1***	collapsin response mediator protein 1	4p16.2
CRYBB3	***CRYBB3***	crystallin beta B3	22q11.23
CTNND2	***CTNND2***	catenin delta 2	5p15.2
CTSB	***CTSB***	cathepsin B	8p23.1
CTXN2	***CTXN2***	cortexin 2	15q21.1
CXORF58	***CXorf58***	chromosome X open reading frame 58	Xp22.11
CYB5D1	***CYB5D1***	cytochrome b5 domain containing 1	17p13.1
CYP27A1	***CYP27A1***	cytochrome P450 family 27 subfamily A member 1	2q35
CYP2J2	***CYP2J2***	cytochrome P450 family 2 subfamily J member 2	1p32.1
CYP39A1	***CYP39A1***	cytochrome P450 family 39 subfamily A member 1	6p12.3
DBH	***DBH***	dopamine beta-hydroxylase	9q34.2
DCXR	***DCXR***	dicarbonyl and L-xylulose reductase	17q25.3
DDAH1	***DDAH1***	dimethylarginine dimethylaminohydrolase 1	1p22.3
DLG4	***DLG4***	discs large MAGUK scaffold protein 4	17p13.1
DNASE1L1	***DNASE1L1***	deoxyribonuclease 1 like 1	Xq28
DOCK11	***DOCK11***	dedicator of cytokinesis 11	Xq24
DTHD1	***DTHD1***	death domain containing 1	4p14
DTX3L	***DTX3L***	deltex E3 ubiquitin ligase 3L	3q21.1
DUPD1	***DUPD1***	dual specificity phosphatase and pro isomerase domain containing 1	10q22.2
DUSP23	***DUSP23***	dual specificity phosphatase 23	1q23.2
DUSP26	***DUSP26***	dual specificity phosphatase 26	8p12
DZIP1L	***DZIP1L***	DAZ interacting zinc finger protein 1 like	3q22.3
E2F5	***E2F5***	E2F transcription factor 5	8q21.2
ECE2	***ECE2***	endothelin converting enzyme 2	3q27.1
ECEL1	***ECEL1***	endothelin converting enzyme like 1	2q37.1
ECHDC2	***ECHDC2***	enoyl-CoA hydratase domain containing 2	1p32.3

ECT2L	***ECT2L***	epithelial cell transforming 2 like	6q24.1
EEF1A1	***EEF1A1***	eukaryotic translation elongation factor 1 alpha 1	6q13
EFCAB6	***EFCAB6***	EF-hand calcium binding domain 6	22q13.2-q13.31
EFHB	***EFHB***	EF-hand domain family member B	3p24.3
ELF3	***ELF3***	E74 like ETS transcription factor 3	1q32.1
ELMOD1	***ELMOD1***	ELMO domain containing 1	11q22.3
EML5	***EML5***	echinoderm microtubule associated protein like 5	14q31.3
ENO4	***ENO4***	enolase 4	10q25.3
EPB41L3	***EPB41L3***	erythrocyte membrane protein band 4.1 like 3	18p11.31
EPDR1	***EPDR1***	ependymin related 1	7p14.1
EPN3	***EPN3***	epsin 3	17q21.33
EZR	***EZR***	ezrin	6q25.3
FABP2	***FABP2***	fatty acid binding protein 2	4q26
FAIM	***FAIM***	Fas apoptotic inhibitory molecule	3q22.3
FAM166B	***FAM166B***	family with sequence similarity 166 member B	9p13.3
FAM183A	***FAM183A***	family with sequence similarity 183 member A	1p34.2
FAM184A	***FAM184A***	family with sequence similarity 184 member A	6q22.31
FAM19A2	***FAM19A2***	family with sequence similarity 19 member A2, C-C motif chemokine like	12q14.1
FAM47E	***FAM47E***	family with sequence similarity 47 member E	4q21.1
FAM83F	***FAM83F***	family with sequence similarity 83 member F	22q13.1
FANK1	***FANK1***	fibronectin type III and ankyrin repeat domains 1	10q26.2
FBXO15	***FBXO15***	F-box protein 15	18q22.3
FBXO16	***FBXO16***	F-box protein 16	8p21.1
FBXO36	***FBXO36***	F-box protein 36	2q36.3
FEM1A	***FEM1A***	fem-1 homolog A	19p13.3
FFAR2	***FFAR2***	free fatty acid receptor 2	19q13.12
FGF21	***FGF21***	fibroblast growth factor 21	19q13.33
FGFBP2	***FGFBP2***	fibroblast growth factor binding protein 2	4p15.32
FGL2	***FGL2***	fibrinogen like 2	7q11.23
FHAD1	***FHAD1***	forkhead associated phosphopeptide binding domain 1	1p36.21
FIP1L1	***FIP1L1***	factor interacting with PAPOLA and CPSF1	4q12
FKBP5	***FKBP5***	FK506 binding protein 5	6p21.31
FOSB	***FOSB***	FosB proto-oncogene, AP-1 transcription factor subunit	19q13.32
FOXN2	***FOXN2***	forkhead box N2	2p16.3
FSIP1	***FSIP1***	fibrous sheath interacting protein 1	15q14

FUT9	***FUT9***	fucosyltransferase 9	6q16.1
GBP1	***GBP1***	guanylate binding protein 1	1p22.2
GLB1L2	***GLB1L2***	galactosidase beta 1 like 2	11q25
GLIS3	***GLIS3***	GLIS family zinc finger 3	9p24.2
GLUL	***GLUL***	glutamate-ammonia ligase	1q25.3
GNPDA1	***GNPDA1***	glucosamine-6-phosphate deaminase 1	5q31.3
GOLGA3	***GOLGA3***	golgin A3	12q24.33
GOLGA4	***GOLGA4***	golgin A4	3p22.2
GOT1L1	***GOT1L1***	glutamic-oxaloacetic transaminase 1 like 1	8p11.23
GPR132	***GPR132***	G protein-coupled receptor 132	14q32.33
GPR143	***GPR143***	G protein-coupled receptor 143	Xp22.2
GRID2IP	***GRID2IP***	Grid2 interacting protein	7p22.1
GRIN2D	***GRIN2D***	glutamate ionotropic receptor NMDA type subunit 2D	19q13.33
GSN	***GSN***	gelsolin	9q33.2
GYG1	***GYG1***	glycogenin 1	3q24
GZMB	***GZMB***	granzyme B	14q12
GZMH	***GZMH***	granzyme H	14q12
HENMT1	***HENMT1***	HEN methyltransferase 1	1p13.3
HHIPL1	***HHIPL1***	HHIP like 1	14q32
HP	***HP***	haptoglobin	16q22.2
HPRT1	***HPRT1***	hypoxanthine phosphoribosyltransferase 1	Xq26.2-q26.3
HSBP1	***HSBP1***	heat shock factor binding protein 1	16q23.3
HSPB11	***HSPB11***	heat shock protein family B (small) member 11	1p32.3
IFT88	***IFT88***	intraflagellar transport 88	13q12.11
IGF1	***IGF1***	insulin like growth factor 1	12q23.2
IL12B	***IL12B***	interleukin 12B	5q33.3
IL1RAPL2	***IL1RAPL2***	interleukin 1 receptor accessory protein like 2	Xq22.3
ITSN2	***ITSN2***	intersectin 2	2p23.3
KCNC1	***KCNC1***	potassium voltage-gated channel subfamily C member 1	11p15.1
KCNC4	***KCNC4***	potassium voltage-gated channel subfamily C member 4	1p13.3
KCNH6	***KCNH6***	potassium voltage-gated channel subfamily H member 6	17q23.3
KCNIP1	***KCNIP1***	potassium voltage-gated channel interacting protein 1	5q35.1
KCNJ1	***KCNJ1***	potassium voltage-gated channel subfamily J member 1	11q24.3
KHDRBS3	***KHDRBS3***	KH RNA binding domain containing, signal transduction associated 3	8q24.23
KIF17	***KIF17***	kinesin family member 17	1p36.12
KIFAP3	***KIFAP3***	kinesin associated protein 3	1q24.2
KLHL24	***KLHL24***	kelch like family member 24	3q27.1

KLHL33	***KLHL33***	kelch like family member 33	14q11.2
KLK13	***KLK13***	kallikrein related peptidase 13	19q13.41
KNDC1	***KNDC1***	kinase non-catalytic C-lobe domain containing 1	10q26.3
KRT20	***KRT20***	keratin 20	17q21.2
LACE1	***AFG1L***	AFG1 like ATPase	6q21
LMNA	***LMNA***	lamin A/C	1q22
LMX1A	***LMX1A***	LIM homeobox transcription factor 1 alpha	1q23.3
LPIN1	***LPIN1***	lipin 1	2p25.1
LRGUK	***LRGUK***	leucine rich repeats and guanylate kinase domain containing	7q33
LRIT1	***LRIT1***	leucine rich repeat, Ig-like and transmembrane domains 1	10q23.1
LRP1B	***LRP1B***	LDL receptor related protein 1B	2q22.1-q22.2
LRP2BP	***LRP2BP***	LRP2 binding protein	4q35.1
LRRC10B	***LRRC10B***	leucine rich repeat containing 10B	11q12.2
LTBP1	***LTBP1***	latent transforming growth factor beta binding protein 1	2p22.3
MAD2L1BP	***MAD2L1BP***	MAD2L1 binding protein	6p21.1
MAGI3	***MAGI3***	membrane associated guanylate kinase, WW and PDZ domain containing 3	1p13.2
MASP1	***MASP1***	mannan binding lectin serine peptidase 1	3q27.3
MATN1	***MATN1***	matrilin 1	1p35.2
MCHR1	***MCHR1***	melanin concentrating hormone receptor 1	22q13.2
MDGA2	***MDGA2***	MAM domain containing glycosylphosphatidylinositol anchor 2	14q21.3
ME1	***ME1***	malic enzyme 1	6q14.2
MEIG1	***MEIG1***	meiosis/spermiogenesis associated 1	10p13
MEOX2	***MEOX2***	mesenchyme homeobox 2	7p21.2
MFAP4	***MFAP4***	microfibril associated protein 4	17p11.2
MMP23B	***MMP23B***	matrix metallopeptidase 23B	1p36.33
MORN2	***MORN2***	MORN repeat containing 2	2p22.1
MORN3	***MORN3***	MORN repeat containing 3	12q24.31
MORN5	***MORN5***	MORN repeat containing 5	9q33.2
MPEG1	***MPEG1***	macrophage expressed 1	11q12.1
MSLN	***MSLN***	mesothelin	16p13.3
MSRB2	***MSRB2***	methionine sulfoxide reductase B2	10p12.2
MTMR7	***MTMR7***	myotubularin related protein 7	8p22
MTTP	***MTTP***	microsomal triglyceride transfer protein	4q23
MUSTN1	***MUSTN1***	musculoskeletal, embryonic nuclear protein 1	3p21.1
MYCBP	***MYCBP***	MYC binding protein	1p34.3
MYCBPAP	***MYCBPAP***	MYCBP associated protein	17q21.33

MYO15A	***MYO15A***	myosin XVA	17p11.2
MYO9A	***MYO9A***	myosin IXA	15q23
MYRIP	***MYRIP***	myosin VIIA and Rab interacting protein	3p22.1
NANOS3	***NANOS3***	nanos C2HC-type zinc finger 3	19p13.13
NCS1	***NCS1***	neuronal calcium sensor 1	9q34.11
NEK3	***NEK3***	NIMA related kinase 3	13q14.3
NELL2	***NELL2***	neural EGFL like 2	12q12
NET1	***NET1***	neuroepithelial cell transforming 1	10p15.1
NEURL3	***NEURL3***	neuralized E3 ubiquitin protein ligase 3	2q11.2
NIPSNAP1	***NIPSNAP1***	nipsnap homolog 1	22q12.2
NLGN3	***NLGN3***	neuroligin 3	Xq13.1
NOTUM	***NOTUM***	notum, palmitoleoyl-protein carboxylesterase	17q25.3
NPNT	***NPNT***	nephronectin	4q24
NRN1L	***NRN1L***	neuritin 1 like	16q22.1
ODF3B	***ODF3B***	outer dense fiber of sperm tails 3B	22q13.33
OGDHL	***OGDHL***	oxoglutarate dehydrogenase like	10q11.23
OPN4	***OPN4***	opsin 4	10q23.2
OSBPL3	***OSBPL3***	oxysterol binding protein like 3	7p15.3
OVGP1	***OVGP1***	oviductal glycoprotein 1	1p13.2
PAQR3	***PAQR3***	progestin and adipoQ receptor family member 3	4q21.21
PARP6	***PARP6***	poly(ADP-ribose) polymerase family member 6	15q23
PAX8	***PAX8***	paired box 8	2q14.1
PBXIP1	***PBXIP1***	PBX homeobox interacting protein 1	1q21.3
PCNT	***PCNT***	pericentrin	21q22.3
PCNXL2	***PCNX2***	pecanex 2	1q42.2
PCSK2	***PCSK2***	proprotein convertase subtilisin/kexin type 2	20p12.1
PDE6C	***PDE6C***	phosphodiesterase 6C	10q23.33
PDE9A	***PDE9A***	phosphodiesterase 9A	21q22.3
PDLIM2	***PDLIM2***	PDZ and LIM domain 2	8p21.3
PELI1	***PELI1***	pellino E3 ubiquitin protein ligase 1	2p14
PFKFB2	***PFKFB2***	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2	1q32.1
PIAS4	***PIAS4***	protein inhibitor of activated STAT 4	19p13.3
PIM3	***PIM3***	Pim-3 proto-oncogene, serine/threonine kinase	22q13.33
PKP2	***PKP2***	plakophilin 2	12p11.21
PLA2G4C	***PLA2G4C***	phospholipase A2 group IVC	19q13.33
PLCXD2	***PLCXD2***	phosphatidylinositol specific phospholipase C X domain containing 2	3q13.2
PLSCR1	***PLSCR1***	phospholipid scramblase 1	3q24
PLXNC1	***PLXNC1***	plexin C1	12q22

PPP1R15B	***PPP1R15B***	protein phosphatase 1 regulatory subunit 15B	1q32.1
PPP4R4	***PPP4R4***	protein phosphatase 4 regulatory subunit 4	14q32.2
PRSS1	***PRSS1***	serine protease 1	7q34
PTGER4	***PTGER4***	prostaglandin E receptor 4	5p13.1
PTPDC1	***PTPDC1***	protein tyrosine phosphatase domain containing 1	9q22.32
PTPRS	***PTPRS***	protein tyrosine phosphatase, receptor type S	19p13.3
PVALB	***PVALB***	parvalbumin	22q12.3
PYY	***PYY***	peptide YY	17q21.31
RAB36	***RAB36***	RAB36, member RAS oncogene family	22q11.23
RABGAP1L	***RABGAP1L***	RAB GTPase activating protein 1 like	1q25.1
RASGRP2	***RASGRP2***	RAS guanyl releasing protein 2	11q13.1
RBM11	***RBM11***	RNA binding motif protein 11	21q11.2
RBM33	***RBM33***	RNA binding motif protein 33	7q36.3
RBM38	***RBM38***	RNA binding motif protein 38	20q13.31
RCAN1	***RCAN1***	regulator of calcineurin 1	21q22.12
RD3	***RD3***	retinal degeneration 3, GUCY2D regulator	1q32.3
RDH8	***RDH8***	retinol dehydrogenase 8	19p13.2
RECQL	***RECQL***	RecQ like helicase	12p12.1
REG1A	***REG1A***	regenerating family member 1 alpha	2p12
RERG	***RERG***	RAS like estrogen regulated growth inhibitor	12p12.3
RFX2	***RFX2***	regulatory factor X2	19p13.3
RIBC1	***RIBC1***	RIB43A domain with coiled-coils 1	Xp11.22
RIC8A	***RIC8A***	RIC8 guanine nucleotide exchange factor A	11p15.5
RILPL2	***RILPL2***	Rab interacting lysosomal protein like 2	12q24.31
RLN3	***RLN3***	relaxin 3	19p13.2
RNF138	***RNF138***	ring finger protein 138	18q12.1
RNF207	***RNF207***	ring finger protein 207	1p36.31
RNF32	***RNF32***	ring finger protein 32	7q36.3
RRM2B	***RRM2B***	ribonucleotide reductase regulatory TP53 inducible subunit M2B	8q22.3
RXRB	***RXRB***	retinoid X receptor beta	6p21.32
S100A11	***S100A11***	S100 calcium binding protein A11	1q21.3
S100P	***S100P***	S100 calcium binding protein P	4p16.1
S100Z	***S100Z***	S100 calcium binding protein Z	5q13.3
SCG5	***SCG5***	secretogranin V	15q13.3
SCN5A	***SCN5A***	sodium voltage-gated channel alpha subunit 5	3p22.2
SDR42E2	***SDR42E2***	short chain dehydrogenase/reductase family 42E, member 2	16p12.1

SEC24C	***SEC24C***	SEC24 homolog C, COPII coat complex component	10q22.2
SESN1	***SESN1***	sestrin 1	6q21
SEZ6L	***SEZ6L***	seizure related 6 homolog like	22q12.1
SHANK3	***SHANK3***	SH3 and multiple ankyrin repeat domains 3	22q13.33
SLC13A1	***SLC13A1***	solute carrier family 13 member 1	7q31.32
SLC16A7	***SLC16A7***	solute carrier family 16 member 7	12q14.1
SLC25A4	***SLC25A4***	solute carrier family 25 member 4	4q35.1
SLC2A10	***SLC2A10***	solute carrier family 2 member 10	20q13.12
SLC3A1	***SLC3A1***	solute carrier family 3 member 1	2p21
SLC43A3	***SLC43A3***	solute carrier family 43 member 3	11q12.1
SLC44A2	***SLC44A2***	solute carrier family 44 member 2	19p13.2
SLC44A4	***SLC44A4***	solute carrier family 44 member 4	6p21.33
SLC7A11	***SLC7A11***	solute carrier family 7 member 11	4q28.3
SMC1A	***SMC1A***	structural maintenance of chromosomes 1A	Xp11.22
SNED1	***SNED1***	sushi, nidogen and EGF like domains 1	2q37.3
SOCS3	***SOCS3***	suppressor of cytokine signaling 3	17q25.3
SPATA17	***SPATA17***	spermatogenesis associated 17	1q41
SPG20	***SPART***	spartin	13q13.3
SPINK2	***SPINK2***	serine peptidase inhibitor, Kazal type 2	4q12
SSX2IP	***SSX2IP***	SSX family member 2 interacting protein	1p22.3
STAB1	***STAB1***	stabilin 1	3p21.1
STAR	***STAR***	steroidogenic acute regulatory protein	8p11.23
STK38	***STK38***	serine/threonine kinase 38	6p21.31
STOM	***STOM***	stomatin	9q33.2
STX19	***STX19***	syntaxin 19	3q11.2
SULT1A1	***SULT1A1***	sulfotransferase family 1A member 1	16p11.2
SYBU	***SYBU***	syntabulin	8q23.2
SYCE2	***SYCE2***	synaptonemal complex central element protein 2	19p13.13
SYNC	***SYNC***	syncoilin, intermediate filament protein	1p35.1
SYNCRIP	***SYNCRIP***	synaptotagmin binding cytoplasmic RNA interacting protein	6q14.3
SYNM	***SYNM***	synemin	15q26.3
SYT2	***SYT2***	synaptotagmin 2	1q32.1
TATDN1	***TATDN1***	TatD DNase domain containing 1	8q24.13
TBC1D14	***TBC1D14***	TBC1 domain family member 14	4p16.1
TBC1D19	***TBC1D19***	TBC1 domain family member 19	4p15.2
TBCE	***TBCE***	tubulin folding cofactor E	1q42.3
TCERG1L	***TCERG1L***	transcription elongation regulator 1 like	10q26.3
TCHH	***TCHH***	trichohyalin	1q21.3
TCTEX1D1	***TCTEX1D1***	Tctex1 domain containing 1	1p31.3

TEC	***TEC***	tec protein tyrosine kinase	4p12-p11
TECR	***TECR***	trans-2,3-enoyl-CoA reductase	19p13.12
TECTA	***TECTA***	tectorin alpha	11q23.3
TGM1	***TGM1***	transglutaminase 1	14q12
TJP2	***TJP2***	tight junction protein 2	9q21.11
TMC5	***TMC5***	transmembrane channel like 5	16p12.3
TMEM106A	***TMEM106A***	transmembrane protein 106A	17q21.31
TMEM107	***TMEM107***	transmembrane protein 107	17p13.1
TMEM232	***TMEM232***	transmembrane protein 232	5q22.1
TMEM63C	***TMEM63C***	transmembrane protein 63C	14q24.3
TMIE	***TMIE***	transmembrane inner ear	3p21.31
TMPRSS13	***TMPRSS13***	transmembrane serine protease 13	11q23.3
TNFAIP1	***TNFAIP1***	TNF alpha induced protein 1	17q11.2
TNNT1	***TNNT1***	troponin T1, slow skeletal type	19q13.4
TOB2	***TOB2***	transducer of ERBB2, 2	22q13.2
TOMM34	***TOMM34***	translocase of outer mitochondrial membrane 34	20q13.12
TP53BP2	***TP53BP2***	tumor protein p53 binding protein 2	1q41
TRAK1	***TRAK1***	trafficking kinesin protein 1	3p22.1
TRIM35	***TRIM35***	tripartite motif containing 35	8p21.2
TRIM40	***TRIM40***	tripartite motif containing 40	6p22.1
TRPM2	***TRPM2***	transient receptor potential cation channel subfamily M member 2	21q22.3
TSGA10	***TSGA10***	testis specific 10	2q11.2
TSN	***TSN***	translin	2q14.3
TSNAXIP1	***TSNAXIP1***	translin associated factor X interacting protein 1	16q22.1
TTC21A	***TTC21A***	tetratricopeptide repeat domain 21A	3p22.2
TTC23	***TTC23***	tetratricopeptide repeat domain 23	15q26.3
TTC26	***TTC26***	tetratricopeptide repeat domain 26	7q34
TTLL11	***TTLL11***	tubulin tyrosine ligase like 11	9q33.2
TTLL3	***TTLL3***	tubulin tyrosine ligase like 3	3p25.3
TTLL6	***TTLL6***	tubulin tyrosine ligase like 6	17q21.32
TTLL7	***TTLL7***	tubulin tyrosine ligase like 7	1p31.1
TUBA1A	***TUBA1A***	tubulin alpha 1a	12q13.12
UBE2E3	***UBE2E3***	ubiquitin conjugating enzyme E2 E3	2q31.3
UBXN11	***UBXN11***	UBX domain protein 11	1p36.11
ULK4	***ULK4***	unc-51 like kinase 4	3p22.1
UNC13D	***UNC13D***	unc-13 homolog D	17q25.3
VWA3A	***VWA3A***	von Willebrand factor A domain containing 3A	16p12.2
WDR27	***WDR27***	WD repeat domain 27	6q27
WNT6	***WNT6***	Wnt family member 6	2q35
XAF1	***XAF1***	XIAP associated factor 1	17p13.2
ZBTB22	***ZBTB22***	zinc finger and BTB domain containing 22	6p21.3

ZDHHC8	***ZDHHC8***	zinc finger DHHC-type containing 8	22q11.21
ZP1	***ZP1***	zona pellucida glycoprotein 1	11q12.2
ZYG11B	***ZYG11B***	zyg-11 family member B, cell cycle regulator	1p32.3
ASIC1	***ASIC1***	acid sensing ion channel subunit 1	12q13.12
ASIC3	***ASIC3***	acid sensing ion channel subunit 3	7q36.1
CEP126	***CEP126***	centrosomal protein 126	11q22.1
CERS3	***CERS3***	ceramide synthase 3	15q26.3
CFAP70	***CFAP70***	cilia and flagella associated protein 70	10q22.2
CRACR2A	***CRACR2A***	calcium release activated channel regulator 2A	12p13.32
EVA1B	***EVA1B***	eva-1 homolog B	1p34.3
GMNC	***GMNC***	geminin coiled-coil domain containing	3q28
IFT22	***IFT22***	intraflagellar transport 22	7q22.1
IZUMO1R	***IZUMO1R***	IZUMO1 receptor, JUNO	11q21
KIAA1407	***CCDC191***	coiled-coil domain containing 191	3q13.31
KIZ	***KIZ***	kizuna centrosomal protein	20p11.23
PALD1	***PALD1***	phosphatase domain containing, paladin 1	10q22.1
SAXO2	***SAXO2***	stabilizer of axonemal microtubules 2	15q25.2
ZC2HC1C	***ZC2HC1C***	zinc finger C2HC-type containing 1C	14q24.3
ANK2	***ANK2***	ankyrin 2	4q25-q26
CFAP54	***CFAP54***	cilia and flagella associated protein 54	12q23.1
C1ORF192	***CFAP126***	cilia and flagella associated protein 126	1q23.3
DNAH1	***DNAH1***	dynein axonemal heavy chain 1	3p21.1
DNAH7	***DNAH7***	dynein axonemal heavy chain 7	2q32.3
DYNLL1	***DYNLL1***	dynein light chain LC8-type 1	12q24.31
DYNLL2	***DYNLL2***	dynein light chain LC8-type 2	17q22
DYNLT1	***DYNLT1***	dynein light chain Tctex-type 1	6q25.3
PIH1D1	***PIH1D1***	PIH1 domain containing 1	19q13.33
PIH1D2	***PIH1D2***	PIH1 domain containing 2	11q23.1
TTC29	***TTC29***	tetratricopeptide repeat domain 29	4q31.22
WDR63	***WDR63***	WD repeat domain 63	1p22.3
WDR78	***WDR78***	WD repeat domain 78	1p31.3
ZMYND12	***ZMYND12***	zinc finger MYND-type containing 12	1p34.2
CFAP36	***CFAP36***	cilia and flagella associated protein 36	2p16.1
DAW1	***DAW1***	dynein assembly factor with WD repeats 1	2q36.3
AK7	***AK7***	adenylate kinase 7	14q32.2
BICC1	***BICC1***	BicC family RNA binding protein 1	10q21.1
C15ORF26	***CFAP161***	cilia and flagella associated protein 161	15q25.1
MAATS1	***MAATS1***	MYCBP associated and testis expressed 1	3q13.33
CASC1	***CASC1***	cancer susceptibility 1	12p12.1
CCDC108	***CFAP65***	cilia and flagella associated protein 65	2q35

CCDC113	***CCDC113***	coiled-coil domain containing 113	16q21
CCDC96	***CCDC96***	coiled-coil domain containing 96	4p16.1
CETN2	***CETN2***	centrin 2	Xq28
DLEC1	***DLEC1***	DLEC1, cilia and flagella associated protein	3p22.2
DNAJB13	***DNAJB13***	DnaJ heat shock protein family (Hsp40) member B13	11q13.4
DNALI1	***DNALI1***	dynein axonemal light intermediate chain 1	1p34.3
DYNLRB2	***DYNLRB2***	dynein light chain roadblock-type 2	16q23.2
EFCAB1	***EFCAB1***	EF-hand calcium binding domain 1	8q11.21
EFHC1	***EFHC1***	EF-hand domain containing 1	6p12.2
EFHC2	***EFHC2***	EF-hand domain containing 2	Xp11.3
FOXJ1	***FOXJ1***	forkhead box J1	17q25.1
IQCG	***IQCG***	IQ motif containing G	3q29
IQUB	***IQUB***	IQ motif and ubiquitin domain containing	7q31.32
KIF9	***KIF9***	kinesin family member 9	3p21.31
LRRC23	***LRRC23***	leucine rich repeat containing 23	12p13.31
LRRC34	***LRRC34***	leucine rich repeat containing 34	3q26.2
MDH1B	***MDH1B***	malate dehydrogenase 1B	2q33.3
MNS1	***MNS1***	meiosis specific nuclear structural 1	15q21.3
NME5	***NME5***	NME/NM23 family member 5	5q31.2
PACRG	***PACRG***	parkin coregulated	6q26
RSPH10B	***RSPH10B***	radial spoke head 10 homolog B	7p22.1
SPA17	***SPA17***	sperm autoantigenic protein 17	11q24.2
SPAG6	***SPAG6***	sperm associated antigen 6	10p12.2
SPAG8	***SPAG8***	sperm associated antigen 8	9p13.3
SPATA18	***SPATA18***	spermatogenesis associated 18	4q12
SPEF1	***SPEF1***	sperm flagellar 1	20p13
STK33	***STK33***	serine/threonine kinase 33	11p15.4
STK36	***STK36***	serine/threonine kinase 36	2q35
TCTE1	***TCTE1***	t-complex-associated-testis-expressed 1	6p21.1
TEKT1	***TEKT1***	tektin 1	17p13.1
TEKT2	***TEKT2***	tektin 2	1p34.3
TEKT4	***TEKT4***	tektin 4	2q11.1
TPPP3	***TPPP3***	tubulin polymerization promoting protein family member 3	16q22.1
TTC25	***TTC25***	tetratricopeptide repeat domain 25	17q21.2
TTC6	***TTC6***	tetratricopeptide repeat domain 6	14q21.1
TTLL9	***TTLL9***	tubulin tyrosine ligase like 9	20q11
AKAP9	***AKAP9***	A-kinase anchoring protein 9	7q21.2
ALDOC	***ALDOC***	aldolase, fructose-bisphosphate C	17q11.2
ANK3	***ANK3***	ankyrin 3	10q21.2
C10ORF53	***C10orf53***	chromosome 10 open reading frame 53	10q11.23

C4ORF47	***C4orf47***	chromosome 4 open reading frame 47	4q35.1
C9ORF135	***C9orf135***	chromosome 9 open reading frame 135	9q21.12
CCP110	***CCP110***	centriolar coiled-coil protein 110	16p12.3
DRC7	***DRC7***	dynein regulatory complex subunit 7	16q21
EML1	***EML1***	echinoderm microtubule associated protein like 1	14q32.2
GOLGB1	***GOLGB1***	golgin B1	3q13.33
IFT57	***IFT57***	intraflagellar transport 57	3q13.12-q13.13
IFT81	***IFT81***	intraflagellar transport 81	12q24.11
KIF27	***KIF27***	kinesin family member 27	9q21.32
KIF3A	***KIF3A***	kinesin family member 3A	5q31.1
MAK	***MAK***	male germ cell associated kinase	6p24.2
MLF1	***MLF1***	myeloid leukemia factor 1	3q25.32
NEK11	***NEK11***	NIMA related kinase 11	3q22.1
NEK5	***NEK5***	NIMA related kinase 5	13q14.3
OSCP1	***OSCP1***	organic solute carrier partner 1	1p34.3
PIFO	***PIFO***	primary cilia formation	1p13.2
PROM1	***PROM1***	prominin 1	4p15.32
PTCH1	***PTCH1***	patched 1	9q22.32
RABL2A	***RABL2A***	RAB, member of RAS oncogene family like 2A	2q14.1
RFX3	***RFX3***	regulatory factor X3	9p24.2
SPAG16	***SPAG16***	sperm associated antigen 16	2q34
SPAG9	***SPAG9***	sperm associated antigen 9	17q21.33
TCP11	***TCP11***	t-complex 11	6p21.31
TRAF3IP1	***TRAF3IP1***	TRAF3 interacting protein 1	2q37.3
TTC30B	***TTC30B***	tetratricopeptide repeat domain 30B	2q31.2
C22ORF23	***C22orf23***	chromosome 22 open reading frame 23	22q13.1
CCDC42B	***CFAP73***	cilia and flagella associated protein 73	12q24.13
DAB1	***DAB1***	DAB1, reelin adaptor protein	1p32.2
ENKD1	***ENKD1***	enkurin domain containing 1	16q22.1
ENKUR	***ENKUR***	enkurin, TRPC channel interacting protein	10p12.1
KATNA1	***KATNA1***	katanin catalytic subunit A1	6q25.1
KATNAL2	***KATNAL2***	katanin catalytic subunit A1 like 2	18q21.1
TTC12	***TTC12***	tetratricopeptide repeat domain 12	11q23.2
WDR92	***WDR92***	WD repeat domain 92	2p14
CFAP52	***CFAP52***	cilia and flagella associated protein 52	17p13.1
CFAP97	***CFAP97***	cilia and flagella associated protein 97	4q35.1
MAPK15	***MAPK15***	mitogen-activated protein kinase 15	8q24.3
AHI1	***AHI1***	Abelson helper integration site 1	6q23.3
ANKS6	***ANKS6***	ankyrin repeat and sterile alpha motif domain containing 6	9q22.33

ARL13B	***ARL13B***	ADP ribosylation factor like GTPase 13B	3q11.1-q11.2
C21ORF2	***CFAP410***	cilia and flagella associated protein 410	21q22.3
C2CD3	***C2CD3***	C2 calcium dependent domain containing 3	11q13.4
C5ORF42	***CPLANE1***	ciliogenesis and planar polarity effector 1	5p13.2
CC2D2A	***CC2D2A***	coiled-coil and C2 domain containing 2A	4p15.32
CEP104	***CEP104***	centrosomal protein 104	1p36.32
CEP120	***CEP120***	centrosomal protein 120	5q23.2
CEP290	***CEP290***	centrosomal protein 290	12q21.32
CEP83	***CEP83***	centrosomal protein 83	12q22
CSPP1	***CSPP1***	centrosome and spindle pole associated protein 1	8q13.1-q13.2
DDX59	***DDX59***	DEAD-box helicase 59	1q32.1
DYNC2H1	***DYNC2H1***	dynein cytoplasmic 2 heavy chain 1	11q22.3
DYNC2LI1	***DYNC2LI1***	dynein cytoplasmic 2 light intermediate chain 1	2p21
EVC	***EVC***	EvC ciliary complex subunit 1	4p16.2
EVC2	***EVC2***	EvC ciliary complex subunit 2	4p16.2
GLI3	***GLI3***	GLI family zinc finger 3	7p14.1
HNF1B	***HNF1B***	HNF1 homeobox B	17q12
IFT122	***IFT122***	intraflagellar transport 122	3q21.3-q22.1
IFT140	***IFT140***	intraflagellar transport 140	16p13.3
IFT172	***IFT172***	intraflagellar transport 172	2p23.3
IFT27	***IFT27***	intraflagellar transport 27	22q12.3
IFT43	***IFT43***	intraflagellar transport 43	14q24.3
IFT80	***IFT80***	intraflagellar transport 80	3q25.33
INPP5E	***INPP5E***	inositol polyphosphate-5-phosphatase E	9q34.3
KIF7	***KIF7***	kinesin family member 7	15q26.1
LBR	***LBR***	lamin B receptor	1q42.12
MUC1	***MUC1***	mucin 1, cell surface associated	1q22
NEK1	***NEK1***	NIMA related kinase 1	4q33
NEK8	***NEK8***	NIMA related kinase 8	17q11.1
NPHP1	***NPHP1***	nephrocystin 1	2q13
NPHP3	***NPHP3***	nephrocystin 3	3q22.1
NPHP4	***NPHP4***	nephrocystin 4	1p36.31
PDE6D	***PDE6D***	phosphodiesterase 6D	2q37.1
PIBF1	***PIBF1***	progesterone immunomodulatory binding factor 1	13q21.33-q22.1
POC1B	***POC1B***	POC1 centriolar protein B	12q21.33
RPGRIP1L	***RPGRIP1L***	RPGRIP1 like	16q12.2
SBDS	***SBDS***	SBDS, ribosome maturation factor	7q11.21

SDCCAG8	***SDCCAG8***	serologically defined colon cancer antigen 8	1q43-q44
TCTEX1D2	***TCTEX1D2***	Tctex1 domain containing 2	3q29
TCTN1	***TCTN1***	tectonic family member 1	12q24.11
TCTN3	***TCTN3***	tectonic family member 3	10q24.1
TMEM138	***TMEM138***	transmembrane protein 138	11q12.2
TMEM216	***TMEM216***	transmembrane protein 216	11q13.1
TMEM231	***TMEM231***	transmembrane protein 231	16q23.1
TMEM237	***TMEM237***	transmembrane protein 237	2q33.1
TMEM67	***TMEM67***	transmembrane protein 67	8q22.1
TTC21B	***TTC21B***	tetratricopeptide repeat domain 21B	2q24.3
UMOD	***UMOD***	uromodulin	16p12.3
WDR34	***WDR34***	WD repeat domain 34	9q34.11
WDR35	***WDR35***	WD repeat domain 35	2p24.1
WDR60	***WDR60***	WD repeat domain 60	7q36.3
ZNF423	***ZNF423***	zinc finger protein 423	16q12.1
CEP164	***CEP164***	centrosomal protein 164	11q23.3
CEP41	***CEP41***	centrosomal protein 41	7q32.2
GLIS2	***GLIS2***	GLIS family zinc finger 2	16p13.3
INVS	***INVS***	inversin	9q31.1
IQCB1	***IQCB1***	IQ motif containing B1	3q13.33
SPATA7	***SPATA7***	spermatogenesis associated 7	14q31.3
TCTN2	***TCTN2***	tectonic family member 2	12q24.31

The gene list are colour coded when blue colour indicates known PCD genes, green colour indicates newly discovered genes implicated in PCD with functional confirmation in human and model organisms, and yellow colour indicates genes with putative mutations in PCD patients without functional characterization so far.

Table A-0-2 List of genes included in the 321 gene motile ciliome panel version

Gene ID	HGNC Approved symbol	Description	Locus
ARMC4	***ARMC4***	armadillo repeat containing 4	10p12.1
C21orf59	***CFAP298***	cilia and flagella associated protein 298	21q22.11
CCDC103	***CCDC103***	coiled-coil domain containing 103	17q21.31
CCDC114	***CCDC114***	coiled-coil domain containing 114	19q13.33
CCDC151	***CCDC151***	coiled-coil domain containing 151	19p13.2
CCDC39	***CCDC39***	coiled-coil domain containing 39	3q26.33
CCDC40	***CCDC40***	coiled-coil domain containing 40	17q25.3
CCDC65	***CCDC65***	coiled-coil domain containing 65	12q13.12
CCNO	***CCNO***	cyclin O	5q11.2
DNAAF3	***DNAAF3***	dynein axonemal assembly factor 3	19q13.42
DNAH11	***DNAH11***	dynein axonemal heavy chain 11	7p15.3
DNAH5	***DNAH5***	dynein axonemal heavy chain 5	5p15.2
DNAI1	***DNAI1***	dynein axonemal intermediate chain 1	9p13.3
DNAI2	***DNAI2***	dynein axonemal intermediate chain 2	17q25.1
DNAL1	***DNAL1***	dynein axonemal light chain 1	14q24.3
DYX1C1	***DNAAF4***	dynein axonemal assembly factor 4	15q21.3
GAS8	***GAS8***	growth arrest specific 8	16q24.3
HYDIN	***HYDIN***	HYDIN, axonemal central pair apparatus protein	16q22.2
LRRC6	***LRRC6***	leucine rich repeat containing 6	8q24.22
MCIDAS	***MCIDAS***	multiciliate differentiation and DNA synthesis associated cell cycle protein	5q11.2
NME8	***NME8***	NME/NM23 family member 8	7p14.1
RPGR	***RPGR***	retinitis pigmentosa GTPase regulator	Xp11.4
RSPH1	***RSPH1***	radial spoke head component 1	21q22.3
RSPH3	***RSPH3***	radial spoke head 3 homolog	6q25.3
RSPH4A	***RSPH4A***	radial spoke head component 4A	6q22.1
RSPH9	***RSPH9***	radial spoke head 9 homolog	6p21.1
SPAG1	***SPAG1***	sperm associated antigen 1	8q22.2
ZMYND10	***ZMYND10***	zinc finger MYND-type containing 10	3p21.31
DNAAF1	***DNAAF1***	dynein axonemal assembly factor 1	16q24.1
KTU	***DNAAF2***	dynein axonemal assembly factor 2	14q21.3
HEATR2	***DNAAF5***	dynein axonemal assembly factor 5	7p22.3
DRC1	***DRC1***	dynein regulatory complex subunit 1	2p23.3
OFD1	***OFD1***	OFD1, centriole and centriolar satellite protein	Xp22.2
PIH1D3	***PIH1D3***	PIH1 domain containing 3	Xq22.3
DNAJB13	***DNAJB13***	DnaJ heat shock protein family (Hsp40) member B13	11q13.4
C11orf70	***CFAP300***	cilia and flagella associated protein 300	11q22.1

DNAH9	***DNAH9***	dynein axonemal heavy chain 9	17p12
IFT74	***IFT74***	intraflagellar transport 74	9p21.2
WDR19	***WDR19***	WD repeat domain 19	4p14
NME7	***NME7***	NME/NM23 family member 7	1q24.2
DNAH2	***DNAH2***	dynein axonemal heavy chain 2	17p13.1
DNAH3	***DNAH3***	dynein axonemal heavy chain 3	16p12.3
CFAP46	***CFAP46***	cilia and flagella associated protein 46	10q26.3
DNAH14	***DNAH14***	dynein axonemal heavy chain 14	1q42.12
DNAH10	***DNAH10***	dynein axonemal heavy chain 10	12q24.31
DNAH12	***DNAH12***	dynein axonemal heavy chain 12	3p14.3
DNAH17	***DNAH17***	dynein axonemal heavy chain 17	17q25.3
DNAH6	***DNAH6***	dynein axonemal heavy chain 6	2p11.2
DNAH8	***DNAH8***	dynein axonemal heavy chain 8	6p21.2
LRRC48	***DRC3***	dynein regulatory complex subunit 3	17p11.2
LRRC56	***LRRC56***	leucine rich repeat containing 56	11p15.5
ROPN1L	***ROPN1L***	rhophilin associated tail protein 1 like	5p15.2
RSPH10B2	***RSPH10B2***	radial spoke head 10 homolog B2	7p22.1
RSPH6A	***RSPH6A***	radial spoke head 6 homolog A	19q13.3
STK36	***STK36***	serine/threonine kinase 36	2q35
RUVBL2	***RUVBL2***	RuvB like AAA ATPase 2	19q13.33
SPAG17	***SPAG17***	sperm associated antigen 17	1p12
SPEF2	***SPEF2***	sperm flagellar 2	5p13.2
WDR66	***WDR66***	WD repeat domain 66	12q24.31
ZBBX	***ZBBX***	zinc finger B-box domain containing	3q26.1
CFAP44	***CFAP44***	cilia and flagella associated protein 44	3q13.2
RUVBL1	***RUVBL1***	RuvB like AAA ATPase 1	3q21.3
ACVR2B	***ACVR2B***	activin A receptor type 2B	3p22.2
CFC1	***CFC1***	cripto, FRL-1, cryptic family 1	2q21.1
CRELD1	***CRELD1***	cysteine rich with EGF like domains 1	3p25.3
GALNT11	***GALNT11***	polypeptide N-acetylgalactosaminyltransferase 11	7q36.1
GDF1	***GDF1***	growth differentiation factor 1	19p13.11
NEK2	***NEK2***	NIMA related kinase 2	1q32.3
NODAL	***NODAL***	nodal growth differentiation factor	10q22.1
NUP188	***NUP188***	nucleoporin 188	9q34.11
ROCK2	***ROCK2***	Rho associated coiled-coil containing protein kinase 2	2p25.1
ZIC3	***ZIC3***	Zic family member 3	Xq26.3
CFAP53	***CFAP53***	cilia and flagella associated protein 53	18q21.1
CCDC105	***CCDC105***	coiled-coil domain containing 105	19p13.12
CCDC24	***CCDC24***	coiled-coil domain containing 24	1p34.1
CCDC78	***CCDC78***	coiled-coil domain containing 78	16p13.3
WDR38	***WDR38***	WD repeat domain 38	9q33.3
CFAP43	***CFAP43***	cilia and flagella associated protein 43	10q25.1
CFAP45	***CFAP45***	cilia and flagella associated protein 45	1q23.2

CFAP57	***CFAP57***	cilia and flagella associated protein 57	1p34.2
CFAP58	***CFAP58***	cilia and flagella associated protein 58	10q25.1
CFAP221	***CFAP221***	cilia and flagella associated protein 221	2q14.2
PPP1R42	***PPP1R42***	protein phosphatase 1 regulatory subunit 42	8q13.1
RSPH14	***RSPH14***	radial spoke head 14 homolog	22q11.22-q11.23
ANK2	***ANK2***	ankyrin 2	4q25-q26
CFAP126	***CFAP126***	cilia and flagella associated protein 126	1q23.3
DNAH1	***DNAH1***	dynein axonemal heavy chain 1	3p21.1
DNAH7	***DNAH7***	dynein axonemal heavy chain 7	2q32.3
DYNLL1	***DYNLL1***	dynein light chain LC8-type 1	12q24.31
DYNLL2	***DYNLL2***	dynein light chain LC8-type 2	17q22
DYNLT1	***DYNLT1***	dynein light chain Tctex-type 1	6q25.3
PIH1D1	***PIH1D1***	PIH1 domain containing 1	19q13.33
PIH1D2	***PIH1D2***	PIH1 domain containing 2	11q23.1
TTC29	***TTC29***	tetratricopeptide repeat domain 29	4q31.22
WDR63	***WDR63***	WD repeat domain 63	1p22.3
WDR78	***WDR78***	WD repeat domain 78	1p31.3
ZMYND12	***ZMYND12***	zinc finger MYND-type containing 12	1p34.2
CFAP36	***CFAP36***	cilia and flagella associated protein 36	2p16.1
DAW1	***DAW1***	dynein assembly factor with WD repeats 1	2q36.3
C22orf23	***C22orf23***	chromosome 22 open reading frame 23	22q13.1
CCDC42B	***CFAP73***	cilia and flagella associated protein 73	12q24.13
DAB1	***DAB1***	DAB1, reelin adaptor protein	1p32.2
ENKD1	***ENKD1***	enkurin domain containing 1	16q22.1
ENKUR	***ENKUR***	enkurin, TRPC channel interacting protein	10p12.1
KATNA1	***KATNA1***	katanin catalytic subunit A1	6q25.1
KATNAL2	***KATNAL2***	katanin catalytic subunit A1 like 2	18q21.1
TTC12	***TTC12***	tetratricopeptide repeat domain 12	11q23.2
WDR92	***WDR92***	WD repeat domain 92	2p14
CFAP52	***CFAP52***	cilia and flagella associated protein 52	17p13.1
CFAP97	***CFAP97***	cilia and flagella associated protein 97	4q35.1
MAPK15	***MAPK15***	mitogen-activated protein kinase 15	8q24.3
AHI1	***AHI1***	Abelson helper integration site 1	6q23.3
ANKS6	***ANKS6***	ankyrin repeat and sterile alpha motif domain containing 6	9q22.33
ARL13B	***ARL13B***	ADP ribosylation factor like GTPase 13B	3q11.1-q11.2
CFAP410	***CFAP410***	cilia and flagella associated protein 410	21q22.3
C2CD3	***C2CD3***	C2 calcium dependent domain containing 3	11q13.4

CPLANE1	***CPLANE1***	ciliogenesis and planar polarity effector 1	5p13.2
CC2D2A	***CC2D2A***	coiled-coil and C2 domain containing 2A	4p15.32
CEP104	***CEP104***	centrosomal protein 104	1p36.32
CEP120	***CEP120***	centrosomal protein 120	5q23.2
CEP290	***CEP290***	centrosomal protein 290	12q21.32
CEP83	***CEP83***	centrosomal protein 83	12q22
CSPP1	***CSPP1***	centrosome and spindle pole associated protein 1	8q13.1-q13.2
DDX59	***DDX59***	DEAD-box helicase 59	1q32.1
DYNC2H1	***DYNC2H1***	dynein cytoplasmic 2 heavy chain 1	11q22.3
DYNC2LI1	***DYNC2LI1***	dynein cytoplasmic 2 light intermediate chain 1	2p21
EVC	***EVC***	EvC ciliary complex subunit 1	4p16.2
EVC2	***EVC2***	EvC ciliary complex subunit 2	4p16.2
GLI3	***GLI3***	GLI family zinc finger 3	7p14.1
HNF1B	***HNF1B***	HNF1 homeobox B	17q12
IFT122	***IFT122***	intraflagellar transport 122	3q21.3-q22.1
IFT140	***IFT140***	intraflagellar transport 140	16p13.3
IFT172	***IFT172***	intraflagellar transport 172	2p23.3
IFT27	***IFT27***	intraflagellar transport 27	22q12.3
IFT43	***IFT43***	intraflagellar transport 43	14q24.3
IFT80	***IFT80***	intraflagellar transport 80	3q25.33
INPP5E	***INPP5E***	inositol polyphosphate-5-phosphatase E	9q34.3
KIAA0586	***KIAA0586***	KIAA0586	14q23.1
KIAA0753	***KIAA0753***	KIAA0753	17p13.1
KIF7	***KIF7***	kinesin family member 7	15q26.1
LBR	***LBR***	lamin B receptor	1q42.12
MUC1	***MUC1***	mucin 1, cell surface associated	1q22
NEK1	***NEK1***	NIMA related kinase 1	4q33
NEK8	***NEK8***	NIMA related kinase 8	17q11.1
NPHP1	***NPHP1***	nephrocystin 1	2q13
NPHP3	***NPHP3***	nephrocystin 3	3q22.1
NPHP4	***NPHP4***	nephrocystin 4	1p36.31
PDE6D	***PDE6D***	phosphodiesterase 6D	2q37.1
PIBF1	***PIBF1***	progesterone immunomodulatory binding factor 1	13q21.33-q22.1
POC1B	***POC1B***	POC1 centriolar protein B	12q21.33
RPGRIP1L	***RPGRIP1L***	RPGRIP1 like	16q12.2
SBDS	***SBDS***	SBDS, ribosome maturation factor	7q11.21
SDCCAG8	***SDCCAG8***	serologically defined colon cancer antigen 8	1q43-q44
TBC1D32	***TBC1D32***	TBC1 domain family member 32	6q22.31
TCTEX1D2	***TCTEX1D2***	Tctex1 domain containing 2	3q29

TCTN1	***TCTN1***	tectonic family member 1	12q24.11
TCTN3	***TCTN3***	tectonic family member 3	10q24.1
TMEM138	***TMEM138***	transmembrane protein 138	11q12.2
TMEM216	***TMEM216***	transmembrane protein 216	11q13.1
TMEM231	***TMEM231***	transmembrane protein 231	16q23.1
TMEM237	***TMEM237***	transmembrane protein 237	2q33.1
TMEM67	***TMEM67***	transmembrane protein 67	8q22.1
TTC21B	***TTC21B***	tetratricopeptide repeat domain 21B	2q24.3
UMOD	***UMOD***	uromodulin	16p12.3
WDR34	***WDR34***	WD repeat domain 34	9q34.11
WDR35	***WDR35***	WD repeat domain 35	2p24.1
WDR60	***WDR60***	WD repeat domain 60	7q36.3
ZNF423	***ZNF423***	zinc finger protein 423	16q12.1
CEP164	***CEP164***	centrosomal protein 164	11q23.3
CEP41	***CEP41***	centrosomal protein 41	7q32.2
GLIS2	***GLIS2***	GLIS family zinc finger 2	16p13.3
INVS	***INVS***	inversin	9q31.1
IQCB1	***IQCB1***	IQ motif containing B1	3q13.33
SPATA7	***SPATA7***	spermatogenesis associated 7	14q31.3
TCTN2	***TCTN2***	tectonic family member 2	12q24.31
DCDC2	***DCDC2***	doublecortin domain containing 2	6p22.3
TAPT1	***TAPT1***	transmembrane anterior posterior transformation 1	4p15.32
TMEM107	***TMEM107***	transmembrane protein 107	17p13.1
KIF27	***KIF27***	kinesin family member 27	9q21.32
PCNT	***PCNT***	pericentrin	21q22.3
TCHH	***TCHH***	trichohyalin	1q21.3
WDR1	***WDR1***	WD repeat domain 1	4p16.1
CCDC18	***CCDC18***	coiled-coil domain containing 18	1p22.1
SLC16A7	***SLC16A7***	solute carrier family 16 member 7	12q14.1
SMC1A	***SMC1A***	structural maintenance of chromosomes 1A	Xp11.22
COL7A1	***COL7A1***	collagen type VII alpha 1 chain	3p21.31
C12orf55	***CFAP54***	cilia and flagella associated protein 54	12q23.1
ANK3	***ANK3***	ankyrin 3	10q21.2
KIF24	***KIF24***	kinesin family member 24	9p13.3
KIAA0754	***KIAA0754***	KIAA0754	1p34.2
DBH	***DBH***	dopamine beta-hydroxylase	9q34.2
CCDC37	***CFAP100***	cilia and flagella associated protein 100	3q21.3
TTC28	***TTC28***	tetratricopeptide repeat domain 28	22q12.1
CCDC74A	***CCDC74A***	coiled-coil domain containing 74A	2q21.1
MASP1	***MASP1***	mannan binding lectin serine peptidase 1	3q27.3
CEP19	***CEP19***	centrosomal protein 19	3q29
CES1	***CES1***	carboxylesterase 1	16q12.2

B4GALNT3	***B4GALNT3***	beta-1,4-N-acetyl-galactosaminyltransferase 3	12p13.33
LTBP1	***LTBP1***	latent transforming growth factor beta binding protein 1	2p22.3
CCDC180	***CCDC180***	coiled-coil domain containing 180	9q22.33
EPB41L3	***EPB41L3***	erythrocyte membrane protein band 4.1 like 3	18p11.31
AK8	***AK8***	adenylate kinase 8	9q34.13
C14orf45	***BBOF1***	basal body orientation factor 1	14q24.3
C11orf65	***C11orf65***	chromosome 11 open reading frame 65	11q22.3
C4orf22	***CFAP299***	cilia and flagella associated protein 299	4q21.21
CAPSL	***CAPSL***	calcyphosine like	5p13.2
CASC1	***CASC1***	cancer susceptibility 1	12p12.1
CCDC170	***CCDC170***	coiled-coil domain containing 170	6q25.1
CETN2	***CETN2***	centrin 2	Xq28
C9orf117	***CFAP157***	cilia and flagella associated protein 157	9q34.11
C15orf26	***CFAP161***	cilia and flagella associated protein 161	15q25.1
C6orf165	***CFAP206***	cilia and flagella associated protein 206	6q15
CFAP61	***CFAP61***	cilia and flagella associated protein 61	20p11.23
CFAP70	***CFAP70***	cilia and flagella associated protein 70	10q22.2
CYB5D1	***CYB5D1***	cytochrome b5 domain containing 1	17p13.1
DNALI1	***DNALI1***	dynein axonemal light intermediate chain 1	1p34.3
DRC7	***DRC7***	dynein regulatory complex subunit 7	16q21
DYNLRB2	***DYNLRB2***	dynein light chain roadblock-type 2	16q23.2
EFHB	***EFHB***	EF-hand domain family member B	3p24.3
EFHC1	***EFHC1***	EF-hand domain containing 1	6p12.2
EFHC2	***EFHC2***	EF-hand domain containing 2	Xp11.3
IFT57	***IFT57***	intraflagellar transport 57	3q13.12-q13.13
IQCG	***IQCG***	IQ motif containing G	3q29
SPATA17	***SPATA17***	spermatogenesis associated 17	1q41
IQUB	***IQUB***	IQ motif and ubiquitin domain containing	7q31.32
KIFAP3	***KIFAP3***	kinesin associated protein 3	1q24.2
LRGUK	***LRGUK***	leucine rich repeats and guanylate kinase domain containing	7q33
LRRC23	***LRRC23***	leucine rich repeat containing 23	12p13.31
LRRC34	***LRRC34***	leucine rich repeat containing 34	3q26.2
MAATS1	***MAATS1***	MYCBP associated and testis expressed 1	3q13.33
MAK	***MAK***	male germ cell associated kinase	6p24.2
ALPK3	***ALPK3***	alpha kinase 3	15q25.3
MLF1	***MLF1***	myeloid leukemia factor 1	3q25.32

MNS1	***MNS1***	meiosis specific nuclear structural 1	15q21.3
MORN2	***MORN2***	MORN repeat containing 2	2p22.1
MORN3	***MORN3***	MORN repeat containing 3	12q24.31
NME5	***NME5***	NME/NM23 family member 5	5q31.2
PACRG	***PACRG***	parkin coregulated	6q26
PROM1	***PROM1***	prominin 1	4p15.32
RABL2A	***RABL2A***	RAB, member of RAS oncogene family like 2A	2q14.1
RFX3	***RFX3***	regulatory factor X3	9p24.2
SAXO2	***SAXO2***	stabilizer of axonemal microtubules 2	15q25.2
SPA17	***SPA17***	sperm autoantigenic protein 17	11q24.2
SPAG16	***SPAG16***	sperm associated antigen 16	2q34
SPAG6	***SPAG6***	sperm associated antigen 6	10p12.2
TCTEX1D1	***TCTEX1D1***	Tctex1 domain containing 1	1p31.3
TEKT1	***TEKT1***	tektin 1	17p13.1
TEKT2	***TEKT2***	tektin 2	1p34.3
TEKT4	***TEKT4***	tektin 4	2q11.1
TPPP3	***TPPP3***	tubulin polymerization promoting protein family member 3	16q22.1
TTC21A	***TTC21A***	tetratricopeptide repeat domain 21A	3p22.2
TTC26	***TTC26***	tetratricopeptide repeat domain 26	7q34
TTC30B	***TTC30B***	tetratricopeptide repeat domain 30B	2q31.2
TUBA1A	***TUBA1A***	tubulin alpha 1a	12q13.12
AGBL2	***AGBL2***	ATP/GTP binding protein like 2	11p11.2
ATXN1	***ATXN1***	ataxin 1	6p22.3
BTBD16	***BTBD16***	BTB domain containing 16	10q26.13
C20ORF96	***C20orf96***	chromosome 20 open reading frame 96	20p13
C21ORF58	***C21orf58***	chromosome 21 open reading frame 58	21q22.3
C3	***C3***	complement C3	19p13.3
CATIP	***CATIP***	ciliogenesis associated TTC17 interacting protein	2q35
DDAH1	***DDAH1***	dimethylarginine dimethylaminohydrolase 1	1p22.3
DLEC1	***DLEC1***	DLEC1, cilia and flagella associated protein	3p22.2
DZIP1L	***DZIP1L***	DAZ interacting zinc finger protein 1 like	3q22.3
EFCAB6	***EFCAB6***	EF-hand calcium binding domain 6	22q13.2-q13.31
FANK1	***FANK1***	fibronectin type III and ankyrin repeat domains 1	10q26.2
FOXJ1	***FOXJ1***	forkhead box J1	17q25.1
GLB1L2	***GLB1L2***	galactosidase beta 1 like 2	11q25
GLIS3	***GLIS3***	GLIS family zinc finger 3	9p24.2
GOLGB1	***GOLGB1***	golgin B1	3q13.33
IFT88	***IFT88***	intraflagellar transport 88	13q12.11

KIAA1407	***CCDC191***	coiled-coil domain containing 191	3q13.31
KIF17	***KIF17***	kinesin family member 17	1p36.12
KNDC1	***KNDC1***	kinase non-catalytic C-lobe domain containing 1	10q26.3
MDH1B	***MDH1B***	malate dehydrogenase 1B	2q33.3
MTTP	***MTTP***	microsomal triglyceride transfer protein	4q23
MUSTN1	***MUSTN1***	musculoskeletal, embryonic nuclear protein 1	3p21.1
NEK11	***NEK11***	NIMA related kinase 11	3q22.1
NEK3	***NEK3***	NIMA related kinase 3	13q14.3
OSBPL3	***OSBPL3***	oxysterol binding protein like 3	7p15.3
RAB36	***RAB36***	RAB36, member RAS oncogene family	22q11.23
RFX2	***RFX2***	regulatory factor X2	19p13.3
RSPH10B	***RSPH10B***	radial spoke head 10 homolog B	7p22.1
SPAG8	***SPAG8***	sperm associated antigen 8	9p13.3
SPATA18	***SPATA18***	spermatogenesis associated 18	4q12
SPEF1	***SPEF1***	sperm flagellar 1	20p13
TCTE1	***TCTE1***	t-complex-associated-testis-expressed 1	6p21.1
TMC5	***TMC5***	transmembrane channel like 5	16p12.3
TOB2	***TOB2***	transducer of ERBB2, 2	22q13.2
TSNAXIP1	***TSNAXIP1***	translin associated factor X interacting protein 1	16q22.1
TTC25	***TTC25***	tetratricopeptide repeat domain 25	17q21.2
TTLL3	***TTLL3***	tubulin tyrosine ligase like 3	3p25.3
TTLL9	***TTLL9***	tubulin tyrosine ligase like 9	20q11
CCP110	***CCP110***	centriolar coiled-coil protein 110	16p12.3
IFT81	***IFT81***	intraflagellar transport 81	12q24.11
SPAG9	***SPAG9***	sperm associated antigen 9	17q21.33
CCDC108	***CFAP65***	cilia and flagella associated protein 65	2q35
CCDC113	***CCDC113***	coiled-coil domain containing 113	16q21
CCDC96	***CCDC96***	coiled-coil domain containing 96	4p16.1
MKS1	***MKS1***	Meckel syndrome, type 1	17q22
IFT22	***IFT22***	intraflagellar transport 22	7q22.1
CEP126	***CEP126***	centrosomal protein 126	11q22.1
AK7	***AK7***	adenylate kinase 7	14q32.2
DPCD	***DPCD***	deleted in primary ciliary dyskinesia homolog (mouse)	10q24.32
ULK4	***ULK4***	unc-51 like kinase 4	3p22.1
ODF2	***ODF2***	outer dense fiber of sperm tails 2	9q34.11
TTLL1	***TTLL1***	tubulin tyrosine ligase like 1	22q13.2
RABL2B	***RABL2B***	RAB, member of RAS oncogene family like 2B	22q13.33
MAPK10	***MAPK10***	mitogen-activated protein kinase 10	4q21.3
LRTOMT	***LRTOMT***	leucine rich transmembrane and O-methyltransferase domain containing	11q13.4

TUBB4B	***TUBB4B***	tubulin beta 4B class IVb	9q34.3
BBS5	***BBS5***	Bardet-Biedl syndrome 5	2q31.1
CAPS	***CAPS***	calcyphosine	19p13.3
SMYD2	***SMYD2***	SET and MYND domain containing 2	1q32.3
PLEKHB1	***PLEKHB1***	pleckstrin homology domain containing B1	11q13.4
PIFO	***PIFO***	primary cilia formation	1p13.2
STOML3	***STOML3***	stomatin like 3	13q13.2
INTU	***INTU***	inturned planar cell polarity protein	4q28.1
MPDZ	***MPDZ***	multiple PDZ domain crumbs cell polarity complex component	9p23
NCALD	***NCALD***	neurocalcin delta	8q22.3
MAPRE3	***MAPRE3***	microtubule associated protein RP/EB family member 3	2p23.3

The gene list are colour coded when blue colour indicates known PCD genes, green colour indicates newly discovered genes implicated in PCD with functional confirmation in human and model organisms, and yellow colour indicates genes with putative mutations in PCD patients without functional characterization so far.

Table A-0-3 List of genes included in this study and the RefSeq accessions of both transcript and protein used in nomenclature of mutations

Gene Name	NCBI RefSeq Transcript	NCBI RefSeq Protein
ARMC4	NM_001290020.1	NP_001276949.1
CCDC103	NM_001258395.1	NP_001245324.1
CCDC114	NM_144577.3	NP_653178.3
CCDC151	NM_145045.4	NP_659482.3
CCDC39	NM_181426.1	NP_852091.1
CCDC40	NM_017950.3	NP_060420.2
CCDC65	NM_033124.4	NP_149115.2
CCNO	NM_021147.3	NP_066970.3
DNAAF1	NM_178452.5	NP_848547.4
DNAAF3	NM_001256715.1	NP_001243644.1
DNAH11	NM_001277115.1	NP_001264044.1
DNAH5	NM_001369.2	NP_001360.1
DNAI1	NM_012144.3	NP_036276.1
DNAI2	NM_023036.4	NP_075462.3
DRC1	NM_145038.4	NP_659475.2
DYX1C1	NM_130810.3	NP_570722.2
HEATR2	NM_017802.3	NP_060272.3
SPAG1	NM_003114.4	NP_003105.2
HYDIN	NM_001270974.1	NP_001257903.1
LRRC6	NM_012472.4	NP_036604.2
MCIDAS	NM_001190787.1	NP_001177716.1
OFD1	NM_003611.2	NP_003602.1
PIH1D3	NM_001169154.1	NP_001162625.1
RPGR	NM_001034853.1	NP_001030025.1
RSPH1	NM_080860.3	NP_543136.1
RSPH4A	NM_001010892.2	NP_001010892.1
RSPH9	NM_152732.4	NP_689945.2
ZMYND10	NM_015896.2	NP_056980.2
C11orf70	NM_032930.2	NP_116319.2
DNAH14	NM_001373.1	NP_001364.1
DNAH2	NM_020877.2	NP_065928.2
DNAH3	NM_017539.2	NP_060009.1
DNAH9	NM_001372.3	NP_001363.2
IFT74	NM_025103.3	NP_079379.2
NME7	NM_013330.4	NP_037462.1
TTC30B	NM_152517.2	NP_689730.2
CFAP46	NM_001200049.2	NP_001186978.2

Table A-0-4 Minor allele frequency and segregation results of bi-allelic and hemizygous mutations in known PCD genes

Study ID	Gene	c.Nomenclature	Exac_MAF	almena_MAF	BinB_MAF	Pathogenicity class	Segregation
G001	***DNAAF3***	c.296del	Not in Exac	Not in al mena	Not in BinB	4	Mo carrier
		c.609_610delinsTGGGA	Not in Exac	Not in al mena	Not in BinB	4	Fa carrier
G002	***DNAH5***	c.13486C>T	0.00005803	Not in al mena	Not in BinB	5	Mo carrier, No Fa DNA
		c.13458_13459insT	0.00005783	Not in al mena	Not in BinB	5	Mo Normal, No Fa DNA
G003	***RSPH9***	c.801_803delGAA	0.00005765	3 / 0.00151	Not in BinB	5	No parents' DNA
G004	***RSPH4A***	c.460C>T	0.00001647	Not in al mena	3/0.00195	5	No parents' DNA
G005	***CCDC40***	c.3181-3C>G	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G006	***LRRC6***	c.630delG	0.000206	Not in al mena	10/0.0065	5	Both parents carrier
G007	***CCDC40***	c.1414del	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
		c.3097A>T	Not in Exac	Not in al mena	Not in BinB	5	No parents' DNA
G008	***CCDC40***	c.2712-1G>T	0.00003315	Not in al mena	Not in BinB	5	Mo Normal, No Fa DNA
		c.248delC	0.0004794	Not in al mena	Not in BinB	5	Mo carrier, No Fa DNA
G009	***DNAAF3***	c.162_164delinsG	Not in Exac	Not in al mena	Not in BinB	4	Both parents carrier
G010	***DNAI1***	c.48+2dupT	0.0004624	Not in al mena	Not in BinB	5	Both parents carrier
G011	***ZMYND10***	c.212T>C	Not in Exac	Not in al mena	Not in BinB	4	3 affected siblings, No parents' DNA
G012	***CCDC114***	c.1093C>T	Not in Exac	Not in al mena	Not in BinB	4	Fa carrier, No Mo DNA
G013	***DNAH5***	c.13175T>G	0.000486	Not in al mena	Not in BinB	3	No parents' DNA
		c.3733C>T	0.0002149	Not in al mena	Not in BinB	3	No parents' DNA
G014	***DNAI1***	c.1612G>A	0.0000742	Not in al mena	Not in BinB	5	2 affected siblings, No parents' DNA
G015	***RSPH1***	c.275-2A>C	0.0003625	Not in al mena	Not in BinB	5	No parents' DNA

G016	***DNAI1***	c.48+2dupT	0.0004624	Not in al mena	Not in BinB	5	No parents' DNA
G017	***HYDIN***	c.11843C>G	Not in Exac	Not in al mena	Not in BinB	3	Not sanger confirmed
		c.10438G>A	Not in Exac	Not in al mena	Not in BinB	3	Not sanger confirmed
G018	***CCDC40***	c.1819_1823delinsT	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G019	***CCDC103***	c.461A>C	0.001261	2 / 0.00101	6/0.0039	5	No parents' DNA
G020	***DNAH5***	c.10615C>T	Not in Exac	Not in al mena	Not in BinB	5	No parents' DNA
		c.10815delT	0.0001483	Not in al mena	Not in BinB	5	No parents' DNA
G021	***HYDIN***	c.6182A>G	Not in Exac	Not in al mena	Not in BinB	3	Not sanger confirmed
		c.2057C>T	Not in Exac	Not in al mena	Not in BinB	3	Not sanger confirmed
G022	***CCNO***	c.538dupC	0.0000249	Not in al mena	Not in BinB	4	No parents' DNA
G023	***DNAH5***	c.6930_6934delinsG	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
		c.13458_13459insT	0.00005783	Not in al mena	Not in BinB	5	No parents' DNA
G024	***RPGR***	c.646G>T	Not in Exac	Not in al mena	N/A	4	2 affected siblings, No parents' DNA
G025	***DNAH11***	c.5593C>T	0.00001657	Not in al mena	1/0.0006	4	No parents' DNA
G026	***DNAH5***	c.6261T>G	Not in Exac	Not in al mena	Not in BinB	4	2 affected siblings, No parents' DNA
G027	***DYX1C1***	Deletion of exon 7	N/A	N/A	N/A	5	Mo carrier, No Fa DNA
G028	***DNAH5***	c.5710-2A>G	0.00002482	Not in al mena	Not in BinB	5	No parents' DNA
G029	***DNAH5***	c.11455+5G>A	Not in Exac	Not in al mena	Not in BinB	4	Mo carrier, No Fa DNA
		c.6261T>G	Not in Exac	Not in al mena	Not in BinB	4	Mo Normal, No Fa DNA
G030	***RSPH9***	c.801_803delGAA	0.00005765	3 / 0.00151	Not in BinB	5	No parents' DNA
G031	***CCDC65***	c.658G>T	0.000008243	Not in al mena	Not in BinB	4	2 affected siblings, Mo carrier, No Fa DNA
G032	***DNAH11***	c.4552C>T	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
		c.5778+1G>A	0.00001193	Not in al mena	Not in BinB	5	No parents' DNA

G033	***LRRC6***	c.630delG	0.000206	Not in al mena	10/0.0065	5	Fa carrier, No Mo DNA
G034	***PIH1D3***	c.266G>A	Not in Exac	Not in al mena	N/A	5	No parents' DNA
G035	***OFD1***	c.2745_2747delinsC	Not in Exac	Not in al mena	N/A	3	No parents' DNA
G036	***DNAH11***	c.8932C>T	0.00001624	Not in al mena	Not in BinB	4	No parents' DNA
		c.853_857delinsG	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G037	***CCDC103***	c.461A>C	0.001261	2 / 0.00101	6/0.0039	5	No parents' DNA
G038	***RSPH4A***	c.1962_1966delinsC	Not in Exac	Not in al mena	Not in BinB	4	Both parents carrier
G039	***DNAAF3***	c.621dupT	0.00008331	Not in al mena	Not in BinB	5	Fa carrier, No Mo DNA
G040	***DNAH5***	c.10815delT	0.0001483	Not in al mena	Not in BinB	5	No parents' DNA
		c.13458_13459insT	0.00005783	Not in al mena	Not in BinB	5	No parents' DNA
G041	***DNAH11***	c.13040T>C	Not in Exac	Not in al mena	Not in BinB	3	No parents' DNA
		Deletion of exons 68-75	N/A	N/A	N/A	3	No parents' DNA
G042	***DYX1C1***	Deletion of exon 7	N/A	N/A	N/A	5	No parents' DNA
G043	***DNAH5***	c.10601T>C	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
		c.13458_13459insT	0.00005783	Not in al mena	Not in BinB	5	No parents' DNA
G044	***CCDC40***	c.1415delC	Not in Exac	Not in al mena	Not in BinB	5	No parents' DNA
G045	***DNAH5***	c.2710G>T	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G046	***LRRC6***	c.630delG	0.000206	Not in al mena	10/0.0065	5	No parents' DNA
G047	***DNAAF3***	c.228+5G>C	0.00002956	Not in al mena	Not in BinB	4	No parents' DNA
G048	***DNAAF3***	c.1030_1031delinsG	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
		c.1273G>T	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G049	***CCDC40***	c.248delC	0.0004794	Not in al mena	Not in BinB	5	No parents' DNA
G050	***DNAH5***	c.232C>T	Not in Exac	Not in al mena	Not in BinB	5	No parents' DNA
		c.10815delT	0.0001483	Not in al mena	Not in BinB	5	No parents' DNA
G051	***DNAH5***	c.5890_5894dup	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
		c.6791G>A	0.000008244	Not in al mena	Not in BinB	5	No parents' DNA

G052	***DNAH5***	c.13285C>T	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
		c.8642C>G	0.0000165	Not in al mena	Not in BinB	5	No parents' DNA
G053	***SPAG1***	c.1519dupA	0.00004118	Not in al mena	Not in BinB	5	No parents' DNA
G054	***DNAH5***	c.6763C>T	0.00002472	Not in al mena	Not in BinB	4	No parents' DNA
		c.9480T>A	0.000008242	Not in al mena	Not in BinB	4	No parents' DNA
G055	***RSPH4A***	c.1351C>T	0.00004121	Not in al mena	Not in BinB	4	No parents' DNA
		c.116C>A	0.00001657	Not in al mena	Not in BinB	5	No parents' DNA
G056	***CCDC39***	c.830_831delCA	0.00006395	Not in al mena	Not in BinB	5	No parents' DNA
G057	***CCDC103***	c.461A>C	0.001261	2 / 0.00101	6/0.0039	5	No parents' DNA
G058	***MCIDAS***	c.332_333delinsG	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G059	***LRRC6***	c.630delG	0.000206	Not in al mena	10/0.0065	5	Mo carrier, No Fa DNA
G060	***ARMC4***	c.1233_1234delinsT	Not in Ecxac	Not in al mena	Not in BinB	4	Fa carrier
		c.1969C>T	Not in Exac	Not in al mena	Not in BinB	5	Mo carrier
G061	***DNAH5***	c.5177T>C	0.000008238	Not in al mena	Not in BinB	5	Mo carrier, No Fa DNA
		c.1730G>C	0.00003591	Not in al mena	Not in BinB	5	Mo Normal, No Fa DNA
G062	***RSPH1***	c.275-2A>C	0.0003625	Not in al mena	Not in BinB	5	No parents' DNA
G063	***DNAH5***	c.8404C>T	Not in Exac	Not in al mena	Not in BinB	5	No parents' DNA
		c.6249G>A	0.000008293	Not in al mena	Not in BinB	3	No parents' DNA
G064	***RSPH4A***	c.1456G>C	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G065	***DNAH5***	c.10815delT	0.0001483	Not in al mena	Not in BinB	5	No parents' DNA
G066	***DNAH5***	c.5557A>T	0.000008243	Not in al mena	Not in BinB	4	No parents' DNA
G067	***ARMC4***	c.1283C>G	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G068	***CCDC103***	c.461A>C	0.001261	2 / 0.00101	6/0.0039	5	Mo carrier, No Fa DNA
G069	***CCDC40***	c.712G>T	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
		c.940-2A>G	0.00001661	Not in al mena	Not in BinB	5	No parents' DNA

G070	***CCDC40***	c.248delC	0.0004794	Not in al mena	Not in BinB	5	No parents' DNA
		c.2450-2A>G	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G071	***LRRC6***	c.630delG	0.000206	Not in al mena	10/0.0065	5	No parents' DNA
G072	***DNAH11***	c.5846G>A	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
		c.13380_13383dup	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G073	***DNAH11***	c.9581_9582del	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
		c.4333C>T	0.00003391	Not in al mena	Not in BinB	4	No parents' DNA
G074	***DNAH11***	c.7472G>C	0.000149	Not in al mena	Not in BinB	4	No parents' DNA
		c.6565C>T	0.00002484	Not in al mena	Not in BinB	4	No parents' DNA
G075	***DNAAF1***	Deletion of exon 1,2&3	N/A	N/A	N/A	4	No parents' DNA
G076	***RSPH1***	c.275-2A>C	0.0003625	Not in al mena	Not in BinB	5	No parents' DNA
		c.281G>A	Not in Exac	Not in al mena	Not in BinB	5	No parents' DNA
G077	***DNAI2***	c.1494+2dupT	0.00000864	Not in al mena	Not in BinB	4	No parents' DNA
G078	***CCDC103***	c.383dup	Not in Exac	Not in al mena	1/0.0006	5	No parents' DNA
G079	***DNAH5***	c.10815delT	0.0001483	Not in al mena	Not in BinB	5	No parents' DNA
		c.10384C>T	0.00004118	Not in al mena	Not in BinB	5	No parents' DNA
G080	***CCDC164***	Deletion of exons 11,12&13	N/A	N/A	N/A	4	No parents' DNA
G081	***CCDC39***	c.1871_1872del	Not in Exac	Not in al mena	Not in BinB	4	2 cousins, Both parents carrier
G082	***CCDC103***	c.104G>C	0.000008241	Not in al mena	Not in BinB	5	Both parents carrier
G083	***LRRC6***	c.436G>C	0.0001977	Not in al mena	Not in BinB	5	Both parents carrier
G084	***MCIDAS***	c.718-1G>A	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G085	***CCDC151***	c.850C>T	Not in Exac	Not in al mena	Not in BinB	4	Both parents carrier
G086	***CCDC40***	c.48A>G	Not in Exac	Not in al mena	Not in BinB	3	Both parents carrier
G087	***CCDC39***	c.1665+1G>T	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA

G088	***CCDC40***	c.387C>G	Not in Exac	Not in al mena	Not in BinB	4	Both parents carrier
G089	***DNAH5***	c.8320T>C	Not in Exac	Not in al mena	Not in BinB	3	Both parents carrier
G090	***LRRC6***	c.974+1G>A	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G091	***CCDC40***	Deletion of exon 11&12	N/A	N/A	N/A	4	No parents' DNA
G092	***DNAH5***	c.11258del	Not in Exac	Not in al mena	Not in BinB	4	Mo carrier
		c.2964_2965del	Not in Exac	Not in al mena	Not in BinB	4	Fa carrier
G093	***DNAH5***	c.10815T>G	Not in Exac	Not in al mena	Not in BinB	3	Mo carrier, No Fa DNA
		c.5157C>T	0.000008241	Not in al mena	Not in BinB	3	Mo Normal, No Fa DNA
G094	***ZMYND10***	c.510+1del	Not in Exac	Not in al mena	Not in BinB	4	2 affected siblings, Both parents carrier
G095	***DNAH5***	c.6782T>G	Not in Exac	Not in al mena	Not in BinB	4	2 affected siblings,Mo carrier, No Fa DNA
		c.1351C>T	Not in Exac	Not in al mena	Not in BinB	4	2 affected siblings,Mo Normal, No Fa DNA
G096	***CCDC40***	c.2824_2825insCTGT	Not in Exac	Not in al mena	Not in BinB	4	Fa carrier
		c.2920C>T	Not in Exac	Not in al mena	Not in BinB	4	Mo carrier
G097	***RSPH1***	c.275-2A>C	0.0003625	Not in al mena	Not in BinB	5	Fa carrier, No Mo DNA
G098	***CCNO***	c.263_267dup	Not in Exac	Not in al mena	Not in BinB	5	Both parents carrier
G099	***ZMYND10***	c.83G>A	Not in Exac	Not in al mena	Not in BinB	3	No parents' DNA
G100	***DNAH11***	c.12646G>T	Not in Exac	Not in al mena	Not in BinB	4	Both parents carrier
G101	***DNAH11***	c.3426-1G>A	0.00003394	Not in al mena	Not in BinB	4	No parents' DNA
		c.2514_2528delinsC	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G102	***CCDC39***	c.1871_1872del	Not in Exac	Not in al mena	Not in BinB	4	Both parents carrier
G103	***CCDC39***	c.1871_1872del	Not in Exac	Not in al mena	Not in BinB	4	Both parents carrier
G104	***DNAH11***	c.563T>C	0.0000127	Not in al mena	Not in BinB	3	2 cousins, Both parents carrier
G105	***DNAI2***	c.883C>T	0.00001647	Not in al mena	Not in BinB	4	Fa carrier, No Mo DNA

G106	***CCDC39***	c.1871_1872del	Not in Exac	Not in al mena	Not in BinB	4	Both parents carrier
G107	***DNAH11***	c.4438C>T	0.00000828	Not in al mena	Not in BinB	4	No parents' DNA
		c.13494_13500del	Not in Exac	Not in al mena	Not in BinB	4	No parents' DNA
G108	***DNAH5***	c.10229C>T	0.00004118	Not in al mena	Not in BinB	4	Fa normal, No Mo DNA
		c.5281C>T	0.00004119	Not in al mena	Not in BinB	4	Fa carrier, No Mo DNA
G109	***RSPH9***	c.760delG	Not in Exac	Not in al mena	Not in BinB	4	Fa affected, Mo carrier
G110	***CCDC39***	c.2182C>T	Not in Exac	Not in al mena	Not in BinB	4	Both parents carrier
G111	***CCDC40***	c.2824_2825insCTGT	Not in Exac	Not in al mena	Not in BinB	4	Both parents carrier
G112	***DNAH11***	c.13494_13500del	Not in Exac	Not in al mena	Not in BinB	4	2 cousins, Both parents carrier
G113	***CCDC39***	c.210+2T>C	0.000008327	Not in al mena	Not in BinB	4	Both parents carrier
G114	***ZMYND10***	c.490C>T	0.00001667	Not in al mena	Not in BinB	4	Both parents carrier
G115	***DNAH11***	c.12196-2A>G	Not in Exac	Not in al mena	Not in BinB	4	2 affected siblings, No parents' DNA
		c.6244C>T	0.00004139	Not in al mena	Not in BinB	5	2 affected siblings, No parents' DNA
G116	***CCDC103***	c.383dup	Not in Exac	Not in al mena	1/0.0006	5	No parents' DNA

Mo; Mother, Fa; Father.

Table A-0-5 TEM results, ethnicity and consanguinity in families with mutations in known PCD genes

Study ID	TEM Defect	Ethnicity Category	Consanguinity	Gene	Genotype
G001	IDA+ODA	South-Asian	No	***DNAAF3***	Comp Heterozygous
G002	ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G003	CC	Arab	Yes	***RSPH9***	Homozygous
G004	CC	Asian-Other	Yes	***RSPH4A***	Homozygous
G005	MT disorg.+IDA	South-Asian	No	***CCDC40***	Homozygous
G006	IDA+ODA	South-Asian	Yes	***LRRC6***	Homozygous
G007	MT disorg.+IDA	White-Caucasian	No	***CCDC40***	Comp Heterozygous
G008	MT disorg.+IDA	White-Caucasian	No	***CCDC40***	Comp Heterozygous
G009	IDA+ODA	South-Asian	No	***DNAAF3***	Homozygous
G010	ODA	White-Caucasian	No	***DNAI1***	Homozygous
G011	IDA+ODA	Unknown	Yes	***ZMYND10***	Homozygous
G012	Not done	South-Asian	Yes	***CCDC114***	Homozygous
G013	Normal	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G014	IDA+ODA	White-Caucasian	Yes	***DNAI1***	Homozygous
G015	Normal	White-Caucasian	No	***RSPH1***	Homozygous
G016	ODA	White-Caucasian	No	***DNAI1***	Homozygous
G017	Normal	South-Asian	No	***HYDIN***	Comp Heterozygous
G018	MT disorg.+IDA	Unknown	No	***CCDC40***	Homozygous
G019	Normal	Unknown	Yes	***CCDC103***	Homozygous
G020	ODA	White-Caucasian	Unknown	***DNAH5***	Comp Heterozygous
G021	Normal	White-Caucasian	No	***HYDIN***	Comp Heterozygous
G022	Lack of Cilia	White-Caucasian	No	***CCNO***	Homozygous
G023	IDA+ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G024	Lack of cilia	White-Caucasian	No	***RPGR***	Hemizygous
G025	Normal	Unknown	Unknown	***DNAH11***	Homozygous
G026	ODA	White-Caucasian	No	***DNAH5***	Homozygous

G027	Lack of Cilia	White-Caucasian	No	***DYX1C1***	Homozygous
G028	Inconclusive	White-Caucasian	No	***DNAH5***	Homozygous
G029	ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G030	MT disorg.	Arab	Unknown	***RSPH9***	Homozygous
G031	MT disorg.	Asian-Other	Yes	***CCDC65***	Homozygous
G032	Normal	White-Caucasian	No	***DNAH11***	Comp Heterozygous
G033	Not done	South-Asian	South-Asian	***LRRC6***	Homozygous
G034	ODA	South-Asian	No	***PIH1D3***	Hemizygous
G035	Normal	Unknown	No	***OFD1***	Hemizygous
G036	Normal	White-Caucasian	No	***DNAH11***	Comp Heterozygous
G037	Normal	South-Asian	No	***CCDC103***	Homozygous
G038	CC	White-Caucasian	No	***RSPH4A***	Homozygous
G039	IDA+ODA	South-Asian	No	***DNAAF3***	Homozygous
G040	ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G041	Inconclusive	South-Asian	Unknown	***DNAH11***	Comp Heterozygous
G042	IDA+ODA	White-Caucasian	No	***DYX1C1***	Homozygous
G043	ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G044	Not done	Asian-Other	Unknown	***CCDC40***	Homozygous
G045	Not done	Other	yes	***DNAH5***	Homozygous
G046	IDA+ODA	South-Asian	No	***LRRC6***	Homozygous
G047	IDA+ODA	White-Caucasian	No	***DNAAF3***	Homozygous
G048	IDA+ODA	White-Caucasian	No	***DNAAF3***	Comp Heterozygous
G049	ODA	White-Caucasian	No	***CCDC40***	Homozygous
G050	ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G051	ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G052	ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G053	ODA	Asian-Other	Yes	***SPAG1***	Homozygous
G054	ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous

G055	CC	White-Caucasian	No	***RSPH4A***	Comp Heterozygous
G056	MT disorg.+IDA	White-Caucasian	Unknown	***CCDC39***	Homozygous
G057	IDA	White-Caucasian	No	***CCDC103***	Homozygous
G058	Lack of Cilia	Other	Yes	***MCIDAS***	Homozygous
G059	IDA+ODA	South-Asian	Yes	***LRRC6***	Homozygous
G060	ODA	White-Caucasian	No	***ARMC4***	Comp Heterozygous
G061	ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G062	CC	White-Caucasian	No	***RSPH1***	Homozygous
G063	Not done	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G064	CC	Arab	Yes	***RSPH4A***	Homozygous
G065	ODA	White-Caucasian	No	***DNAH5***	Homozygous
G066	ODA	South-Asian	Yes	***DNAH5***	Homozygous
G067	ODA	Other	Unknown	***ARMC4***	Homozygous
G068	IDA	South-Asian	Unknown	***CCDC103***	Homozygous
G069	IDA	White-Caucasian	No	***CCDC40***	Comp Heterozygous
G070	Not done	Other	No	***CCDC40***	Comp Heterozygous
G071	IDA+ODA	South-Asian	Unknown	***LRRC6***	Homozygous
G072	Normal	White-Caucasian	No	***DNAH11***	Comp Heterozygous
G073	Normal	White-Caucasian	No	***DNAH11***	Comp Heterozygous
G074	Normal	White-Caucasian	No	***DNAH11***	Comp Heterozygous
G075	IDA+ODA	Asian-Other	No	***DNAAF1***	Homozygous
G076	CC	White-Caucasian	Unknown	***RSPH1***	Comp Heterozygous
G077	Normal	Unknown	Unknown	***DNAI2***	Homozygous
G078	IDA+ODA	South-Asian	Yes	***CCDC103***	Homozygous
G079	ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G080	IDA	South-Asian	Yes	***CCDC164***	Homozygous
G081	MT disorg.+IDA	Arab	Yes	***CCDC39***	Homozygous
G082	IDA+ODA	Arab	Yes	***CCDC103***	Homozygous

G083	IDA+ODA	Arab	Yes	***LRRC6***	Homozygous
G084	Lack of Cilia	South-Asian	Unknown	***MCIDAS***	Homozygous
G085	ODA	Arab	Yes	***CCDC151***	Homozygous
G086	MT disorg.+IDA	Arab	Yes	***CCDC40***	Homozygous
G087	Not done	South-Asian	Unknown	***CCDC39***	Homozygous
G088	Not done	Arab	Yes	***CCDC40***	Homozygous
G089	Not done	Arab	Yes	***DNAH5***	Homozygous
G090	Not done	Arab	Yes	***LRRC6***	Homozygous
G091	Not done	Arab	Yes	***CCDC40***	Homozygous
G092	Not done	Arab	Yes	***DNAH5***	Comp Heterozygous
G093	ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G094	IDA+ODA	White-Caucasian	No	***ZMYND10***	Homozygous
G095	Normal	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G096	MT disorg.	White-Caucasian	No	***CCDC40***	Comp Heterozygous
G097	MT disorg.	White-Caucasian	No	***RSPH1***	Homozygous
G098	Lack of Cilia	White-Caucasian	No	***CCNO***	Homozygous
G099	Inconclusive	Other	yes	***ZMYND10***	Homozygous
G100	Normal	Arab	Yes	***DNAH11***	Homozygous
G101	Normal	Asian-Other	Unknown	***DNAH11***	Comp Heterozygous
G102	MT disorg.	Arab	yes	***CCDC39***	Homozygous
G103	MT disorg.	Arab	yes	***CCDC39***	Homozygous
G104	Normal	Arab	Yes	***DNAH11***	Homozygous
G105	ODA	White-Caucasian	No	***DNAI2***	Homozygous
G106	MT disorg.	Arab	yes	***CCDC39***	Homozygous
G107	Normal	White-Caucasian	No	***DNAH11***	Comp Heterozygous
G108	ODA	White-Caucasian	No	***DNAH5***	Comp Heterozygous
G109	Not done	Arab	Yes	***RSPH9***	Homozygous
G110	Not done	Arab	No	***CCDC39***	Homozygous

G111	Not done	Arab	No	***CCDC40***	Homozygous
G112	Not done	Arab	Yes	***DNAH11***	Homozygous
G113	Not done	Arab	Yes	***CCDC39***	Homozygous
G114	Not done	Arab	Yes	***ZMYND10***	Homozygous
G115	Normal	South-Asian	Unknown	***DNAH11***	Comp Heterozygous
G116	IDA+ODA	South-Asian	Yes	***CCDC103***	Homozygous
G117	IDA+ODA	White-Caucasian	Unknown	***DNAH5***	Single Heterozygous
G118	Normal	White-Caucasian	No	***DNAH11***	Single Heterozygous
G119	Normal	White-Caucasian	No	***DNAH11***	Single Heterozygous
G120	Normal	South-Asian	No	***HYDIN***	Single Heterozygous
G121	IDA+ODA	South-Asian	No	***HEATR2***	Single Heterozygous
G122	ODA	White-Caucasian	Unknown	***DNAI1***	Single Heterozygous
G123	Inconclusive	White-Caucasian	No	***DNAH11***	Single Heterozygous
G124	Normal	White-Caucasian	No	***HYDIN***	Single Heterozygous
G125	Normal	Unknown	Unknown	***HYDIN***	Single Heterozygous
G126	IDA+ODA	Unknown	Unknown	***DNAH5***	Single Heterozygous
G127	Normal	South-Asian	No	***HYDIN***	Single Heterozygous
G128	ODA	White-Caucasian	No	***DNAH5***	Single Heterozygous
G117	IDA+ODA	White-Caucasian	Unknown	***DNAH5***	Single Heterozygous
G118	Normal	White-Caucasian	No	***DNAH11***	Single Heterozygous

Table A-0-6 In silico predication of the missense mutations identified in known PCD genes

c.Nomen	p.Nomen	Gene	Polyphen	SIFT	Mutation Taster	CADD
c.10615C>T	p.Arg3539Cys	***DNAH5***	1.000	D	Damaging	35
c.13040T>C	p.Leu4347Pro	***DNAH11***	0.999	D	Damaging	34
c.7472G>C	p.Arg2491Pro	***DNAH11***	0.992	D	Damaging	34
c.5665G>T	p.Gly1889Trp	***DNAH11***	0.960	D	Damaging	34
c.104G>C	p.Arg35Pro	***CCDC103***	0.999	D	Damaging	33
c.5846G>A	p.Arg1949Gln	***DNAH11***	0.957	D	Damaging	33
c.1612G>A	p.Ala538Thr	***DNAI1***	0.963	D	Damaging	33
c.1048C>T	p.Arg350Trp	***HEATR2***	0.982	D	Damaging	32
c.6923A>T	p.Asp2308Val	***HYDIN***	0.967	D	Damaging	32
c.436G>C	p.Asp146His	***LRRC6***	0.999	D	Damaging	31
c.83G>A	p.Gly28Asp	***ZMYND10***	0.957	NS	Damaging	31
c.8320T>C	p.Trp2774Arg	***DNAH5***	1.000	D	Damaging	29.8
c.13175T>G	p.Phe4392Cys	***DNAH5***	0.974	D	Damaging	29.7
c.3733C>T	p.Arg1245Cys	***DNAH5***	0.994	D	Damaging	29.7
c.6791G>A	p.Ser2264Asn	***DNAH5***	0.948	D	Damaging	29.3
c.6782T>G	p.Leu2261Arg	***DNAH5***	0.998	D	Damaging	29.1
c.10601T>C	p.Phe3534Ser	***DNAH5***	0.998	D	Damaging	28.7
c.1456G>C	p.Ala486Pro	***RSPH4A***	0.987	D	Damaging	28.1
c.10229C>T	p.Thr3410Met	***DNAH5***	0.996	D	Damaging	28
c.212T>C	p.Leu71Pro	***ZMYND10***	0.994	D	Damaging	26.6
c.2057C>T	p.Ala686Val	***HYDIN***	0.998	T	Damaging	26
c.11843C>G	p.Pro3948Arg	***HYDIN***	0.966	D	Damaging	25.9
c.9805T>G	p.Tyr3269Asp	***HYDIN***	0.899	D	Polymorphism	25.9
c.950G>A	p.Arg317Gln	***HYDIN***	0.652	T	Damaging	25.3
c.10815T>G	p.Asp3605Glu	***DNAH5***	1.000	D	Damaging	24.1
c.6249G>A	p.Met2083Ile	***DNAH5***	0.998	D	Damaging	23.9
c.563T>C	p.Met188Thr	***DNAH11***	0.039	D	Damaging	23.5
c.5177T>C	p.Leu1726Pro	***DNAH5***	0.994	D	Damaging	23
c.1730G>C	p.Arg577Thr	***DNAH5***	0.063	T	Damaging	22.7
c.8642C>G	p.Ala2881Gly	***DNAH5***	0.003	T	Damaging	20.2
c.10438G>A	p.Val3480Met	***HYDIN***	0.077	D	Polymorphism	15.54
c.461A>C	p.His154Pro	***CCDC103***	0.596	T	Damaging	15.16
c.6182A>G	p.Asn2061Ser	***HYDIN***	0.002	T	Polymorphism	0.001

c.Nomen; c.Nomenclature, p.Nomen; p.Nomenclature, D; Damaging, T; Tolerant, NS; Not Socred.