1290

UNIVERSIDADE Ð COIMBRA

Cátia Sofia Novais Brandão

INTERNSHIP REPORT AT THE ROYAL BROMPTON HOSPITAL |
CAN VARIANTS IN NON-CODING REGIONS LEAD TO PRIMARY CILIARY DYSKINESIA?

September 2022

Acknowledgments

First, I would like to express my sincere gratitude to Dr. Debbie Morris-Rosendahl for all her knowledge, constant support, and guidance throughout this placement. Thank you for your indispensable recommendations and comments on this dissertation.

A massive thank you to Lizi Briggs for all her patience, encouragement, guidance and all the practical and theoretical knowledge. Without your help the completion of the research project would never be possible.

I also would like to say a special thank you to Professor Joana de Melo for all her valuable feedback and to everyone who has been involved in this MSc program. Your knowledge and skills have prepared me to undertake and succeed at this placement.

To all of the CGGL team, who have been incredibly kind to me throughout the last nine months and have taken some of their time to explain and show me what their roles and daily tasks involve.

I must also extend a special thanks to the bioinformatics team at the CGGL as well as the PCD diagnostic team at the Royal Brompton Hospital, who have contributed immensely to the research project.

To all my university colleagues who have questioned, doubted, and despaired with me. Somehow, we got to the finishing line!

A massive thank you to my friends who have put up with my drama and overactive crisis. Who have always had the patience to hear me despair and have always motivated and cheered me up.

Finally, I would like to demonstrate my eternal gratitude to my parents and family who have always been there for me. To my mom, which has been my biggest support no matter what, who as showered me with her wise words when things weren't going as planned and who has always believed in me.

Resumo

Parte I

Estabelecido em 2015 e sediado no Royal Brompton Hospital, o Laboratório de Genética Clínica e Genómica (CGGL) oferece testes genéticos de diagnóstico a famílias e indivíduos em risco de doenças cardíacas e respiratórias hereditárias. O laboratório realiza testes para doenças como fibrose cística, aortopatias, vasculopatias, discinesia ciliar primária e hipercolesterolemia familiar, bem como cardiomiopatias e arritmias hereditárias.

Durante nove meses realizei o meu estágio no CGGL, tendo ido a oportunidade de acompanhar o trabalho diário realizado num laboratório de genética. Tive a oportunidade de trabalhar com vários softwares de análise de dados, bancos de dados de doentes e com o sistema CGGL LIMS. Além disso, pude participar nas reuniões semanais do laboratório, "jornal club" e reuniões de diagnóstico entre equipas multidisciplinares.

Este relatório descreve técnicas moleculares genéticas realizadas no CGGL de modo a obter um diagnóstico genético de doenças cardíacas e respiratórias hereditárias, nomeadamente Sequenciação de Sanger, MLPA, NGS e ddPCR. Adicionalmente, apresentam-se casos de doentes analisados - usando técnicas de modo a ilustrar as suas potenciais aplicações e respetiva análise de resultados.

Palavras-Chave

NGS, Sequenciação de Sanger, MLPA, ddPCR

Parte II

A discinesia ciliar primária (DCP) é uma doença caracterizada por heterogeneidade fenotípica e alélica e é herdada maioritariamente de forma autossómica recessiva. A maioria das variantes que causam PCD são privadas e o *clustering* é muito incomum. O diagnóstico da DCP depende de uma combinação de características clínicas, testes ciliares e testes genéticos moleculares conduzidos por sequenciação direcionada. Atualmente, quase um terço dos doentes permanece sem diagnóstico genético. Esta baixa taxa de diagnóstico pode ser explicada pela identificação de apenas uma variante patogénica ou a identificação de variantes de significado incerto.

Variantes não codificantes que desempenham um papel significativo na regulação de genes foram sugeridas como responsáveis por uma significativa parte das doenças genéticas raras. Variantes não codificantes penetrantes podem alterar o normal padrão de

splicing de mRNA, criando" cryptic donor/acceptor splice sites", resultando na inclusão de fragmentos intrónicos não funcionais – "pseudoexons". Funcionalmente, essas variantes podem resultar em "frameshift" ou a introdução prematura de codões stop e, subsequentemente, levar a "nonsense mediated decay" ou truncamento de proteínas.

Este projeto de investigação incluiu 23 casos de doentes previamente identificados heterozigotas para uma única variante patogénica ou provavelmente patogénica. Foram sequenciados genes inteiros dos usando um painel personalizado de NGS contendo as regiões codificantes e não codificantes dos genes relevantes. As variantes identificadas foram analisadas usando ferramentas de bioinformática e *in silico* "splice predictions". Mediante a disponibilidade de amostras, a caracterização do transcriptoma foi realizado usando RNA extraído de células epiteliais nasais.

Entre os 23 doentes, 4 variantes patogénicas foram identificadas, uma variante sinónima, uma variante intrónica, uma grande deleção genómica do cromossoma 16 e uma deleção do exão 9 no gene *CCDC39*. Outras 5 variantes potencialmente patogénicas foram identificadas, no entanto, análises moleculares complementares são necessárias para confirmar os resultados.

Um diagnóstico genético completo permitirá identificar portadores em membros da família, orientará o planeamento familiar e possibilitará a inclusão do doente em futuros ensaios clínicos relevantes de terapia genica. Além disso, os resultados deste projeto apoiam a relevância e utilidade da inclusão da avaliação de variantes não codificantes no diagnóstico molecular, no futuro.

Palavras-Chave

Discinesia ciliar primária," Whole gene sequencing", Regiões não codificantes, Análise *In Silico* de "splice predictions", "Splice Sites"

Abstract

Part I

Established in 2015 and based at The Royal Brompton Hospital, the Clinical Genetics and Genomics Laboratory (CGGL) offers diagnostic genetic testing for families and individuals at risk of inherited cardiac and respiratory disorders. The laboratory conducts testing for disorders such as cystic fibrosis, aorthopathies, vasculopathies, primary ciliary dyskinesia and familial hypercholesterolaemia, as well as inherited cardiomyopathies and arrhythmias.

Throughout nine months I conducted my placement at the CGGL, and I got the opportunity to follow the daily task performed at a genetics lab. I had the opportunity to work with several data analysis software's, patient databases and the CGGL LIMS system. Additionally, I got to assist to lab meetings, journal clubs and diagnostic multidisciplinary team meetings.

This report describes genetic molecular techniques performed at the CGGL in order to achieve a genetic diagnosis of inherited cardiac and respiratory disorders such as Sanger Sequencing, MLPA, NGS and ddPCR. Moreover, patient cases analysed using specific techniques have also been added to illustrate the techniques potential applications and their respective analysis of results.

Keywords

NGS, Sanger Sequencing, MLPA, ddPCR

Part II

Primary ciliary dyskinesia is a disorder characterized by phenotypic and allelic heterogeneity and it is primarily inherited in an autosomal recessive fashion. Most of the PCD-causing variants are private and clustering is very uncommon. Diagnosis of PCD relies on a combination of clinical features, a series of diagnostic features and molecular genetic testing conducted by targeted sequencing. At present, nearly one-third of the patients remain without a genetic diagnosis, either because results are negative, only one variant has been identified or their results demonstrate variants of unknown significance.

Noncoding variants which play a significant role in gene regulation have been suggested to account for a significant burden of causal variants in rare genetic diseases. Penetrant noncoding variants may disrupt the normal pattern of mRNA splicing by creating

cryptic donor or acceptor sites, which may result in the inclusion of nonfunctional intronic fragments – pseudoexons. Functionally, these variants can result in frameshift or introduction of premature termination codons and subsequently lead to nonsense mediated mRNA decay or protein truncation.

This research project investigated 23 cases of patients who had previously been found to be heterozygous for a single pathogenic or likely pathogenic variant. Whole genes were sequenced using a custom next-generation sequencing panel containing the coding and non-coding regions of the relevant genes. Variants identified were analysed using bioinformatic and *in silico* splice predictions tools. Where samples were available, transcript characterization was conducted using RNA extracted from nasal epithelial cells.

Out of the 23 patients, 4 disease causing variants were identified, a synonymous variant, an intronic variant, a large genomic deletion of chromosome 16 and a *CCDC39* exon 9 deletion. Another 5 potentially disease-causing variants were identified, however further confirmational molecular analysis is required.

A complete genetic diagnosis will allow carrier testing in family members, guide family planning and enable the inclusion of the patient in future relevant gene therapy procedures. Furthermore, the results of this project support the relevance and usefulness of the inclusion of the assessment of non-coding variants at the clinical setting, in the future.

Keywords

"Whole-gene" sequencing, PCD, Intronic Variants, Splice sites, *In Silico* splice prediction analysis

Contents

Abbreviations

AATD – **α**1- **A**nti**T**rypsin **D**eficiency

ACD – **A**lveolar **C**apillary **D**ysplasia

Array CGH – **A**rray **C**omparative **G**enomic **H**ybridisation

BP – **B**ase **P**airs

BS – **B**rugada **S**yndrome

CBF – **C**iliary **B**eat **F**requency

CBP – **C**iliary **B**eat **P**attern

cDNA – **c**omplementary **D**eoxyribo**N**ucleic **A**cid

CGGL – **C**linical **G**enetics and **G**enomics **L**aboratory

CNV – **C**opy Number **V**ariation

CPVT - **C**atecholaminergic **P**olymorphic **V**entricular **T**achycardia

CVD – **C**ardio**V**ascular **D**isease

ddNTPs -**did**eoxy**N**ucleotide **T**ri**P**hosphates

ddPCR – **D**roplet **D**igital **PCR**

DNA – **D**eoxyribo**N**ucleic **A**cid

dNTPs - **d**eoxy**N**ucleotide **T**ri**P**hosphates

EPR – **E**lectronic **P**atient **R**ecord

LIMS – **L**aboratory **I**nformation **M**anagement **S**ystem

FAM - **F**luorescein **AM**idites

FH – **F**amilial **H**ypercholesterolemia

FISH - **F**luorescence *In Situ* **H**ybridisation

GATK – **G**enome **A**nalysis **T**ool **K**it

GWAS – **G**enome **W**ide **A**ssociation **S**tudies

HEX - **HEX**achloro-fluorescein

HGP – **H**uman **G**enome **P**roject

HSVA – **H**igh **S**peed **V**ideo **A**nalysis

ICC – **I**nherited **C**ardiovascular **C**onditions

IDA – **I**nner **D**ynein **A**rm

IF – **I**mmuno**f**luorescence

IG – **I**nformation **G**overnance

ILD – **I**nterstitial **L**ung **D**isease

MAF – **M**inor **A**llele **F**requency

MLPA - **M**ultiplex **L**igation- **D**ependent **P**robe **A**mplification

MSc CLG – **M**aster's in **C**linical **L**aboratory **G**enetics

N-DRC – **N**exin-**D**ynein **R**egulatory **C**omplex

NGS – **N**ext **G**eneration **S**equencing

NGS- **N**ext **G**eneration **S**equencing

NMD – **N**onsense **M**ediated **D**ecay

nNO – **N**asal **N**itric **O**xide

NO – **N**itric **O**xide

ODA – **O**uter **D**ynein **A**rm

OME – **O**titis **M**edia with **E**ffusion

PCD – **P**rimary **C**iliary **D**yskinesia

PCR – **P**olymerase **C**hain **R**eaction

PH – **P**ulmonary **H**ypertension

RBH – **R**oyal **B**rompton **H**ospital

RIN – **R**NA **I**ntegrity **N**umber

RT – **R**everse **T**ranscription

SCD – **S**udden **C**ardiac **D**eath

SNP – **S**ingle **N**ucleotide **P**olymorphism

SS – **S**anger **S**equencing

SSNVs – **S**ynonymous **S**ingle **N**ucleotide **V**ariants

SV – **S**tructural **V**ariant

TEM – **T**ransmission **E**lectron **M**icroscopy

UKAS – **U**nited **K**ingdom **A**ccreditation **S**ervice

VUS – **V**ariant of **U**nknown **S**ignificance

WES – **W**hole **E**xome **S**equencing

WGS – **W**hole **G**enome **S**equencing

List of Figures

LIST OF TABLES

Introductory Note

The present coursework represents a curriculum internship report, conducted as an integral part of the master's in clinical laboratory genetics (MSc CLG), Faculty of Medicine, University of Coimbra. The MSc in CLG provides advanced specialised education in several areas of diagnosis and research in genetics. Furthermore, the annual internship is conducted with the aim of consolidating the acquired theoretical knowledge; the acquisition of new competences required to partake in the routine of a clinical genetics laboratory by validating and interpreting several methodologies and its results and; the development of both written and oral communication skills in the area of clinical laboratory genetics.

My internship began on the 9th of September of 2021 and was successfully concluded on the 31st of May of 2022. Throughout nine months I was given the opportunity to observe and conduct several research and diagnostic techniques at the Royal Brompton Hospital, one of the major UK specialist centres in cardiac and respiratory disorders.

I was initially introduced to the team and described the roles and responsibilities of each one of them. Moreover, I was explained the layout of the laboratory and introduced to several data analysis software's (e.g., Alamut) and patient databases (Electronic Patient Record - EPR). I began by observing sample reception and insertion of patient data into the laboratory's LIMS (Laboratory Information Management System) – "Clarity" (Illumina, USA) and "SQVD" (a CGGL custom database). I eventually began observing DNA extraction and other molecular techniques. Furthermore, I was also given the opportunity to assist to diagnostic MDT (Multidisciplinary Team Meetings) meetings where several PCD cases were discussed and to journal clubs where relevant cardiac and respiratory papers were presented and discussed.

Due to ethical reasons, I was also required to complete the Information Governance (IG) training. IG training is required and essential to try to prevent data breaches and it is a key aspect to help respect patients and other staff personal information and data security.

This report is structured in two main parts. The first part contains the techniques observed and conducted at the Clinical Genetics and Genomics Laboratory at the Royal Brompton Hospital. The second part focuses on a specific research project to which I was enquired about whether I would like to take part in. The research project is regarding a rare, respiratory disorder- primary ciliary dyskinesia, and was initiated by one of the genetic technologists, Lizi Briggs, while she was conducting her own MSc.

PART I

Techniques performed at the Clinical Genetics

and Genomics Laboratory at the RBH

1.1   The Royal Brompton Hospital

The Royal Brompton Hospital (RBH) situated in Chelsea, West London, was founded in 1841 by 25-year-old solicitor Philip Rose in the era before universal free healthcare. Initially built with the intention to admit and treat people with tuberculosis, the hospital is now a national and international leader in the treatment of cardiac and pulmonary disorders (RB&HH, 2021).

In 1964, the RBH established Europe's first adult cystic fibrosis clinic and it is currently the largest centre in Europe for the treatment and management of this disorder. There are over 3600 staff members at RBH, working across 2 hospital sites in London and Harefield, 11 dedicated operating theatres, one hybrid theatre, 9 catheter laboratories and attending 380 patients' beds over surgical, intensive care, respiratory and cardiology areas including paediatric and paediatric intensive care patients (Royal Brompton and Harefield hospitals, 2021).

1.2   Clinical Genetics and Genomics Laboratory

Established in 2015 and based at The Royal Brompton Hospital, the Clinical Genetics and Genomics Laboratory (CGGL) offers diagnostic genetic testing for families and individuals at risk of inherited cardiac and respiratory disorders. In 2018, the laboratory became part of the South East Genomic Laboratory Hub, based at Guy's and St. Thomas' NHS Foundation Trust, as part of a new national NHS Genomics Medicine Service.

The laboratory is accredited by the United Kingdom Accreditation Services (UKAS) to ISO 15189 standards and offers next-generation sequencing including copy number variant analysis, Sanger sequencing, Multiplex Ligation-Dependent Probe Amplification (MLPA), droplet digital PCR (ddPCR), microsatellite analysis and targeted mutation analysis for common pathogenic variants in the *CFTR* gene for the diagnosis of cystic fibrosis (CF), improving the diagnosis of inherited cardiac and respiratory disorders. The laboratory conducts testing for disorders such as CF, aorthopathies, vasculopathies, primary ciliary dyskinesia and familial hypercholesterolaemia, as well as inherited cardiomyopathies and arrhythmias.

Tests are often performed on blood samples collected in EDTA tubes; however, saliva and tissue samples may also be used. A request form (figure 1) must be completed, and patient information and referral information must also be provided in full and accompanied by a consent form. Assessment and reporting of detected variants follow the guidelines of the American College of Medical Genetics and the UK association of Clinical Genomic Science, as well as up-to-date literature (Royal Brompton and Harefield hospitals, 2021).

NEXT GENERATION SEQUENCING - *Testing for the conditions below utilises Next Generation Sequencing (NGS). Data will be generated and stored on all genes in each panel. Comprehensive bioinformatic analysis, including copy number variant analysis, clinical interpretation and variant confirmation will be reported only on the genes of clinical relevance to the disease category requested below.*

Inherited Cardiac and Respiratory Diseases

For full details of the genes included on each subpanel please refer to our website: www.rbht.nhs.uk/ggl

Aortopathy and connective tissue genes
- ☐ Alport syndrome, X-linked (*COL4A5*)
- ☐ Cutis laxa (~4 genes)
- ☐ Ehlers-Danlos syndrome (EDS) (~15 genes)
- ☐ Familial thoracic aortic aneurysm (FTAA) (~26 genes)
- ☐ Loeys-Dietz syndrome (LDS) (~5 genes)
- ☐ Marfan syndrome (MFS) (*~5 genes*)
- ☐ Weill-Marchesani syndrome (*ADAMTS10, ADAMTS17, LTBP2*)
- ☐ All Aortopathy and connective tissue genes (~63 genes)

Arrhythmia genes
- ☐ Andersen-Tawil syndrome (*KCNJ2*)
- ☐ Brugada syndrome (BrS) (~13 genes)
- ☐ Catecholaminergic polymorphic ventricular tachycardia (CPVT) (~4 genes)
- ☐ Long QT syndrome (LQTS) (~14 genes)
- ☐ Short QT syndrome (~6 genes)
- ☐ All Arrhythmia genes (~38 genes)

Cardiomyopathy genes
- ☐ Arrhythmogenic cardiomyopathy (ACM) (~14 genes)
- ☐ Arrhythmogenic right ventricular dysplasia/cardiomyopathy (ARVD/C) (~8 genes)
- ☐ Dilated cardiomyopathy (DCM) (~38 genes)
- ☐ Hypertrophic cardiomyopathy (HCM) (~29 genes)
- ☐ Laminopathy (*LMNA*)
- ☐ Noncompaction cardiomyopathy (LVNC) (~8 genes)
- ☐ Fabry disease (*GLA*)
- ☐ All Cardiomyopathy genes (~88 genes)

- ☐ **Familial Hypercholesterolemia (FH) (~4 genes + 14 SNPs)**

Other cardiac conditions and genes
- ☐ Alagille syndrome (*JAG1*)
- ☐ Carney complex (*PRKAR1A*)
- ☐ Heterotaxy/situs ambiguous (HTX) (~30 genes)
- ☐ Holt-Oram syndrome (*TBX5*)
- ☐ *NKX2-5*-related disorders
- ☐ Noonan spectrum disorders (~11 genes)
- ☐ *SALL4*-related disorders

Vasculopathy genes
- ☐ Birt-Hogg-Dubé syndrome (Primary spontaneous pneumothorax) (*FLCN*)
- ☐ Capillary malformation-arteriovenous malformation/Parkes-Weber syndrome (*RASA1*)
- ☐ Hereditary Haemorrhagic Telangiectasia (HHT) (~4 genes)
- ☐ Homocystinuria (*MTHFR, CBS*)
- ☐ Microcephaly Capillary Malformation syndrome (*STAMBP*)
- ☐ Venous Malformations (*GLMN, TEK*)
- ☐ All Vasculopathy genes (~12 genes)

Bronchiectasis genes
- ☐ Cystic Fibrosis targeted mutation analysis - 36 most common *CFTR* mutations in EU populations
- ☐ Sequencing of the *CFTR* gene (exons)
- ☐ Non-CF Bronchiectasis (4 x ENAC genes)
- ☐ Primary Ciliary Dyskinesia (PCD) (~43 genes)
- ☐ All Bronchiectasis genes (~48 genes including PCD genes and *CFTR*)

Ciliopathy genes
- ☐ Joubert syndrome (JS) (~20 genes)
- ☐ Orofaciodigital syndrome (OFD) (~6 genes)
- ☐ Short rib thoracic dysplasia (Jeune syndrome) (SRTD) (~13 genes)
- ☐ All Ciliopathy genes (including PCD) (~76 genes)

Congenital respiratory condition genes
- ☐ Alveolar capillary dysplasia (*FOXF1*)
- ☐ Ataxia telangiectasia (*ATM*)
- ☐ Central Hypoventilation syndrome (~7 genes)
- ☐ Periventricular nodular heterotopia and lung disease (*FLNA*)
- ☐ Primary pulmonary hypoplasia (*ZFPM2*)
- ☐ Pulmonary alveolar microlithiasis (PAM) (*SLC34A2*)
- ☐ All Congenital respiratory condition genes (~12 genes)

***Emphysema* genes**
- ☐ Alpha-1-Antitrypsin deficiency (AAT) (*SERPINA1*)
- ☐ All Emphysema genes (~5 genes)

Immunodeficiency genes
- ☐ Agammaglobulinemia (*PIK3R1, BTK*)
- ☐ Autoimmune lymphoproliferative syndrome (*CTLA4*)
- ☐ Autoinflammation, antibody deficiency and immune dysregulation syndrome (*PLCG2*)
- ☐ Candidiasis, familial (*CARD9, IL17R, IL17F*)
- ☐ Hyper-IgE recurrent infection (*STAT3, DOCK8*)
- ☐ Immunodeficiency, common variable (~20 genes)
- ☐ Immunodysregulation, polyendocrinopathy & enteropathy (*FOXP3*)
- ☐ Susceptibility to Aspergillosis (*CLEC7A*)
- ☐ All Immunodeficiency genes (~31 genes)

Interstitial Lung Disease (ILD) genes
- ☐ Childhood ILD (ChILD) including surfactant genes (~7 genes)
- ☐ Hermansky-Pudlak Syndrome (HPS) (~8 genes)
- ☐ Pulmonary fibrosis, familial (FPF) (~26 genes)
- ☐ Tuberous sclerosis (TS) (*TSC1, TSC2*)
- ☐ All Interstitial Lung Disease (ILD) genes (~36 genes)

- ☐ **Molecular autopsy (Sudden Cardiac Death, SCD) (~115 genes)**
- ☐ **Pulmonary Hypertension (~6 genes)**

- ☐ **All Inherited Cardiac Condition genes (~169 genes)**
  Only available after discussion with the laboratory
- ☐ **All Inherited Respiratory Condition genes (~171 genes)**
  Only available after discussion with the laboratory

TESTING FOR A KNOWN FAMILIAL VARIANT:
Please provide a copy of the familial report or full details of the proband if tested at RBH

- ☐ Diagnostic/confirmatory testing (has phenotype consistent with familial disease-causing variant)
- ☐ Predictive/pre-symptomatic testing (has no or unknown phenotype)
- ☐ Family studies (for variant interpretation)   **Variant details:**

☐ **DNA STORAGE ONLY (no test will be performed until requested)**

Samples and completed forms should be sent to the lab (address overleaf) packaged appropriately according to UN3373 guidelines. All samples should be sent by first class post, courier or hospital transport.

Figure 1. Request form of the Clinical Genetics and Genomics Laboratory

1.3 Inherited Cardiovascular disorders

Cardiovascular disease (CVD) is a broad, umbrella term used to describe several linked pathologies which are commonly defined as coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic and congenital heart diseases and venous thromboembolism (Stewart, Manmathan and Wilkinson, 2017). Accounting for over 3.8 million deaths every year, CVD is the most common cause of death in the European region (Townsend et al, 2022).

In recent years there has been an increased recognition of genetic causes for many types of cardiovascular disease, such as cardiomyopathies, arrhythmic disorders, channelopathies, vascular disorders, and lipid disorders such as familial hypercholesterolemia. Genetic testing is informative and presents with significant implications for patient management depending on the specific genetic condition (Otto, Savla and Hisama, 2020).

At present, there are over 40 inherited cardiovascular conditions (ICCs), comprising a wide and heterogenous spectrum of diseases of the heart, that have been proven to occur due to single-gene defects. ICCs span all aspects of cardiovascular disease and affect all parts of the heart structure (Kelly and Semsarian, 2009). With a combined estimated prevalence ranging from 1 in 200 to 10000 in the general population, ICCs are often characterised by marked clinical, genetic and allelic heterogeneity (Blanch et al, 2017). Furthermore, they often present with an autosomal dominant inheritance pattern and are associated with highly variable expressivity and penetrance (Pua et al, 2016).

ICCs may manifest at any age and individuals with a hereditary predisposition to the disease can remain asymptomatic or may also manifest symptoms of severe heart failure and arrhythmias (Van den Heuvel et al, 2019). ICCs are common causes of sudden cardiac death (SCD) in young people (Cadrin-Tourigny and Tatros, 2022). SCD is defined as a non-traumatic, non-violent, unforeseen, and unexpected heart function loss which occurs within 1 hour of symptom onset (Primorac et al, 2021). Nearly half of all SCD victims have no previously diagnosed heart disease, and therefore in several cases SCD presents as the first manifestation of an underlying cardiac disease (Holkeri et al, 2020). Early identification of individuals at risk is therefore crucial to prevent a lethal episode (Coll et al, 2018).

Due to the extraordinary progress achieved in molecular genetics over the last decades at identifying the molecular basis of many cardiac disorders, genetic testing is now considered a main component of clinical management of inherited cardiovascular conditions. In current practice, genetic testing can be used to confirm a diagnosis on a clinically symptomatic patient;

to clarify the underlying aetiological basis of a disease, particularly in cases where cardiac manifestations are part of wider syndromes; to elucidate disease risk amongst clinically unaffected relatives by conducting cascade screening and can also be useful for molecular autopsy in cases of sudden unexpected death (Girolami et al, 2018; Ingles et al, 2020).

Owing to the heterogenous nature of cardiac disorders with several overlapping genes, and taking into consideration family history and clinical screening, genetic testing may range from single -variant tests for predictive screening to a few genes most likely involved or even a large multigene panel (Cirino et al, 2017). Table 1 lists the major genes associated with the most common inherited cardiovascular diseases.

Table 1 List of genes to be considered for genetic testing for the most frequent inherited cardiovascular diseases. Adapted from Musunuru et al, 2020

Condition	Genes
HCM, definitive evidence	*MYBPC3, MYH7, TNNT2, TNNI3, TPM1, ACTC1, MYL2, MYL3*
HCM, moderate evidence	*CSRP3, TNNC1, JPH2*
HCM, definitive syndromic genes for which isolated left ventricular hypertrophy can be seen	*PLN, CACNA1C, DES, FHL1, FLNC, GLA, LAMP2, PRKAG2, PTPN11, RAF1, RIT1, TTR*
DCM	*TTN, LMNA, MYH7, TNNT2, BAG3, RBM20, TNNC1, TNNI3, TPM1, SCN5A, PLN*; for testing, all HCM and ARVC genes are recommended to be included
ARVC	*DES, DSC2, DSG2, DSP, JUP, LMNA, PKP2, PLN, RYR2, SCN5A, TMEM43, TTN*; consider full DCM panel
Restrictive cardiomyopathy	*TTR*; consider HCM or DCM panel
LVNC	Use the gene panel for the cardiomyopathy identified in association with the LVNC phenotype
Long-QT syndrome	*KCNQ1, KCNH2, SCN5A*
Short-QT syndrome	*KCNH2, KCNQ1, KCNJ2*
Brugada syndrome	*SCN5A*
Catecholaminergic polymorphic ventricular tachycardia	*RYR2, CASQ2*
HTAD, definitive or strong evidence	*ACTA2, COL3A1, FBN1, MYH11, SMAD3, TGFB2, TGFBR1, TGFBR2, MYLK, LOX, PRKG1*
HTAD, potentially diagnostic	*EFEMP2, ELN, FBN2, FLNA, NOTCH1, SLC2A10, SMAD4, SKI*
FH	*LDLR, APOB, PCSK9*
Phenotypic overlap with FH	*LDLRAP1, LIPA, ABCG5, ABCG8, APOE*

ARVC – arrhythmogenic right ventricular cardiomyopathy; DCM – dilated cardiomyopathy; FH – familial hypercholesterolemia; HCM – hypertrophic cardiomyopathy; HTAD – heritable thoracic aortic aneurysm or dissection; LVNC – left ventricular noncompactation

At present, the CGGL provides NGS gene panels for the diagnosis of cardiomyopathies, arrhythmias, vasculopathies, familial hypercholesterolemia and aorthopathies and connective tissue disorders (Figure 1). Additionally, cascade screening may also be conducted by Sanger Sequencing to test for familial cardiovascular variants.

Familial hypercholesterolemia (FH) is a common genetic disorder of lipoprotein metabolism resulting in elevated serum low-density lipoprotein cholesterol levels that lead to atherosclerotic plaque deposition in the coronary arteries and proximal aorta contributing to an increased risk for premature cardiovascular disease (Varghese, 2014). FH predominantly follows an autosomal inheritance pattern, with variants on *LDL-R, APOB* and *PCSK9* representing the major causes of the condition. Rarely, variants in *LDLRAP* which follow a recessive inheritance pattern maybe also lead to a phenotypic expression of FH (Vrablik et al, 2020).

Arrhythmias are the abnormalities or perturbations in the normal activation or beating of heart myocardium and may be characterised by irregular heartbeats, skipped beats, tachycardia, or bradycardia (Fu, 2015). Long QT syndrome, short QT syndrome, catecholaminergic polymorphic ventricular tachycardia (CPVT) and Brugada syndrome (BS) result from mutations in several genes encoding ion channels or proteins involved in their regulation. These are rare disorders which account for the main inherited cardiac arrthymias and are often the underlying cause of sudden cardiac death in the young (Schwartz et al, 2020).

Cardiomyopathies are a heterogenous group of diseases of the myocardium associated with mechanical and/or electrical dysfunction that usually exhibit inappropriate ventricular hypertrophy or dilation (Maron et al, 2006). Manifestations of cardiomyopathy can range from microscopic alterations in cardiac myocytes to fulminant heart failure with inadequate tissue perfusion, fluid accumulation and cardiac rhythm dysfunction (Brieler, Breeden and Tucker, 2017). Cardiomyopathies may result from an array of factors, mainly genetic, and are classified according to ventricular morphology and pathophysiology (McKenna and Elliott, 2020).

Vasculopathies represent a wide variety of disorders that affect the blood vessels often resulting in structural vessel wall changes such as arterial stenosis, vessel wall thickening, luminal irregularities, occlusion, dissection, aneurysm formation and/or dilation causing multi-organ ischaemia and significant cardiac and cerebral complications (Khosla et al, 2016). The genetic causes of vascular disease can either alter the structure or function of

veins and arteries and are often associated with abnormalities of the vasculature in several parts of the body or additional syndromic features (Yonker et al,2020).

Aorthopathies and connective tissue disorders – The inherited aorthopathies denote a group of conditions characterised by aortic wall weakness and abnormal aortic hemodynamic profiles which consequently lead to aortic dilation, aneurysm formation and acute aortic complications which can often lead to fatal consequences (Fletcher et al, 2020). The connective tissue disorders are a heterogenous group of disorders brought about by defects in structure and synthesis of extracellular matrix elements such as collagen, elastin, as well as proteins that modify these component elements and proteins that affect protein glycosylation (Murphy-Ryan, Psychogios and Lindor, 2010). The aorthopathies and connective tissue disorders may present as syndromic and nonsyndromic forms and occur due to one of the following pathomechanisms: perturbation of the TGF-β signalling pathway, disruption of the vascular smooth muscle cell contractile apparatus, and impairment of extracellular matrix synthesis (Andelfinger, Loyes and Dietz, 2015; Renner et al, 2019).

Although, genetic testing may be helpful in guiding therapy and assessing prognosis, particularly for long QT syndrome (Wilde and Behr, 2013), the psychosocial consequences may they be due to uncertainty, lifetime periodic clinical surveillance or other factors must always be considered. Therefore, when performing predictive genetic testing, genetic counselling must always be conducted *a priori*.

1.4 Inherited respiratory disorders

Respiratory diseases are diseases of the airways and other structures of the lung and are among the leading causes of morbidity, mortality, economic and societal burden, and disability adjusted life years (Shukla et al, 2020). The aetiology of respiratory diseases may be due to interactions between genotype and environmental factors such as allergens, smoking, diet, and drugs. Examples of this multifactorial disorders include asthma, chronic obstructive pulmonary disease, pulmonary fibrosis and sarcoidosis. Alternatively, respiratory diseases may be classified as a monogenic/ Mendelian disorder in which the clinical phenotype is brought about by single gene variation. Common examples comprise cystic fibrosis and α1-antitrypsin deficiency (European lung, 2022). Due to the wide spectrum of clinical presentation of monogenic diseases in respiratory medicine this may be further subclassified into: pulmonary disorders, comprising airway disease, pulmonary parenchymal disease, and pulmonary vascular disease; sleep disorders; and monogenic diseases with respiratory involvement, including primary deficiency disease, neuromuscular disease, certain syndromes, and inherited metabolic diseases (Yao and Shen, 2017).

Advances in technology have improved the recognition of genetic aetiologies of disease which has in turn impacted diagnosis and management of respiratory disease patients. At present, the CGGL provides NGS gene panels for the diagnosis of ciliopathies, bronchiectasis, congenital respiratory conditions, emphysema, immunodeficiency, interstitial lung disease and pulmonary hypertension (Figure 1). Moreover, cascade screening is also available by Sanger Sequencing for known familial variants.

Ciliopathies comprise a heterogenous group of disorders which occur due to the dysfunction of motile and/or non-motile cilia or by abnormal cilia biogenesis. Ciliopathies entail a group of around 35 reported disorders and are caused by a multitude of largely unrelated genes (Modarage, Malik and Goggolidou, 2022). The effects of the abnormal formation of function of cilia are seen in several organs with prevalent phenotypes including polycystic kidney disease, retinal degeneration, obesity, skeletal malformations, and brain abnormalities (Lee and Gleeson, 2011; Reiter and Leroux, 2017). Furthermore, along with a Mendelian inheritance pattern, oligogenicity, genetic modifications, epistatic interactions and retrotransposon interactions have also been described when defining the ciliopathy phenotype (Focșa, Budișteanu and Bălgrădean, 2021). Examples of ciliopathies include Joubert syndrome, Jeune syndrome and primary ciliary dyskinesia which will be further described in detail later on in this dissertation.

Bronchiectasis is a heterogenous chronic condition characterised by abnormal dilation of one or more bronchi and a damaged epithelium. Airway dilation can lead to failure of mucus clearance and facilitate bacterial infections (Derbyshire and Calder, 2021). Bronchiectasis presents with a broad array of clinical symptoms, ranging from asymptomatic radiological changes detected by incidentally to chronic sputum production and recurrent exacerbations (Macfarlane et al, 2021).

Several different conditions known to cause or be associated with bronchiectasis have been described, however most of them present with common pathophysiologic pathways leading to the remodelling of the airways and dilation (Flume, Chalmers and Olivier,2018). Bronchiectasis may occur due to conditions such as chronic obstructive disease, obstruction of single airway or post-infectious bronchiectasis (Börekçi and Müsellim, 2021). Moreover, immunodeficiency, impairments of mucociliary clearance and intrinsic airway anomalies have been given increased significance as determinants of disease (Gold, Freeman and Olivier, 2012). Additionally, from a genetic point of view, the main causes of bronchiectasis include ciliary abnormalities such as primary ciliary dyskinesia; abnormal epithelial chloride transport, which is a characteristic of cystic fibrosis; abnormal sodium transport as seen in pseudo-hypoaldosteronism due to mutations in *SCNN1A, SCNN1B* and *SCNN1G*; and primary immunodeficiencies which lead to impaired immune responses to infection (Yonker et al, 2020).

Emphysema is defined as irreversible enlargement of airspaces distal to the terminal bronchioles due to alveolar wall destruction without obvious fibrosis (McLoud and Boiselle, 2010). α1- Antitrypsin deficiency (AATD) is a hereditary disorder and it is one of the most common genetic causes of emphysema. The main pathobiological mechanism implicated in the development of AATD-related emphysema, is caused by a protease-antiprotease imbalance when smoking-induced release of neutrophil elastase in the lung is inadequately inhibited by the deficient levels of AAT (Abboud, Ford and Chapman, 2005; Goel et al, 2021).

Interstitial lung disease (ILD) is an umbrella term used for a large group of diffuse parenchymal lung disorders of known or unknown cause, with varying degrees of inflammation or fibrosis (Wong, Ryerson and Guler, 2020). ILDs encompass more than 100 different pulmonary diseases with many clinical, radiological and physiological similarities (Garcia, 2004). ILDs can be subdivided into 2 major groups: systemic disorders such as tuberous sclerosis, neurofibromatosis, Hermansky-Pudlak syndrome and Gaucher disease; and disorders that primarily affect the lung such as surfactant dysfunction disorders and familial pulmonary fibrosis. Interstitial disease can therefore originate from different pathobiological mechanisms associated with a broad variety of genetic variants (Devine and Garcia, 2012).

Immunodeficiency diseases, represent a heterogenous group including over 350 distinct disorders. These are caused by a quantitative and/or functional disorder of the immune system, which may lead to a greater risk of infections, immune dysregulation, inflammation, and autoimmune phenomena. The underlying causes of the defects may involve different mechanisms of the innate and/or adaptive immune response and subsequently this may lead to a broad spectrum of respiratory phenotype (Cinetto et al, 2018; Soler-Palacín et al, 2018).

Pulmonary hypertension (PH) is a hemodynamic and pathophysiologic state defined by an improper elevation of pressure in the pulmonary vascular system leading to right ventricular failure and premature death (Simoneau et al, 2013; Austin and Loyd, 2014). PH can be subclassified into: pulmonary arterial hypertension, PH due to left-sided heart disease, PH due to chronic lung disease; chronic thromboembolic PH and PH with an unclear and/or multifactorial mechanisms (Mandras, Mehta and Vaidya, 2020).

Congenital respiratory conditions are a group of rare and extremely heterogenous conditions. The clinical spectrum of this group of disorders is wide and includes clinical features such as respiratory distress, calcific deposits in lung parenchyma, and congenital malformations/underdevelopment of the lungs and blood vessels (Liners, Médart and Collignon, 2019; Chopra, Tendolkar and Vardhan, 2019).

1.5 Techniques observed

1.5.1 MLPA

Genomic variation is both a source of phenotypic diversity and a cause of many genetic disorders. It has been long thought that single nucleotide polymorphisms (SNPs) account for most of the variation is humans. Yet, with approximately 12% of the genome in human populations being subjected to copy number changes, copy number variation (CNV) is now thought to have equal contribution (Zhao et al, 2013). CNVs are a group of structural variants of DNA segments, displayed as microscopically invisible deletions or duplications ranging from 1 kilobase to several megabases in size. The discrepancy in the dosage of genomic segments is brought about by several structural variations within the genome such as deletions, duplications, insertions or unbalanced translocations and inversions (Shaikh, 2017). CNVs may include genes and/or noncoding sequences, and they account for morphological variation, altered metabolic states, susceptibility to infectious diseases, host-microbiome interactions, as well as rare genetic disorders/syndromes (Pös et al, 2021).

Ever since CNVs were shown to be causative of Mendelian traits in the early 1990s (Zhang et al, 2009), several thousand of CNVs have now been catalogued and shown to be involved in the aetiology of monogenic and complex disorders such as autism, Alzheimer's disease, cancer, epilepsy and schizophrenia. Over the years, different techniques have been introduced to detect genomic imbalances. Since neither conventional cytogenetics analysis or DNA sequencing is able to detect gene deletions/duplications and CNVs, the detection of these structural alterations was initially conducted by Southern Blotting or Fluorescence *in situ* hybridisation (FISH). However, these techniques have proven to be expensive, time-consuming, labour intensive and unable to detect small intragenic rearrangements (Ceulemans, Van der Ven and Del-Favero, 2011). At present, the gold standards for CNV detection in genetic diagnosis are multiplex ligation-dependent probe amplification and array comparative genomic hybridisation (aCGH). However, and even though array CGH provides a genome wide CNV screening capability, its resolution is low and it often requires confirmation and validation by alternative quantitative PCR methods (Stuppia et al, 2012; Kerkhof et al, 2017).

Multiplex ligation-dependent probe amplification (MLPA®, MRC Holland), is a high-throughput semiquantitative method developed to determine variations in the copy number of up to 60 genomic DNA sequences in a single multiplex PCR-based reaction (Eijk-Van Os and Schouten, 2011). Aside from determining the presence of deletions or duplications, MLPA can also be used in the prenatal molecular diagnosis of genetic of aneuploidies, common

microdeletion syndromes and subtelomeric copy-number changes, identification of marker chromosomes. Moreover, MLPA may also be used in specific circumstances to evaluate methylation changes in genetic imprinting disorders and promoter methylation status in cancer (Willis, Van den Veyver and Eng, 2012).

MLPA kits are designed by MRC Holland, and at present there are over 350 probe sets commercially available. Current kits in use at the RBH include:

- DNAH5 P238-B2 – contains one probe for each DNAH5 exon. Currently used for PCD patients.
- TGFBR1-TGFBR2 P148-B3 – contains probes for the TGFBR1 and TGFBR2 genes, including upstream region.
- FBN1 P065-B1/ P066-B2 – The P065-B1/ P066-B2 probe mixes contain one probe for each exon of the FBN1 gene. Moreover, one additional probe for exon 2, 65 and 66 are included in these probe mixes. Furthermore, the P065-B1 probe mix contains one probe for each exon of the TGFBR2 gene, with one extra probe for exon 1.
- TNNt2- BAG3 P196-B1 – Contains one probe of each exon for TNNT2 (except exon 13) and one probe for each exon of BAG3.
- DMD P034-B2 and P035-B1 – Cover each exon of the DMD gene, spread over 2 kits
- CFTR P091-D2 – Covers each exon of the CFTR gene, and detects the common pathogenic p.(Phe508del) variant.
- FOXF1 P431-A1 – Covers several alveolar capillary dysplasia (ACD) related genes, although only FOXF1 causes isolated ACD, and has two probes per exon of FOXF1 (2 exons) and covers the upstream shared deleted region, deletions of which can cause ACD.

The basis of the MLPA technique can be divided into 6 steps: denaturation, hybridisation, ligation, amplification, capillary electrophoresis, and data analysis. The laboratory protocol spans over 2 days and requires as little as 20ng of DNA to be performed. Not only it is a quick and easy technique, it also can be conducted in most laboratories since it only requires a thermocycler and capillary sequencing equipment to be performed. The protocol includes the use of normal control DNA samples, negative controls and if available will also include the use of a positive control.

In this method, DNA is denatured and incubated overnight with a mixture of gene-specific probes. These probes consist of two immediately adjacent oligonucleotides hemiprobes (a synthetic one and one derived from the M13 bacteriophage) per target exon, each containing

one of the PCR primer sequences. One of the probes also contains a nonhybridizing "stuffer sequence" at the 3' end to give the PCR product of desired length.

Following hybridisation, probes are ligated by a DNA ligase enzyme, and the fragments are amplified by PCR using a fluorescently labelled universal primer pair. Only those pairs that have both hybridised to their adjacent target sequences are ligated. Amplification products, typically between 130 and 480bp in length, can then be separated by size, identified and quantified by capillary electrophoresis (using the ABI 3500, Applied Biosystems, USA) due to the differences in length of the "stuffer" sequences. The number of probe ligation products correlates directly to the number of target sequences in the sample. Deletions and duplications of the targeted regions are detected by comparing probe peaks to the relative probe peak heights of reference probes and peak heights of reference samples. A relative decrease (below 0.7) in the peak height of a particular probe pair reflects aa deletion of that target region, whilst a relative increase (above 1.4) in peak height reflects a duplication of that target region.
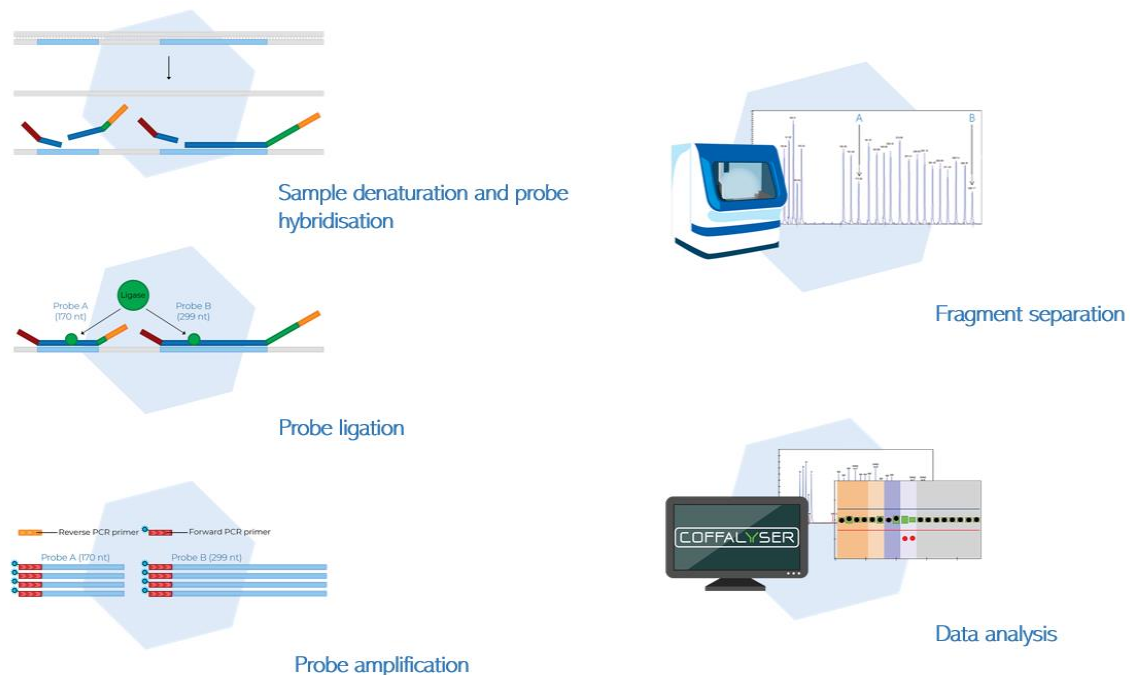
Figure 2. Overview of the steps involved multiplex ligation-dependent probe amplification technology for copy number detection. Images available from MRC Holland, 2022

Although MLPA will detect heterozygous and homozygous deletions and duplications it may not accurately define the numbers of duplicated/deleted copies. It is also unable to detect point, cell ploidy alterations, low levels of mosaicism, and balanced rearrangements. MLPA is also very sensitive to contaminants and PCR inhibitors such as phenols. Also, it generally does not cover the entire coding regions of all exons of a specific gene; the probes are targeted to certain regions of genes which have previously been shown to harbour CNVs Moreover, probe signals can also be affected by SNPs under primer sites, and therefore there is a low risk of a false positive result.

1.5.1.1 Patient cases

Patient A, a 30-year-old woman with some minor repolarization abnormalities, underwent molecular genetic testing due to a family history of long QT syndrome. Affected relatives have been found to be heterozygous for a deletion of *KCNQ1* exon 5. MLPA analysis of the KCNQ1 gene was conducted using the MRC Holland SALSA MLPA Probemix P114 Long-QT. The deletion was not detected on patient A, and she is therefore not predicted to be at an increased risk to developed LQTS caused by it, however she remains at population risk.
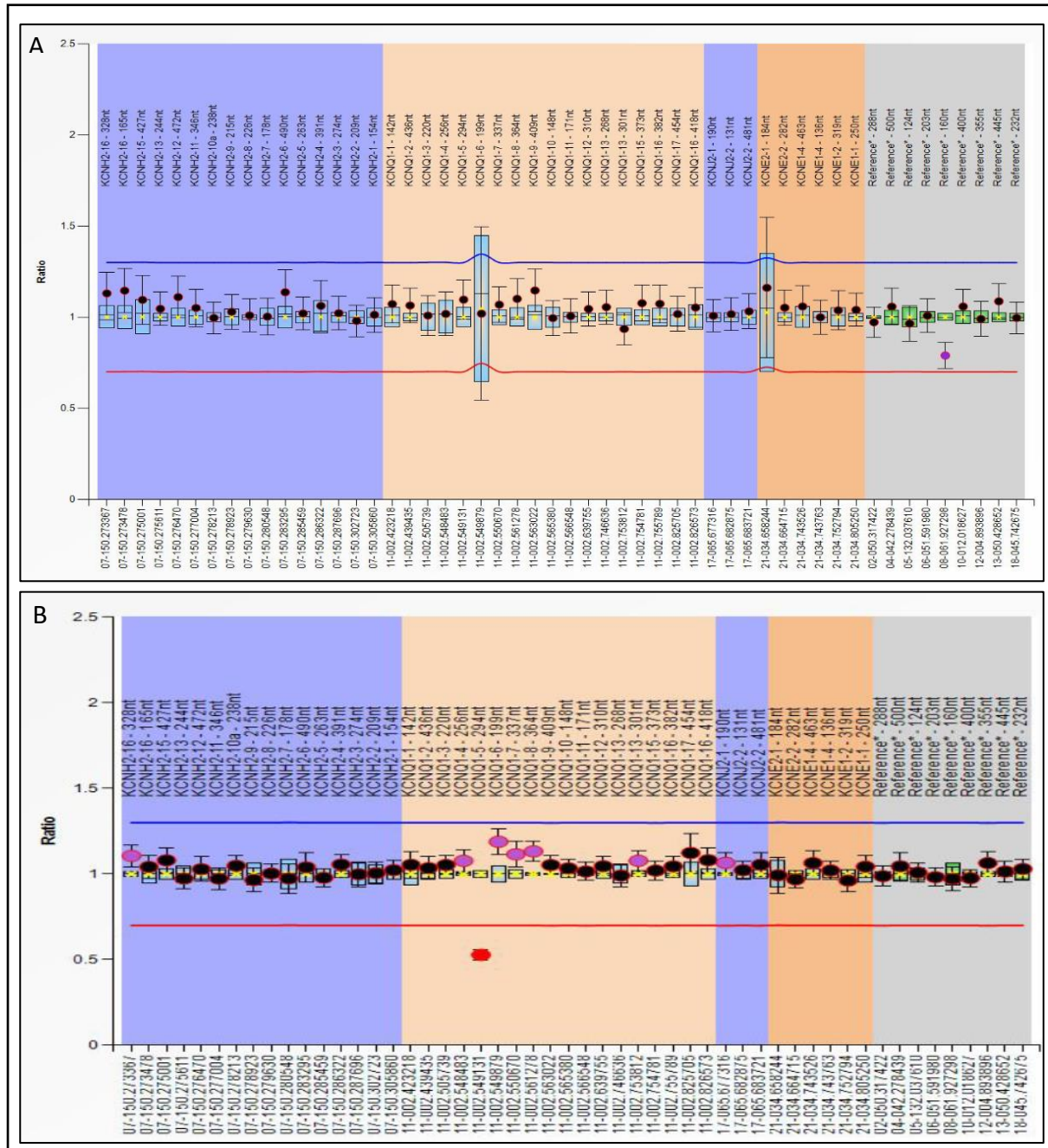
Figure 3.. Multiplex Ligation-dependent Probe Amplification (MLPA) for gene dosage of KCNQ1 gene. A- Result of patient A showed a normal ratio (1.1) for exon 5 of KCNQ1 gene; B- Result of previously tested relative showed a heterozygous deletion (ratio =0.53) of exon 5 of KCNQ1 gene.

1.5.2 Droplet Digital PCR

Droplet-digital PCR (Bio-Rad, USA) is a highly sensitive and specific technique, with absolute quantification without a standard curve, high reproducibility, good tolerance to PCR inhibitor, and high efficacy compared to conventional molecular methods (Kojabad et al, 2021). ddPCR is currently used for absolute allele quantification, rare mutation detection, analysis of copy number variations, DNA methylation, and gene rearrangements in different kinds of clinical samples (Olmedillas-López, García-Arranz and García-Olmo, 2017).

ddPCR is a method available since 2011 which allows a direct quantification of genetic material. The technique relies on the creation of micro-reaction units and microfluidic technology. It depends on division of the reaction batch prior to amplification into several thousand separate oil-droplet-based PCR reactions to quantitatively measure changes in a DNA sample (Kuhlmann, Cieselki and Schumann, 2021). The technique is accurate down to the single molecule level and due to sample partitioning, each droplet has a unique fluorescent signature, since each droplet can be read and recorded individually.

ddPCR uses fluorescein amidites (FAM) fluorescence and hexachloro-fluorescein (HEX) labelled probes to show differences between genes of interest and a reference gene through simple exon PCR reactions. The gene of interest has a complementary fluorescently labeled DNA probe created for the region of interest plus a forward and reverse primer for the PCR reaction. The probe is labelled with FAM fluorescent dye, whereas the reference gene used for comparison is labelled with HEX (Vessies et al, 2021). The reference gene used is RPP30, which is a housekeeping gene with no known copy number variation (tested and checked in the ExAC CNV control population of 38250 individuals: 1 deletion of exons 1-6 is noted in this population, therefore a probe covering exon 7 is used), recommended by the manufacturer. Following PCR amplification, the absolute concentration of the target of interest in terms of the number of copies per microliter, becomes possible to determine based on the positive and negative fraction and following the Poisson distribution (Villamil et al, 2020).

1.5.2.1   Probe and primer design

Probes can be designed using the primerquest design tool available at the IDT website https://sg.idtdna.com/site . These should have <30 nucleotides between the fluorophore and the quencher and they must not have a G at its 5' end. Furthermore, the T_m of each hydrolysis probe should be 5-10° C higher than that of the corresponding primers and the GC content must be between 30-80%. The probe sequence must also be checked for any possibility of

sequence similarity to other areas of the genome using the BLAST function available at www.ensembl.org.

Primers should have a GC content of 50-60% and a T_m between 50 and 65°C. Primer design tools are also available at the IDT website. The specificity of the primers for the target sequencing can also be assessed using BLAST. Forward and reverse primers must also be checked for SNPs using SNP checker https://secure.ngrl.org.uk/SNPCheck/snpcheck.htm. A SNP can be included if it has a frequency of less than 1/2000 (Bio-Rad, 2022).

1.5.2.2 ddPCR procedure and data analysis

Prior to droplet generation, DNA fragmentation using restriction enzymes such as EcoRI and BamHI is conducted in order to reduce sample viscosity, improve template accessibility and to enable optimal accuracy by separating tandem gene copies (Bio-Rad, 2015). Positive as well as negative controls are used to assess the reliability and validity of the procedure.

Using the QX200 Droplet generator, droplets are then generated in a water-oil emulsion forming partitions that separate the template DNA molecules and from which the genetic material can be identified and quantified. Partitioning is achieved using microfluidic circuits and surfactant chemistry, thereafter, leading to 20 - 25 thousand droplets, each containing one or no copy of the template DNA (Saumyal, D'Gama, and Walsh, 2016). A PCR reaction is then carried out within each droplet in a compatible thermal cycler.

Afterwards, samples are placed in the QX200 droplet reader, which analysis each droplet individually using a two-color detection system, enabling multiplexed analysis for different targets in the same sample. A simple threshold based on fluorescence amplitude, then determines each droplet as positive or negative. Analysis of results is then conducted by the QuantaSoft™ software (Bio-Rad, USA) which uses statistical models in order to determine the target DNA concentration in the form of copy number per μl in the sample (Bio-Rad, 2022).

Figure 4 Representative 2D scatter plot of a ddPCR result. Groups are automatically assigned into distinct areas (clusters) by the software, with the blue cluster representing the probe of interest, the green cluster the RPP30 probe, the black cluster the negative droplets and the orange clusters representing the droplets with both probes present.

Once the grouping has been determined for each sample, QuantaSoft™ will then assess how many droplets are present for each of the probes and generate a 2D interpretable graph displaying concentration (copies/µl) and copy number.
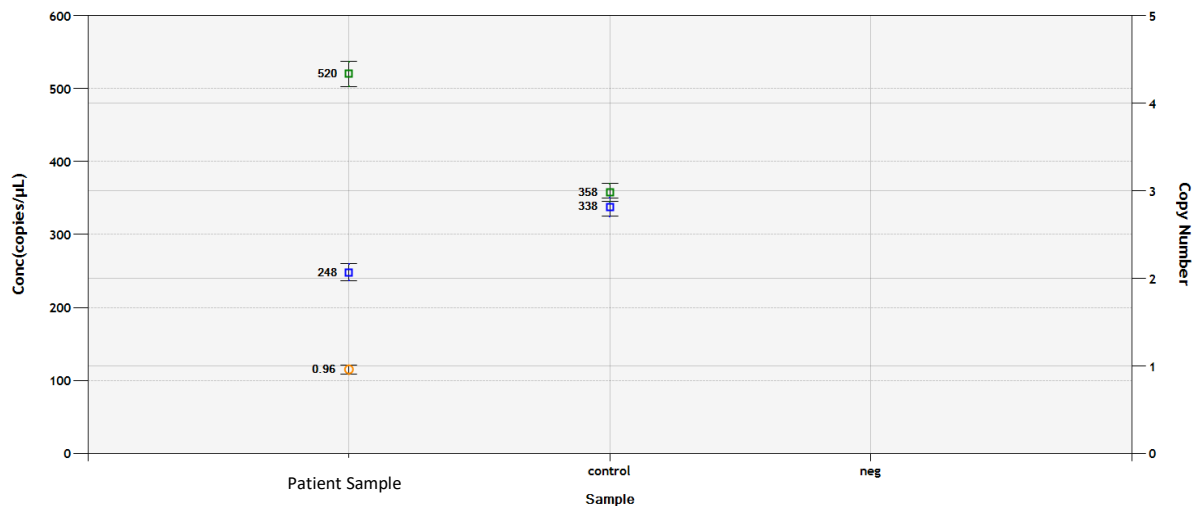
Figure 5. Representative graph demonstration concentration and copy number of the sample, and positive and negative controls. Since the probe of interest is approximately half of the value of the reference probe, this indicates that a deletion is present on the patient sample.

Analysis and interpretation of the results is based on the ratio between the probe of interest and the reference probe. A ratio between 0.40 and 0.55 represents a deletion (apart from a homozygous deletion, that would be represented by a ratio of 0); a ratio between 0.9 and 1.1 represents a wild-type sample; and a ratio above 1.4 represents a duplication. Negative and positive controls should also be in line with these values. The normal control must show comparable values (ratio between 0.9 and 1.1) for reference gene and gene of interest for the probe to be considered successful. Values outside the above ranges may either indicate technical issues or real values (e.g., mosaicism or higher ploidy of duplication).

1.5.2.3 Patient cases

Patient B, a 19-year-old male was referred for molecular genetic testing for familial aortopathy and presented with ectopia lentis and some Marfanoid features, although no apparent family history was determined. Molecular genetic analysis of a 63 gene sub-panel associated with inherited aortopathy and vasculopathy conditions as well as bioinformatic copy number analysis was conducted. Following genetic testing the patient was found to contain a heterozygous likely pathogenic duplication of exons 17 to 44 of the *FBN1* gene. This structural variant is likely to cause the production of an abnormal protein, with altered or deficient function. Confirmation of the structural variant was undertaken by ddPCR (Fig.6) using exons 17 and 44 to represent the entire duplication, furthermore parental genetic testing was also offered to determine inheritance.
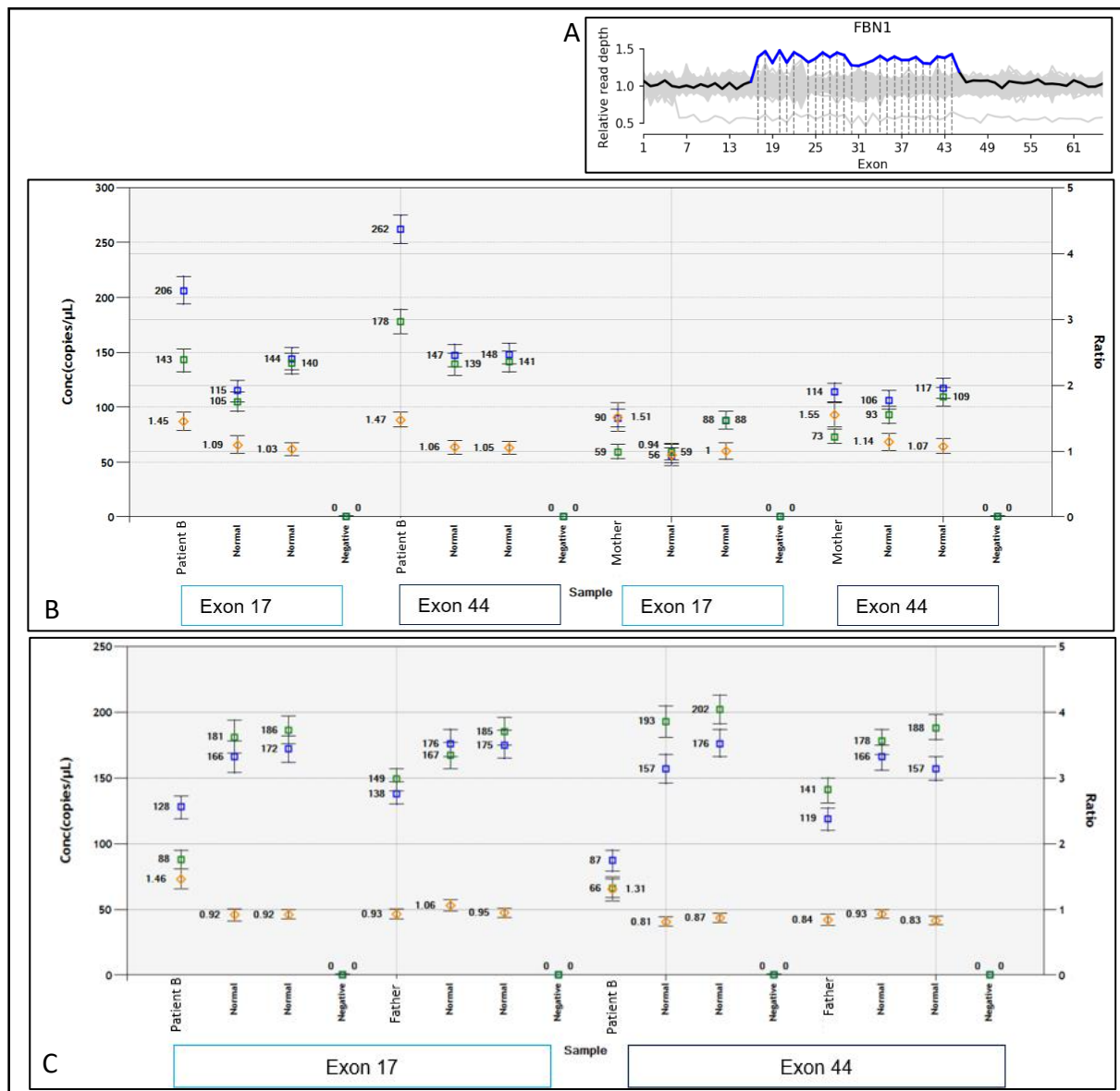
Figure 6. CNV molecular testing of patient B duplication of exons 17-44 of the FBN1 gene. A – Graph depicting relative copy number exons of FBN1 in patient B identified by bioinformatic analysis. B – ddPCR assay of the FBN1 exon 17 and 44 on patient B and patient's mother. The assay demonstrates patient's B maternal inheritance of the exon 17 and 44 duplication. C- – ddPCR assay of the FBN1 exon 17 and 44 on patient B and patient's father. Patient's father demonstrates a normal copy number ratio for exons 17 and 44 of the FBN1 gene.

1.5.3 Sanger Sequencing

Sanger sequencing (SS) is a "first-generation" sequencing method based on selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase and developed in 1977 by Dr. Frederick Sanger and colleagues (Sanger, Nicklen and Coulson, 1977). A few improvements have been made since then, which primarily involved the replacement of radiolabelling with fluorometric based detection, an improved detection through capillary-based electrophoresis and the development of software to interpret and analyse the sequences (Heather and Chain, 2016). Most of these improvements took place during the Human Genome Project (HGP) which began in 1990 and was conducted as an international effort to completely sequence the first human genome (International Human Genome Sequencing Consortium, 2001).

Sanger sequencing (Fig. 7) is a robust technique, capable of virtually detecting any variant in the targeted region and can typically achieve read lengths of up to 1 kb with relatively low cost (Al-Turkmani, Deharvengt and Lefferts, 2020). However, its high cost, labour intensiveness, low sensitivity and its low throughput represent the major limitations of this technique (Ishige, Itoga, and Matsushita, 2018). Nonetheless, after more than 40 years since it's invention and despite the implementation of next-generation sequencing, Sanger Sequencing remains one of the most common techniques used in any genetics laboratory.

At present sanger sequencing is used for the following reason in the CGGL:

- To confirm clinically actionable variants that are identified by NGS.
- For familial cascade screening and of variants reported in a family member sequenced using the NGS assy. Sanger sequencing and ddPCR are cheaper and quicker methods for checking if the family member has the variant in question.
- Sequencing for gap filling of exons of core genes that are not covered to 100% at 20x coverage. Different core genes apply to the different sub-panels used by the laboratory.
- For screening for variants in genes which are not included in any current NGS panel offered by the lab, where appropriate and small enough to be feasible.

Prior to Sanger sequencing, the region of the genome of interest must be amplified using a conventional polymerase chain reaction (PCR) protocol. After denaturation, an oligonucleotide primer is then annealed adjacent to the single-stranded sequence of interest and elongated by DNA polymerase (Deharvengt et al, 2020). Primers are generally designed to amplify the entire exon in which a variant is located. In order to ensure that there is no

contamination, and that the PCR product size is as expected, a 1% agarose gel is made. The gel can then be visualised using the Geldoc system and analysed using the Genesys software.

Following agarose gel electrophoresis, the sample is cleaned-up to remove excess primer and dNTPs, and BigDye terminator cycle sequencing is used to sequence amplicons in the forward and reverse directions separately. In general, primers used for the initial PCR are also used as sequencing primers.

Primers which anneal to the single-stranded DNA template are elongated by DNA Taq polymerase, whilst fluorescent-labelled deoxynucleotide triphosphates (dNTPs), which provide the needed adenine, cytosine, thymine, and guanine nucleotides are introduced one at a time and the primer is extended in a template-dependent manner. Fluorescent-labelled dideoxynucleotide triphosphates (ddNTPs) for each nucleotide are also included at rate-limiting concentrations (Gomes and Korf, 2018).

Since the polymerase enzyme is unable to distinguish dNTPs and ddNTPs, the deoxyribonucleotide analogs which lack the 3'- hydroxyl group begin to be randomly incorporated into the growing chain. However, when the ddNTP is incorporated into the growing strand of DNA, further chain elongation cannot occur because it becomes impossible to form a phosphodiester bond with the 5'- phosphate end with the next dNTP. This results in a mixture of truncated fragments of variable length (Belhassan and Granadillo, 2021). Subsequently, the different length products are then separated by capillary electrophoresis (using the ABI 3500 Genetic Analyzer) in accordance to their size. Terminating nucleotides are identified by a laser due to their electrophoretic mobility and colour of the chain-terminating base (Ishige, Itoga, and Matsushita, 2018). A sequence chromatogram with fluorescent peaks corresponding to the incorporation of the four different fluorescent dyes coupled to ddATP, ddCTP, ddGTP and ddTTP is then generated. Analysis software (SeqPatient) can then compare the results with reference sequences and highlight abnormalities which represent variants and polymorphisms (Franklin et al, 2014). The convention for electrophoresis colours is green for 'A', red for 'T', black for 'G' and blue for 'C' (Gomes and Korf, 2018).

Negative control reactions are performed for each primer used in a PCR reaction. This ensures that ay contamination is identified immediately, minimising the risk of false positive results. For cascade screening, a positive control from the same pedigree (usually the proband) must be included (when possible) in the assay. The variant of interest must be identified in the positive control during analysis and the sequence compared with the test samples and a reference sequence of the correct region. This ensures that the primers are

working correctly, and the correct variant is being identified, minimising the risk of a false negative result.
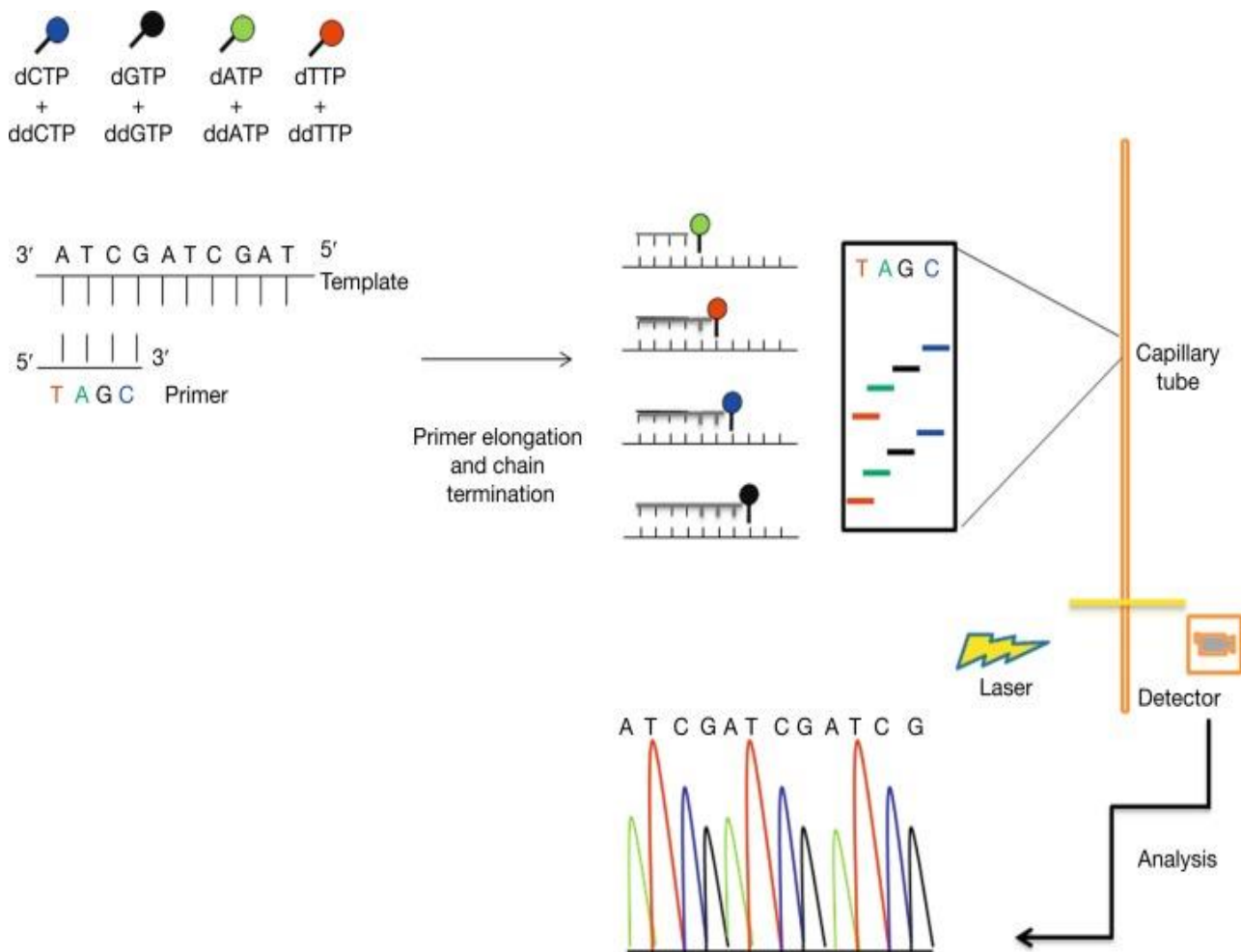
Figure 7. Schematic overview of the Sanger Sequencing method. Following amplification and elongation, each chain termination reactions is attached with a dideoxynucleotide labelled with a different fluorophore. Each fragment is then analysed by a fluorescence detector, which identifies which ddNTP is present. The information is then analysed by a software analysis program and chromatogram with different peaks and colours is generated. Adapted from: Zhang and Fernandes, 2014.

1.5.3.1 Patient cases

Patient C, a 9-year-old boy, with *situs inversus totalis*, outer dynein arm cilia defects on electron microscopy and a clinical diagnosis of PCD was found to be compound heterozygous for two variants in the *DNAH5* gene c.6261T>G p.(Tyr2087Ter) and c.11455+5G>A. The pathogenic c.6261T>G variant has been detected in 1 allele in control populations (gnomAD database) compatible with the prevalence of autosomal recessive disease by pathogenic variants in this gene. The variant is predicted to cause premature protein truncation and the mRNA produced might be targeted for nonsense mediated decay, leading to loss of function,

a known disease mechanism for *DNAH5*. The VUS c.11455+5G>A has not been detected in control populations nor in literature reports. In silico tools predict this variant will disrupt the donor splice site of exon 66 and cause premature protein truncation leading to loss of function, although this prediction has not been tested with any functional studies.

Patient C has a brother with PCD who is also compound heterozygous for the two variants. Parental testing was undertaken by forward and reverse Sanger sequencing (Fig.8) to confirm bi-allelic inheritance of the *DNAH5* variants.
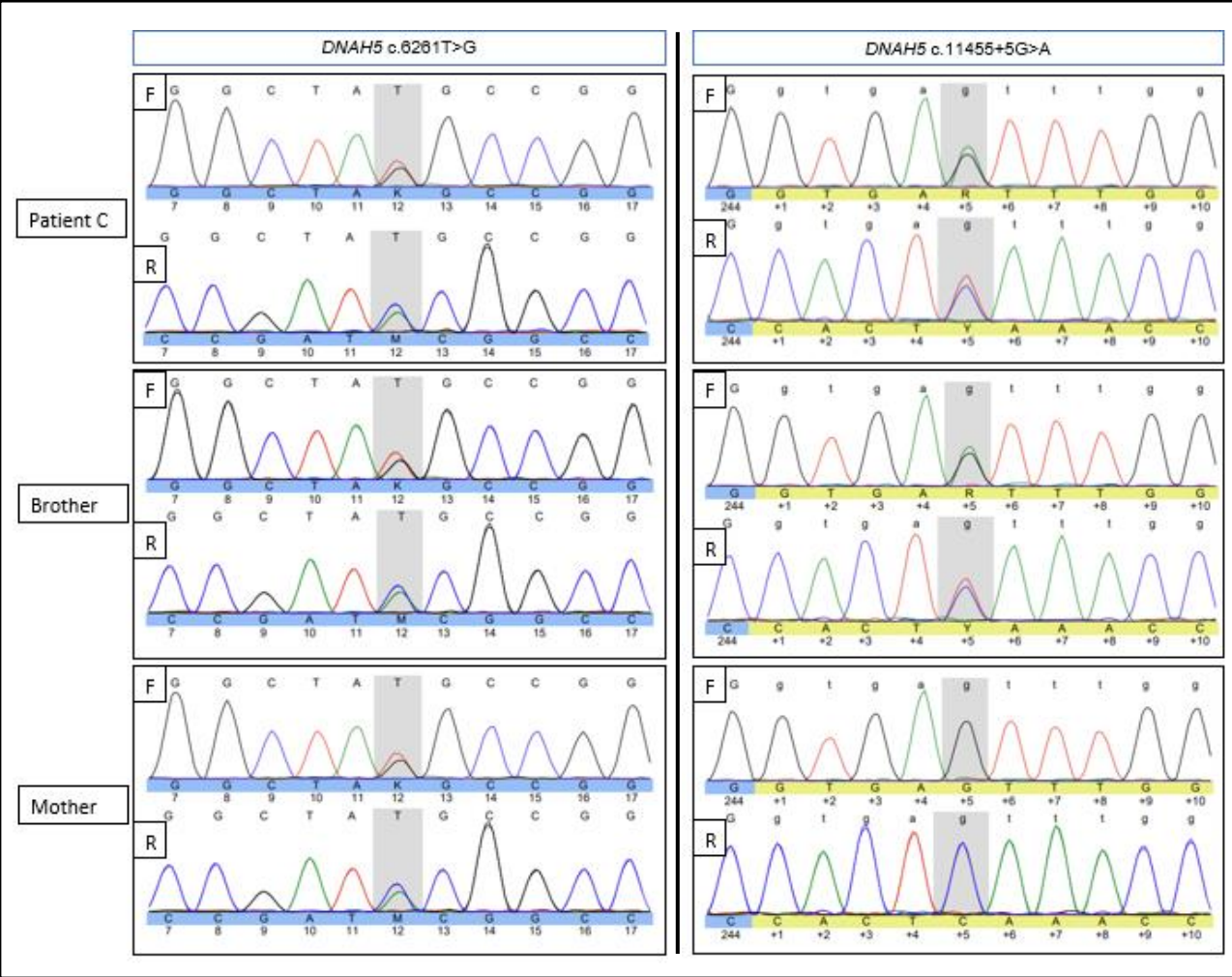
Figure 8. Cascade screening analysis (using SeqPatient) of the DNAH5 c.6261T>G and c.11455+5G>A variants. On the left side, sequencing of the reverse and forward fragments demonstrates that all 3 individuals are heterozygous for the pathogenic DNAH5 c.6261T>G variant. This corroborates the fact that the pathogenic variant was maternally inherited. On the right side, sequencing of the reverse and forward fragments demonstrates that both the patient and patient's brother are heterozygous for the VUS DNAH5 c.11455+5G>A, whereas the mother is homozygous for the wildtype variant. This demonstrates that the variant was not maternally inherited, and suggests a possible paternal inheritance, however genetic testing of the father would be required to confirm this.

1.5.4 Next Generation Sequencing

Since the completion of the HGP in 2003, substantial changes have occurred in the approach to genome sequencing. Although, SS is highly accurate and effective at sequencing a few short DNA fragments, it becomes costly and extremely time-consuming when dealing with large amounts of DNA fragments. To overcome these method limitations, new and improved technologies prompted the development of high-throughput sequencing technologies (Gupta and Verma, 2019).

Next-Generation Sequencing (NGS) is a generic term used to describe several different techniques that enable parallel multiplexed analysis of DNA sequences on a massive scale (Lindeman, Fletcher, and Longtine, 2021). In recent years, NGS has dramatically revolutionized both research and diagnostic areas. Due to its ultra-high throughput, robustness, and speed, incredible progress in the genetic diagnosis of several inherited disorders has been accomplished. Current use of NGS expands beyond the diagnosis of inherited disorders with its applications being successfully reported in the diagnosis of infectious diseases, immune disorders, non-invasive prenatal diagnosis, and therapeutic decision making for somatic cancers (Di Resta et al, 2018).

Although the implementation of NGS in the clinical setting does present with some challenges, such as the requirement of sophisticated analyses tools, complex bioinformatic interpretation and extensive data storage requirements, NGS presents several advantages when compared to single gene approaches by Sanger Sequencing (Pereira, Oliveira, and Sousa, 2020). Besides its higher sequencing capacity, shorter turnaround times and lower costs, NGS also allows clonal sequencing of single molecules, multiplexing of samples and workflow miniaturization (Grumbt et al, 2013).

NGS methods typically generate millions of short reads on the order of 50-250bp, and besides being able to detect single nucleotide variants (SNVs) and small insertions/deletions (indels) (Abel and Duncavage, 2013), NGS can also identify large structural rearrangements, balanced translocations, uniparental isodisomy, copy number variations and mosaicism. Furthermore, the technique also offers the opportunity to interrogate noncoding regions of DNA and identify important sequence variants that influence gene expression (Haworth, Savage and Lench, 2016).

NGS can be performed at different levels of complexity to include whole genome sequencing, whole exome sequencing, whole transcriptome sequencing (mRNA sequencing), and targeted sequencing of multigene panels (Nikiforova and Nikiforov, 2019). Whole genome

sequencing (WGS) provides the most comprehensive analysis of the entire genome; however, it can be a cumbersome task to decipher complex and sometimes ambiguous genetic information. Therefore, suitable alternative approaches such as whole exome sequencing (WES) or targeted sequencing may be conducted to reduce costs and data burden (Vecchio et al, 2017). WES targets approximately 3% of the whole genome and it is restricted to the coding region (Suwinski et al, 2019) whereas targeted sequencing focuses on a panel of genes or targets know to have strong associations with pathogenesis of disease and/or clinical relevance (Bewicke-Copley et al, 2019).

NGS technologies may also be implemented for resequencing the genome of an individual for which there is a reference genome, and this may discern some knowledge on the relationship between sequence variation and normal or disease phenotypes (Stratton, 2008). Moreover, NGS can also be implemented for *de novo* assembly sequencing which may contribute to a better understanding at the genomic level and may even assist at predicting genes, protein coding regions, and pathways. Additionally, NGS technologies have also been used to analyse small RNAs, including identification of differentially expressed micro RNAs (miRNAs), prediction of novel miRNAs, and annotation of other small non-coding RNAs (Lee et al, 2013).

Different approaches to NGS exist, and although several common features are shared, these will in turn affect sequencing quality, quantity, and choice of application (Kanzi et al, 2020). A NGS run typically involves DNA extraction; DNA fragmentation; library preparation; massive parallel sequencing; bioinformatics analysis and variant annotation and interpretation (Qin, 2019).

At present, the CGGL uses the NextSeq 550 system by Illumina which utilizes a sequencing be synthesis technology. Following DNA extraction, the DNA is fragmented either by physical or enzymatic methods, and uniform adapter sequences are ligated covalently to the ends of each fragment by a DNA ligase. These adapters contain unique sequences (index tags / "sample barcodes") that are used to identify each individual sample, which allows the DNA from multiple samples to be pooled and analysed concurrently. The oligonucleotide adapters also contain universal primer recognition sequences and can be used to polymerase-amplify the library fragments during specific steps of the process (Lindeman, Fletcher, and Longtine, 2021). Amplification is required to provide sufficient signal from each of the DNA sequencing reaction steps (Mardis, 2013). Moreover, the adapters contain a sequence that is complementary to a second sequence located on the inside surface of a flow cell lane. This allows bridge amplification to take place and generates a double-strand bridge (Pereira, Oliveira, and Sousa, 2020). After, denaturation occurs and the initial library molecules are

removed and the copied, flow cell-attached fragments are used to generate a cluster of identical template molecules using isothermal amplification. This process which is continuously repeated through cyclic alternations of three specific buffers that mediate the denaturation, annealing and extension steps results in the generation of several million dense clusters of DNA in each channel of the flow cell (Buermans and Dunnen, 2014).
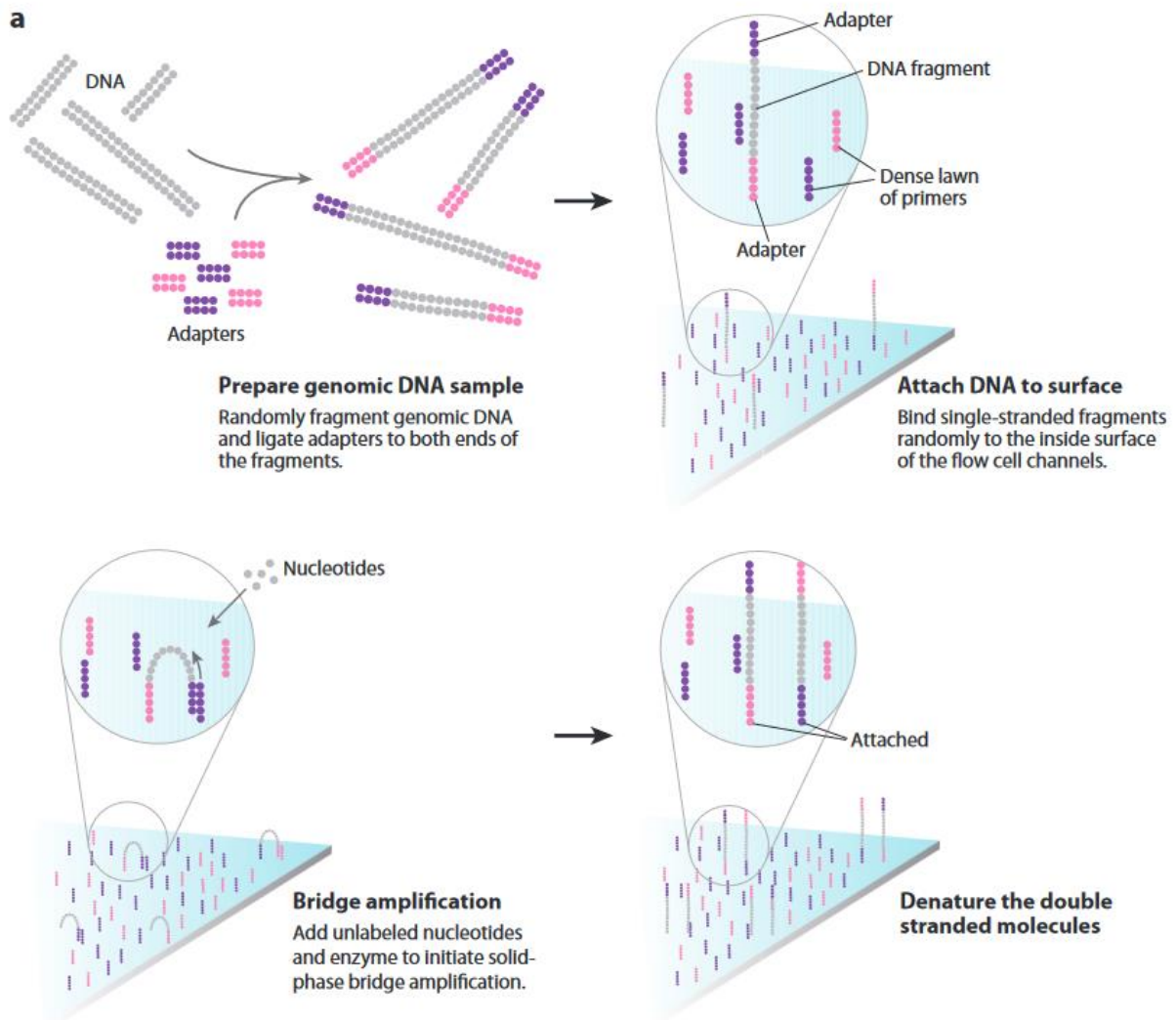
Figure 9. The Illumina initial sequencing-by-synthesis approach. Following DNA fragmentation, adapters, sequencing primer binding sites, index sequences, and regions complementary to the flow cell oligonucleotides are added to the fragments. To increase the signal for sequencing, the short DNA fragments are amplified through a process called bridge amplification. Adapted from Mardis, 2008.

For sequencing, clusters are denatured, and a subsequent chemical cleavage reaction and wash leave only forward strands for single-end sequencing. Sequencing of the forward strands is initiated by hybridizing a primer complementary to the adapter sequences, followed by addition of polymerase and a mixture of four differently colored fluorescent reversible dye terminators (Castiblanco, 2013). During each cycle, a mixture of the individually labelled and 3'-blocked dNTPs are added. After the incorporation of a single dNTP to each elongation complementary strand, unbound dNTPs are washed, and the surface is imaged to identify

which dNTP was incorporated at each cluster. After imaging the fluorescent group is cleaved off and the blocking group can then be enzymatically washed, and a new cycle is performed. The terminators are incorporated according to sequence complementarity in each strand in a clonal cluster. After incorporation, excess reagents are washed away, the cluster are optically interrogated, and the fluorescence is recorded. With successive chemical steps, the reversible dye terminator is unblocked, the fluorescent labels are cleaved and washed away, and the next sequencing cycle is performed (Goodwin, McPherson and McCombie, 2016).
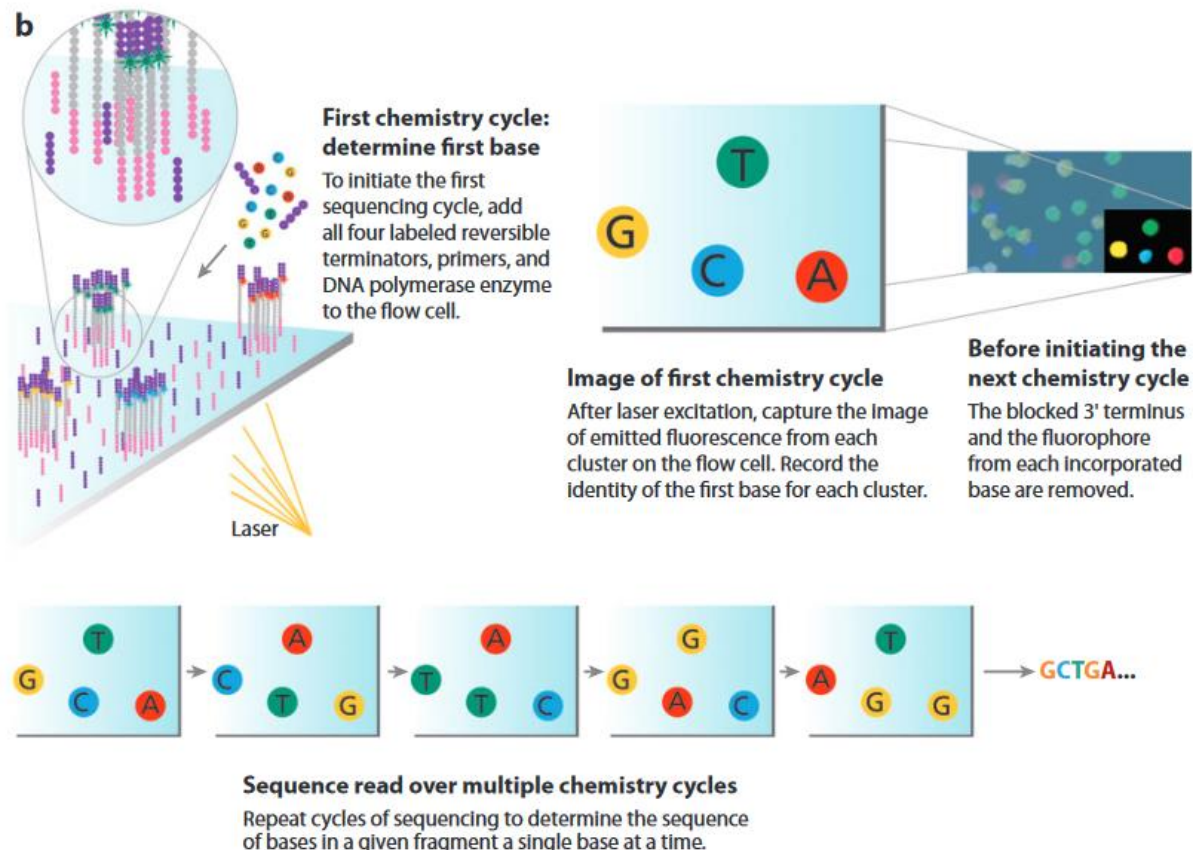
Figure 10. The Illumina final sequencing-by-synthesis approach. For each cluster of DNA on the flow cell, a supply of sequencing primers, fluorescently-labeled dNTPs, and DNA polymerase are added. After the first complementary nucleotide is added to the end of the primer, the fluorophore blocks DNA polymerase from adding any more nucleotides. Afterwards, the reagents are washed from the cell flow and the cycle is repeated. Adapted from Mardis, 2008.

Following sequencing, the large amounts of data generated must be analysed by multiple computationally intensive steps. NGS data is processed by a NGS bioinformatics pipeline which consists of the following major steps: sequence generation, sequence alignment, variant calling, variant filtering, variant annotation, and variant prioritization (Roy et al, 2018).

Variant interpretation and assessment are then conducted following the guidelines produced by the American College of Medical Genetics and the UK association of Clinical Genomic Science, as well as up-to-date literature on the assessment of variants in specific genes.

PART II

Primary Ciliary Dyskinesia

Whole gene sequencing identifies copy number variations and intronic variants with potential functional impact in patients with PCD

2.1 Introduction

Primary ciliary dyskinesia (PCD) (MIM no.244400) is a clinically and genetically heterogenous disorder resulting from ciliary dysfunction, predominantly inherited in an autosomal-recessive way, however rare cases of autosomal-dominant and X-linked recessive inheritance patterns have also been observed (Narayan et al, 1994; Horani et al, 2015). PCD is clinically characterised by upper and lower airway disease, infertility caused by ciliary (or flagellar) dysfunction and organ laterality defects (Knowles, Zairwala and Leigh,2016; Kuehni and Lucas, 2017).

First described in 1933 by Kartagener, as a syndrome based on the triad of chronic sinusitis, bronchiectasis, and situs inversus (Kartagener, 1933), PCD was then acknowledged by Afzelius, forty years later, to be the result of ultrastructural defects of cilia leading to ciliary immotility or abnormal ciliary motility (Afzelius, 1976). Nowadays, it is known that clinical disease may also result from functional ciliary impairment without ultrastructural deformities, as well as motile cilia with obvious uncoordinated and ineffective movement patterns (Lobo, Zairwala and Noone, 2015).

The prevalence of PCD is estimated to be approximately 1 in 10000 people, although its true prevalence may be higher due to limitations in diagnostic methods that focus on testing ciliary ultrastructure and function, as well as an often under-recognition of the syndrome (Knowles et al, 2013). Furthermore, a higher incidence of PCD cases can be diagnosed amongst certain ethnic groups with elevated rates of consanguinity, such as, the British Asian population, the Volendam population in the Netherlands and the Amish and Mennonite communities in the USA (Damseh et al, 2017).

Diagnosing PCD can be challenging and relies on a combined approach of clinical data, cilia ultrastructure and function analysis, nasal nitric oxide levels and genetic testing (Mirra, Werner and Santamaria, 2017). At present, pathogenic variants in approximately 50 genes are known to cause primary ciliary dyskinesia. Causative mutations leading to PCD include genes encoding for axonemal, cytoplasmic and regulatory proteins involved in ciliary biogenesis, ciliary assembly and preassembly, as well as structure, regulation and function of motile cilia (Horani, Brody and Ferkol, 2014; Zhao et al, 2021). Advances in molecular genetics and genomics have improved immensely in the last few years, with a genetic cause now being able to be identified in 70% of cases (Wheway et al, 2021). Nowadays, there are no PCD therapies validated, and treatment has not yet been standardized.

2.2 Function and structure of cilia

Cilia are complex and highly conserved organelles, in both protein composition and structural organization, that protrude from the surface of almost every cell present on the human body (Wirschell et al, 2011). These dynamic structures play an essential role in several cellular functions and normal development processes and can be classified as either motile or primary cilia (Ostrowski, Dutcher and Lo, 2011).

Motile cilia are hair-like structures, with a highly organized internal microtubule-based structure (axoneme), that protrude out from the apical surface of epithelial cells lining the respiratory tract, the ventricles of the brain, the Fallopian tubes, sperm flagellum, the Eustachian tube of the ear and the nodal cells in the early embryo (Hyland and Brody, 2022). The axoneme (Fig.11) is comprised of nine peripheral microtubule doublets surrounding a central pair of microtubule singlet and hence this ultra-structure is referred to as "9+2" (Popatia, Haver and Casey, 2014). Attached to these microtubules there a variety of other structures, such as, the outer and inner dynein arms (ODAs and IDAs, respectively) whose interaction can convert chemical energy stored within ATP into mechanical work, and thus creating motion through the sliding of the peripheral microtubular doublets (Leigh et al, 2013). Additional key components are nexin-dynein regulatory complexes (N-DRC) which connect neighboring doublets, act to stabilize the axonemal core structure and regulate dynein motor proteins (ODA+IDA), thereby controlling ciliary motility (Gui et al, 2019; Lee and Ostrowski, 2021) and radial spoke complexes which are simultaneously connected to the outer doublet microtubules and the mobile central pair apparatus, this way providing a mechanochemical signal transduction capable of regulating ciliary beating and waveform(Smith and Yang, 2004; Mitchinson and Valente, 2017). These two mechano-regulatory complexes are essential for a regular beating pattern and work as a mechanical feedback loop that spatially and temporarily inhibit the dynein motor proteins (Gui et al, 2021).
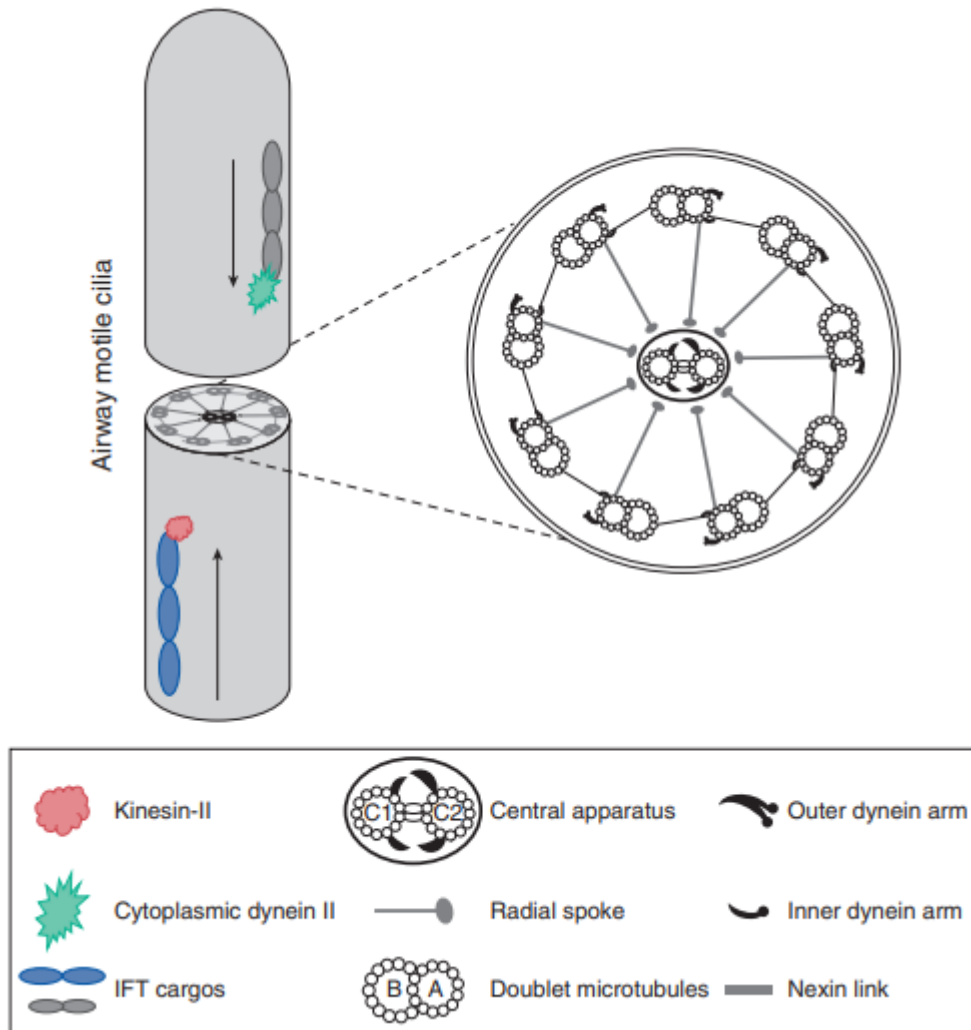
Figure 11. Schematic representation of the normal axoneme structure of the motile cilia adapted from Mianné et al, 2018.

Each airway epithelial cell has approximately 200 motile cilia projecting from its surface, which are oriented in the same direction and move in unison at a constant rate that ranges between 8-20Hz, with the aim of clearing bacteria and toxic substances from the airways (Mirra, Werner and Santamaria, 2017). Factors such as the depth and viscosity of the apical surface fluid, triphosphate nucleotides, changes in redox conditions and exposure to various pathogens and airborne pollutants can lead to changes in ciliary microenvironment and consequently alter the regulation of motile cilia beating pattern (Horani, Brody and Ferkol, 2014; Leigh et al, 2019).

During early embryonic development a subtype of motile cilia, the nodal cilia, are present on the embryonic nodal plate cells. These specialized motile cilia have 9 peripheral doublets and dynein arms; however, they lack the central pair of microtubules. Functionally, this leads to a rotatory motion that generates left-ward movement of extracellular fluid across the surface of the embryonic node which is essential for the development of organ laterality during embryogenesis. In the absence of normal nodal ciliary function, organ placement becomes random and may therefore lead to laterality defects such as situs inversus totalis, situs

ambiguous and heterotaxis syndromes (Damseh et al, 2017; Leigh et al, 2019). This randomness of laterality in PCD explains why approximately 50% of patients have situs inversus, even within genetically identical twins (Noone et al, 1999).

Another type of cilia present in almost all types of cells during interphase is the primary or sensory cilia, these are solitary and non-motile organelles which display a 9+0 microtubule configuration (Kim and Dynlacht, 2013). Primary cilia display an important role in multiple signaling pathways and play a pivotal part in sensing the cellular microenvironment and transducing them into decisions regarding proliferation, migration, differentiation, polarity, nerve growth and tissue maintenance (Schoentgen, 2020). Furthermore, primary cilia can also act as photo-, mechano-, osmo-, thermo-, hormone and olfactory receptors (Hildebrandt, Benzing and Katsanis, 2013).

2.3   Clinical features of PCD

PCD clinical presentation is heterogenous and includes a broad range of symptoms that are commonly found in other disorders, which often complicates and delays the diagnosis. Clinical symptoms of PCD affect the entire respiratory tract with most symptoms starting soon after birth. At least 80% of newborns with PCD develop neonatal respiratory distress despite a full-term gestation, with unexplained respiratory distress manifesting as tachypnoea, prolonged oxygen requirement and prevalence of upper and middle lobe atelectasis on chest radiographs (Mullowney et al, 2014).

In healthy individuals, a coordinated whip-like cilia beating pattern allows the occurrence of mucociliary clearance from the lungs, paranasal sinuses and middle ears, which is an important primary innate defense mechanism against bacteria and debris (Lucas et al, 2014). Due to impaired cilia motility and insufficient mucociliary clearance this results in pathogens being trapped in layers of mucus which cannot be adequately cleared, leading over time to a persistent wet cough, recurrent chest infections, perennial rhinosinusitis and eventually bronchiectasis (Lucas et al, 2020). The dysfunctional mucociliary clearance in the Eustachian tube and middle ear cleft results in several early onset otologic features (Prulière-Escabasse, Coste and Chauvin, 2010). Most children with PCD suffer from binaural chronic or recurrent otitis media with effusion (OME), which often persists until adulthood. The degree of severity of OME varies, however mild to moderate conductive hearing loss is reported in approximately 75% of the children with PCD, and so this may lead to possible delays in speech development (Kreicher et al,2018).

A spectrum of organ laterality defects can occur within PCD patients. Without functional nodal cilia, left-right body asymmetry becomes random and approximately half of the PCD sufferers displaying *situs inversus totalis* with reversal of the thoracic and abdominal organs. More rarely, *situs ambiguus* has also been described in 6.3% of PCD patients and is sometimes associated with congenital heart disease, asplenia and polysplenia (Kennedy et al, 2007; Shapiro et al, 2015; Fretzayas and Moustaki, 2016).

Male subfertility or infertility are also common and are additional non-respiratory complications. Since sperm flagella are structurally similar to respiratory cilia and are powered by dynein arm heavy chains, impaired motility in sperm cells as well as dysmotility in the efferent duct of the male reproductive tract leads to fertility problems in roughly 75% of males (Vanaken et al, 2017; Jayasena and Sironen, 2021). Females with PCD may also experience fertility problems and/or ectopic pregnancies due to immotile fallopian tube cilia (Behan et al, 2017).

Furthermore, cilia dysfunction is also involved in polycystic hepatic and renal disease, skeletal disorders, psychomotor developmental delay, diabetes, biliary atresia, retinal degeneration and rare syndromes (Bush et al, 2007). Hydrocephalus has also been demonstrated to be associated with PCD; although it is believed that impaired cerebrospinal fluid flow secondary to dysfunctional motor cilia that line the ventricular ependymal cells may be the cause of it, the full cause and mechanism is not yet fully understood (Greenstone et al, 1984; Sakamoto et al, 2021).

2.4 Diagnostic approaches

The diagnosis of PCD involves a complex, time-consuming and expensive diagnostic process. There have been several advances in our knowledge of genetics and pathogenesis of the disease in last few years, which have contributed tremendously to an increase in quality and innovation of diagnostic techniques. Furthermore, the clinical features of PCD are nowadays well recognised, yet these can often be very broad and overlap with symptoms of other disorders, which can delay and complicate an accurate diagnosis. The European Respiratory Society and the American Thoracic Society have published guidelines in order to improve diagnosis of PCD. Nowadays, there is no single "gold standard" diagnostic test and therefore there must be a combination of investigations, including clinical phenotype, nasal nitric oxide, high speed video-microscopy analysis (HSVA), transmission electron microscopy (TEM), genetic analysis and immunofluorescence (IF) staining of ciliary proteins in order to achieve a PCD diagnosis (Lucas et al, 2017; Shapiro et al, 2018).

Despite the presence of typical symptoms early in life, PCD is often diagnosed at a mean age roughly around 5 years of age. However, advances in diagnostic techniques and perhaps an increased awareness of this rare disease, appear to have decreased the time taken to achieve diagnosis to a median age of 2.6 years (Rubbo et al, 2020). This is an extremely important improvement since, it is thought that an early diagnosis can reduce long-term pulmonary morbidity and prevent unnecessary investigations and unhelpful treatments (Kuehni et al, 2010).

2.4.1 Nasal nitric oxide

Nitric oxide (NO) is continuously synthesised in the respiratory epithelium and is upregulated in response to infection or inflammation. Nitric oxide is a highly reactive gaseous molecule with numerous signaling roles within the airways, such as vascular homeostasis, immune cell activity and bronchomotor tone. Nitric oxide is upregulated during infection, via increased inducible nitric oxide synthase transcription and activity. Despite recurrent respiratory infection, nasal concentrations of nitric oxide are markedly reduced in the vast majority of PCD patients, compared with those without the disorder. Nasal NO (nNO) can be measured using simple non-invasive techniques via the reaction of NO with ozone and subsequent detection by chemiluminescence through commercially available equipment (Walker et al, 2012; Walker et al, 2013).

The use of nNO for PCD diagnosis is a widely common test across PCD specialised centres, since levels of nNO are extremely low in several patients with PCD (10-15% of healthy individuals) (Leigh, O'Callaghan and Knowles, 2011). Measurement of nNO has also the advantage of yielding immediate results and it is less invasive and costly when compared to TEM and HSVA.

A systematic literature review conducted by Collins et al in 2014, demonstrated nNO cut-off values that varied between 25-126 nL-min^{-1} presented with a sensitivity of 75-100% and specificity of 88-100%. At present, there are still no standardised disease specific cut-off values mostly because levels of NO vary significantly with age (Walker et al, 2012). Young children have lower nasal NO values than older children and adults, therefore age-specific cut-off values are needed when using nNO as a screening test for PCD. Additionally, young children are often unable to comply with the protocol requirements and thereafter measurement of nNO is usually only performed on children >5 years old (Jackson et al, 2016).

Low levels of nNO are not an exclusive feature of PCD and do not warrant an immediate diagnosis. Since low levels of nNO may occur in other disorders such as Cystic Fibrosis, nasal

blockage, nasal polyposis, panbronchiolitis or even in smokers, confirmation of diagnosis is necessary to be undertaken using other techniques (Barbato et al, 2009). Moreover, it is also important to mention that a normal result does not exclude the diagnosis of PCD, since certain patients present normal levels of nNO. Normal levels are associated with some PCD caused genes such as *RSPH1, GAS8, RPGR, CCNO, CCDC103, CFAP221, DNAH9, FOXJ1, GAS2L2, LRRC56, NEK10, SPEF2, STK36 and TTC12* (Shapiro et al, 2020).

2.4.2 Ciliary beat pattern and frequency - HSVA

Evaluation of ciliary beat pattern (CBP) and ciliary beat frequency (CBF) can be assessed by using high-speed video microscopy analysis. HSVA is a subjective and qualitative technique in which the respiratory cilia are visualized ex-vivo with a light microscope and recorded with a high-speed video camera (Rubbo et al, 2019). Prior to HSVA, a good quality epithelial sample must be obtained from the upper or lower airways by nasal brushing.

Pattern abnormalities such as static, slow, hyperkinetic and stiff cilia can be detect by HSVA, however, this assessment must be conducted by experienced and skilled specialists. Moreover, an analysis of HSVA can be extremely challenging due to lack of standardization; the existence of ciliary defects secondary to infection or inflammation and the existence of PCD-associated genes (*HYDIN, CCDC164, DNAH9, GAS8, CCNO and MCIDAS*) with normal or very subtle abnormalities in HSVA (Raidt et al, 2014; Shapiro et al, 2020).

2.4.3 Ciliary structural analysis – TEM

Transmission electron microscopy (TEM) analysis of cilia is an arduous and expensive technique performed in order to assess and quantify ultrastructure abnormalities of motile cilia. Ultrastructural defects are classified into two main classes, according to the International Consensus guideline for reporting transmission electron microscopy results in the diagnosis of PCD (Shoemark et al, 2020).

Class 1 defects – Hallmark diagnostic defects:

- ODA defects
- Outer and inner dynein arm defect
- Microtubular disorganization and inner dynein arm defect

Class 2 defects - Indicate a PCD diagnosis with other supporting evidence:

- Central complex defect
- Mislocalisation of basal bodies with few or no cilia

- Microtubular disorganization defect with inner dynein arm present
- Outer dynein arm absence from 25-50% cross sections
- Combined inner and outer dynein arm absence from 25-50% cross sections

Although TEM can be considered highly specific, it must be used as part of a wider set of investigations. Since TEM is normal in ~30% of patients with PCD, lack of ciliary defects does not exclude the diagnosis of PCD. Defects of nexin link components, central pair components, ciliary biogenesis defects and defects cause by *DNAH11* cannot be visualized by TEM (Lucas et al, 2016). Furthermore, false positives may also occur due to inflammation, infection, exposure to airborne pollutants or poor sample handling (Dixon and Shoemark, 2017).

2.4.4 Immunofluorescent antibody staining of ciliary proteins

IF microscopy analysis is a technique consisting of labelling ciliary proteins in human respiratory cells with fluorescent tags and indirectly examining the presence and localization of specific ciliary proteins along the cilia by fluorescent or confocal microscopy (Baz-Redón et al, 2020). IF was firstly used as a research tool in 2005 by Fliegauf *et al*, in order to improve understanding of the impact of disease-causing genes on ciliary proteins. Nowadays, an important number of antibodies against different cilia proteins defective in PCD have been developed and validated, including *DNAH5* (an outer dynein arm heavy chain), *DNAL1* (an inner dynein arm light chain), *GAS8* (a component of the nexin-dynein regulatory complex) and *RSPH4A, RSPH9 and RSPH1* (components of the radial spoke). Recently, it has been shown that PCD variants with normal ultrastructure (e.g. caused by *HYDIN, DNAH11 and SPEF2* mutations) can be diagnosed by IF (Shoemark et al, 2017).

IF is cheaper, faster, and easier than other techniques such as TEM, and its use as a diagnostic test in PCD is likely to increase as more antibodies become available (Hogg and Bush, 2021). IF has also its limitations – it does not provide antibodies to all defective proteins and the technique may fail due to the absence of cilia or interference of mucus or blood in the sample (Baz-Redón et al, 2019).

2.4.5 PICADAR

Recently, a simple and quick diagnostic prediction tool for PCD has been developed: PICADAR (Primary CiliARy DyskinesiA Rule). Although it demonstrates good accuracy, further validation is still required in different settings and populations. A PICADAR score is obtained based on seven clinical parameters including: congenital cardiac defects, situs inversus, ear symptoms, chronic rhinitis, neonatal intensive care admittance, neonatal chest symptoms and full-term gestation. Development of PICADAR aims to provide a practical clinical diagnostic

tool to identify patients requiring testing and be of extreme usefulness particularly in countries with limited resources (Behan et al, 2016).

2.5 Genetics of PCD

Primary ciliary dyskinesia is a rare, mainly respiratory, ciliopathy that is characterized by variable disease progression, phenotypic heterogeneity, and different age of onset, yet it does not have an apparent racial nor gender predilection (Horani and Ferkol, 2018). PCD is primarily inherited in an autosomal recessive fashion although hemizygous X-linked and heterozygous *de novo* dominant inheritance patterns have also been reported (Bhatt and Hogg, 2020). Most recently, a study conducted in 2016 by Li *et al*, suggested that inheritance pattern and genetic causes of PCD may also be linked to trans-heterozygous pathogenic variants.

Although significant progress has been achieved in the last few years, there is still much to unveil regarding the genes and mechanisms linked to PCD. Scattered throughout the human genome, there are over 250 potential motile cilia genes, each encoding a protein present within a single cilium. Thereafter, and given the number of genes encoding proteins for cilia structure and function it can be anticipated that the number of PCD-causative genes will continue to increase (Gravesande and Omran, 2005).

To date, mutations in at least 50 genes have been identified as pathogenic in PCD. It is estimated that the reported genetic mutations in known genes, account for only 70% of the PCD cases. Most of the genes associated with PCD encode proteins that are involved in axonemal motors, structure and regulation or ciliary assembly and preassembly (Horani and Ferkol, 2021). Most of the variants (~80%) with a high pathogenic impact are loss-of- function variants (nonsense, frame shift or defective splice variants), whereas the remaining variants accounting for PCD are conservative missense variants or in-frame deletions and are much less reported (Lobo, Zairwala and Knowles, 2015). Additionally, most of the mutations are private and clustering of mutations in specific genetic regions is less frequent than in other genetic disorders (Werner, Onnebrink and Omran, 2015). Another aspect worthy of mention is that, although dozens of genes have been described as causative in PCD, nearly 50% of the genetically solved cases in Europe and North America have been accounted to occur due to variants in *DNAH5, DNAH11, CCDC39 and CCDC40* (Lucas et al, 2020).

The identification of PCD genes has relied on a combination of methodologies such as functional candidate gene testing, homozygosity mapping followed by positional candidate gene analysis and comparative computational analysis involving comparative genomics,

transcriptomics, and proteomics (Leigh et al, 2009). Linkage analysis and homozygosity mapping have been used to indicate the regions of interest in the genome of PCD families and cohorts (Kurkowiak, Ziętkiewicz and Witt, 2015), whilst proteomic profiling studies have sought to define the components of motile cilia by dysregulating such transcriptional regulators and analysing the proteome of isolated cilia preparations using mass spectrometry (Patir et al, 2020).

Until 2008, the identification of the first genes associated with PCD relied tremendously on the use of model organisms such as *Chlamydomonas reinhardtii* and Zebrafish. By taking advantage of the evolutionary conservation of genes and the homology between the axonemal structure of these model animals and the human respiratory cilia and sperm tails, several candidate genes with roles in structural elements of the ciliary axoneme and preassembly of dynein arms, have been identified due to this approach (Omran et al, 2000; Horani et al, 2012, Kristof et al, 2017).

Since 1999, when *DNAI1* was the first PCD – causing gene to be identified based (Fig.12) on a candidate gene approach, remarkable efforts have been made to define the genetic causes of PCD (Guo et al, 2017). With the advent of Next Generation Sequencing (NGS) in the last decade, the discovery of PCD-associated genes has been accelerated, with over 50 genes now identified and more awaiting validation. Traditionally genetic testing of a single gene candidate was conducted by Sanger Sequencing (SS), however due to the extensive heterogeneity of PCD and the large size of many of the causative genes, SS has become costly and time-consuming (Djakow et al, 2016). Therefore, NGS approaches have become extremely attractive not only for the use of research of novel genes, but also for the use at a clinical setting where it is expected to increase the accuracy and robustness of PCD diagnosis (Lucas and Leigh, 2014; Boaretto et al, 2016).
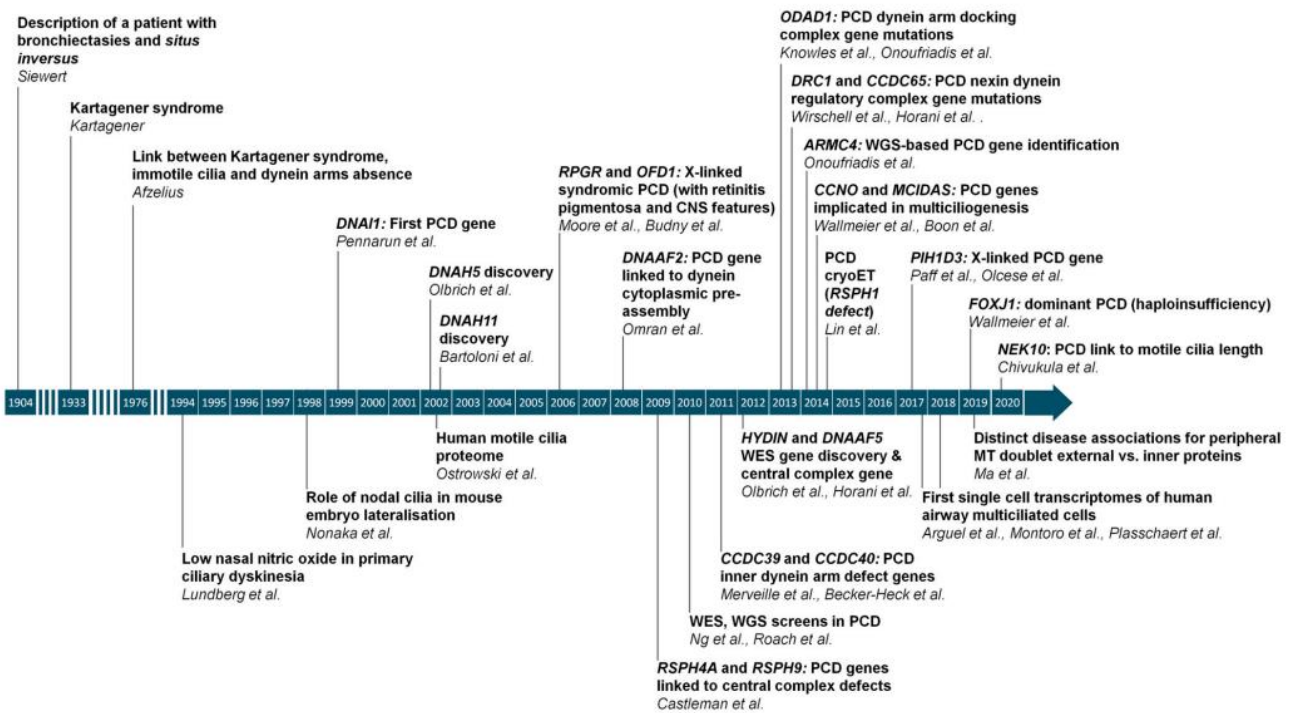
Figure 12. Timeline of PCD clinical description, respiratory and ciliary investigations, and significant steps in the discovery of its molecular causes. Adapted from Legendre, Zaragosi and Mitchinson, 2021

Nowadays, molecular genetic analysis is emerging as an accessible, accurate and standardized diagnostic test and is being embraced as a first-line diagnostic tool by several specialized PCD diagnostic centres since no prior knowledge of candidate genes or pathways is required. Several NGS approaches are available to capture regions of interest, including known and potential candidate genes for PCD, the ciliome, the exome or the entire genome (Zariwala, Omran and Ferkol, 2011).

Currently, there are several commercially available gene panels that provide coverage of most known genes associated with PCD. Even though targeted sequencing is much preferred in the diagnostic setting when compared to Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS) since it is more affordable, presents with faster turnaround times and yields higher coverage of the targeted regions, it also presents its limitations (Xuan et al, 2013). Because of the continuation of gene discovery, the use of gene panels is hampered due to disadvantages such as limited gene-detection number, unknown genes that are absent from the panel, technical issues affecting coverage depth, bioinformatic challenges to identify CNVs and the difficulty to identify pathogenic variants in a highly conserved region of *HYDIN* due to a nonfunctional duplication, the pseudogene *HYDIN2* (Fassad et al, 2020; Dutcher and Brody, 2020; Zhao et al, 2021).

The use of WES is an efficient, practical, and increasingly economical genetic analysis method to analyse the entire coding part of the human genome (~3%) and a way to overcome some of the limitations imposed by targeted sequencing (Gileles-Hillel et al, 2020). WES has

emerged as a more practical and cost-effective alternative to WGS, since more than 80% of all variants reported in ClinVar, and more than 89% of variants reported to be pathogenic come from the protein-coding part of the genome (Barbitoff et al, 2020). However, with the use of techniques such as WGS and WES, new challenges arise. Due to the large volume of genetic data generated the need of sophisticated bioinformatic algorithms emerges (Collins, Walker and Lucas, 2014). Moreover, interpretation of the results may be challenging, and expertise is necessary to distinguish pathogenic variants from rare polymorphisms. Lastly, it should be highlighted that some of the limitations of WES include incomplete coverage which may result in false negatives, as well as the current inability to reliably assess certain disease mechanisms such as variation in repetitive elements and structural variants, which can represent a significant proportion of PCD pathogenic variants (Boycott et al, 2015; Marshall et al, 2015).

PCD diagnosis requires that two pathogenic variants must be found in a single known PCD gene on two opposite chromosomes or on one allele in the case of X-linked or autosomal dominant forms (O'Connor, Horani and Shapiro, 2021). However, nearly one-third of patients remain without a genetic diagnosis, either because their results are negative, only one variant has been identified or their results demonstrated variants of unknown significance (VUS). Furthermore, an accurate and reliable genetic diagnosis may be missed due to interpretive limitations of missense, inframe and close to the splice-site intronic variants (Rosenfeld, Ostrowski and Zariwala, 2018). It is estimated that additional genes and synonymous, missense changes, non-coding, or structural variants (SV) in known genes are presumed to account for the missing heritability. Identification of non-coding and SVs in known PCD genes can be conducted by WGS which may otherwise be missed by WES or standard gene panels (Wheway et al, 2021).

Whole exome sequencing has revolutionised the clinical diagnosis of patients and families with rare genetic disorders. Yet, for a substantial proportion of patients, sequence information restricted to exons and exon-intron boundaries fails to identify a genetic cause for the disease (Vaz-Drago, Custódio and Carmo-Fonseca, 2017). Noncoding variants play a significant role in gene regulation and have been suggested to account for a significant burden of causal mutations in rare genetic diseases, after unbiased genome wide association studies (GWAS) showed that more than 90% of disease-associated single nucleotide polymorphisms (SNPs) are located outside of the coding genes (Jaganathan et al, 2019). Indeed, penetrant noncoding variants that disrupt the normal pattern of mRNA splicing contribute to at least 15% of disease-causing mutations and in certain genes, up to 50% of all described variants in that gene (Qian et al, 2021).

Recently, a pre-print by Ellingford *et al* described a non-coding pathogenic variant in *DNAH11* as a cause of PCD after conducting WGS. Precise splicing of mRNA, essential for appropriate protein translation, is dependent on the recognition of the exon-intron boundaries and regulatory sequences (promoters, insulators, and enhancers) by the splicing machinery (Abramowicz and Gos, 2018). Mutations in the regulatory sequences can lead to aberrant splicing, resulting in exon skipping, alternative 5'/3'splicing and intron retention, which in turn can lead to abolished or altered and often nonfunctional proteins. Deep intronic variants can create cryptic donor or acceptor sites, which result in the inclusion of nonfunctional intronic fragments – pseudoexons. Functionally, these variants can result in a frameshift or the introduction of premature termination codon and subsequently lead to nonsense mediated mRNA decay (NMD) or a truncated protein (Reeskamp et al, 2018; Fitzgerald et al, 2020; Li et al, 2021).

2.5.1 Genotype- phenotype correlations

Whilst many efforts have been made to establish a clear relationship between clinical features and genotype, much remains uncertain because not only can different variants in the same PCD-causative gene result in disease phenotypic variability, but also the same variant in two different patients can be expressed with phenotypic differences (Eksi et al, 2022). However, a few genotype-phenotype correlations particularly with regards to variability in severity and progression of lung disease are starting to emerge (Leigh et al, 2019).

Studies have shown that children who have biallelic variants in *CCDC39 and CCDC40*, which are characterized by IDA defects with microtubular disorganization, typically have worse pulmonary function by both functional and structural assessment and are associated with earlier presentation and poorer nutritional status, when compared to similar aged children with ODA defects (Davis et al, 2015; Davis et al, 2018). Conversely, patients with biallelic variants in *RSPH1* are associated with milder disease phenotype, higher levels of nasal nitric oxide and better lung function (Knowles et al, 2014). Moreover, patients with variants that encode the central pair (*HYDIN*), radial spokes (*RSPH1, RSPH4A and RSPH9*) or are involved in the generation of multiple motile cilia (*MCIDAS and CCNO*) do not present with situs abnormalities, since they retain the normal rotatory function of nodal cilia, essential for left-right symmetry (Werner et al, 2015; Damseh et al, 2017).

2.5.2 Genotype correlations with ciliary diagnostic findings

There are currently 50 genes associated with PCD, many of which have been associated with specific structural elements such as those that encode proteins in the outer dynein arms, the inner dynein arms, the central apparatus, and radial spokes (Fig. 13).
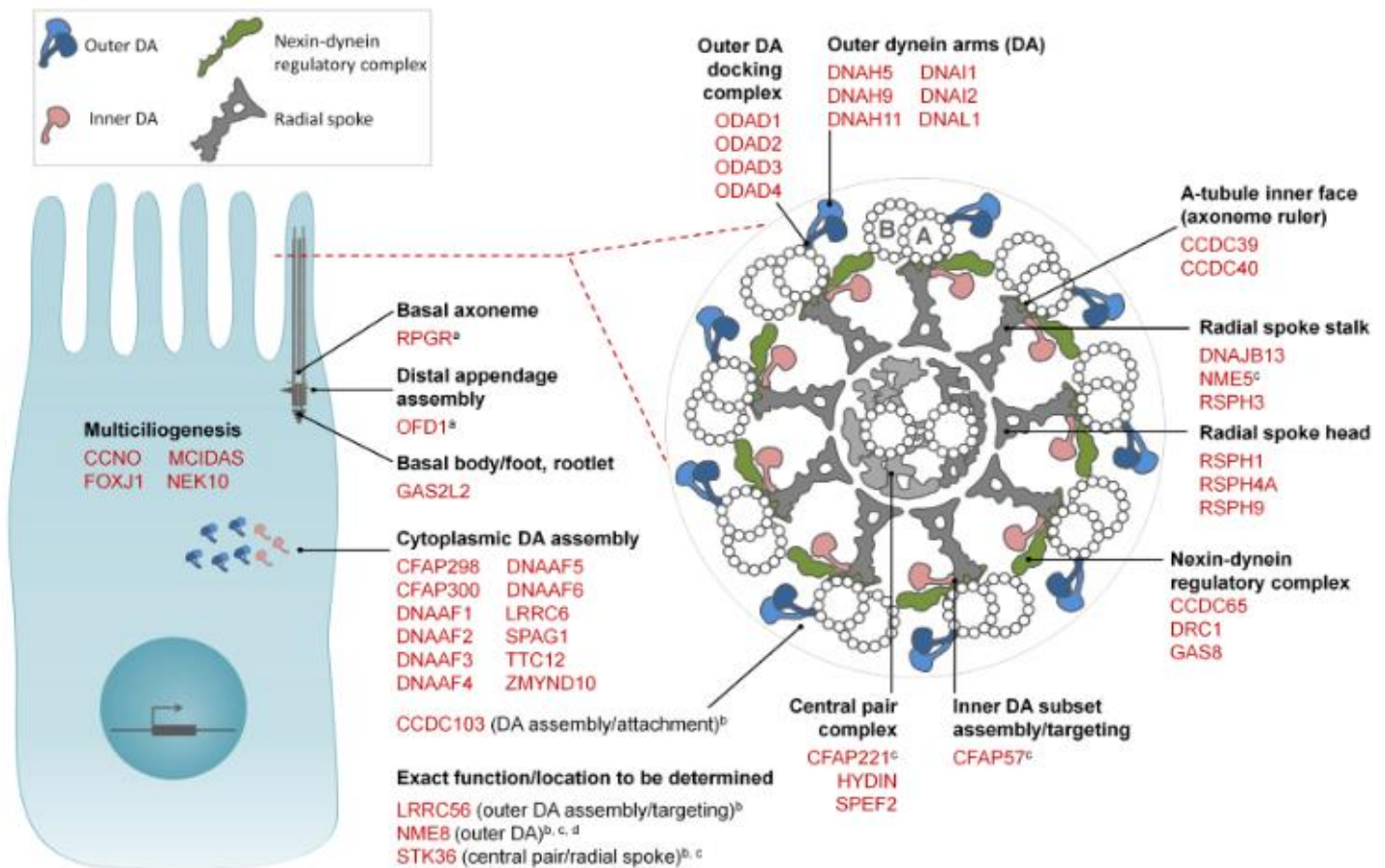
Figure 13. Location and function of proteins encoded by PCD genes. Adapted from Legendre, Zaragosi and Mitchinson, 2021. a- syndromic PCD forms; b- proposed functions; c – genetic results so far based upon data from one human family; d- not upregulated during ciliogenesis

A tight correlation exists between the implicated gene in PCD patients and the axonemal ultrastructural phenotype of their respiratory cilia, the ciliary beat pattern and frequency and even the levels of nasal nitric oxide (Table 2). For example, gene variants that lead to loss of functional cilia lead to low levels of nasal nitric oxide. Combined outer and inner dynein arm defects leads to virtually immotile cilia, whereas isolated ODA absence is associated with residual movement and vast areas of static cilia and central apparatus defects create circular, whirling patterns (Chilvers, Rutman and O'Callaghan, 2003; Raidt et al, 2014; Rubbo and Lucas, 2017).

Table 2 Correlation between specific genes and ultrastructural findings and other diagnostic and clinical features. Adapted from Brennan, Ferkol and Davis, 2021 and Hyland and Brody, 2022.

ULTRASTRUTURE	GENE	CHROMOSOME LOCUS	nNO	MOTILITY	RESPIRATORY SYMPTOMS	LATERALITY DEFECTS	SUBFERTILITY	PCD CAUSATIVE GENE PERCENTAGE
OUTER DYNEIN ARM DEFECT	DNAH5	5p15.2	LOW	IMMOTILE OR STIFF	YES	YES	YES	15-29%
	DNAI1	9p13.3	LOW	REDUCED, MINIMAL MOVEMENTS	YES	YES	YES	2-10%
	DNAJ2	17q25.1	NR	REDUCED, MINIMAL MOVEMENTS	YES	YES	YES	NR
	DNAL1	14q24.3	LOW	IMMOTILE OR WEAK	YES	YES	NR	NR
	NME8	7p14.1	NR	NORMAL	YES	YES	YES	NR
	CCDC114	19q13.33	LOW	IMMOTILE OR FLICKERING	YES	YES	NR	NR
	CCDC151	19q13.2	NR	IMMOTILE	YES	YES	NR	<3%
	ARMC4	10q12.1	LOW	FLICKERING	YES	YES	NR	<3%
	TTC25	17q21.2	LOW	IMMOTILE OR FLICKERING	YES	YES	NR	<1%
	CCDC103 (a)	17q21.31	LOW OR NORMAL	IMMOTILE OR NORMAL	YES	YES	YES	<4%
OUTER AND INNER DYNEIN ARM DEFECTS	DNAAF1	16q24.1	NR	IMMOTILE	YES	YES	YES	NR
	DNAAF2	14q21.3	LOW	IMMOTILE	YES	YES	YES	<1%
	DNAAF3	19q13.42	LOW	IMMOTILE	YES	YES	NR	<1%
	LRRC6	8q24.22	LOW	IMMOTILE	YES	YES	YES	<1%
	HEATR2/ DNAAF5	7p22.3	LOW	IMMOTILE OR MINIMAL MOVEMENT	YES	YES	YES	NR
	ZMYND10	3p21.31	LOW	IMMOTILE	YES	YES	YES	<2-4%
	DYX1C1	15q21.3	LOW	IMMOTILE	YES	YES	YES	<1%
	SPAG1	8q22.2	LOW	IMMOTILE	YES	YES	YES	<4%
	PIH1D3 (a)(f)	Xq22.3	LOW	IMMOTILE	YES	YES	YES	NR
	CFAP300	11q22.1	LOW	IMMOTILE	YES	YES	YES	NR
	CFAP298	21q22.11	LOW	IMMOTILE	YES	YES	NR	NR
INNER DYNEIN ARM DEFECT WITH AXONEMAL DISORGANISATION	CCDC39	3q26.33	LOW	IMMOTILE	YES	YES	YES	4-9%
	CCDC40	17q25.3	LOW	IMMOTILE OR STIFF	YES	YES	YES	3-4%
	TTC12	11q23.2	LOW OR NORMAL	NORMAL, IMMOTILE OR REDUCED BEATING ANGLE	YES	NR	YES	NR
OLIGOCILIA	CCNO	5q11.2	LOW	HYPOKINETIC	YES	NR	YES	<1%
	MCIDAS	5q11.2	LOW	IMMOTILE	YES	NR	YES	<1%
	FOXJ1 (g)	17q25.1	NORMAL	NORMAL	YES	YES	YES	<1%
NORMAL OR NEAR NORMAL CILIARY ULTRASTRUCTURE	DNAH1	3q21.1	NR	NR	YES	YES	YES	<1%
	DNAH9	17p12	LOW OR NORMAL	HYPOKINETIC OR REDUCED DISTAL BENDING	YES	YES	YES	NR
	DNAH11	7p15.3	LOW	HYPERKINETIC	YES	YES	YES	6-9%
	DNAH17	17q25.3	NR	NR	NR	NR	YES	NR
	LRRC56	11p15.5	LOW	IMMMOTILE, STIFF OR TWITCHING	YES	YES	NR	NR
	GAS2L2 (d)	17q12	LOW OR NORMAL	HYPERKINETIC	YES	NR	NR	NR
	HYDIN	16q22.2	LOW OR NORMAL	IMMOTILE, RIGID, ROTATIONAL	YES	NR	YES	<1%
	CFAP221	2q14.2	NORMAL	CIRCULAR PATTERN	YES	NR	NR	NR
	SPEF2	5p13.2	LOW	STIFF, ROTATIONAL PATTERN	YES	NR	YES	NR
	STK36	2q35	NORMAL	STIFF	YES	NR	YES	NR
	CCDC164	2p23.3	LOW	STIFF	YES	NR	NR	NR
	CCDC65	12q13.12	LOW	STIFF, HYPERKINETIC	YES	NR	NR	NR
	GAS8	16q24.3	LOW OR NORMAL	NORMAL OR SUBTLE BEAT ABNORMALITY	YES	NR	YES	NR
	RSPH1 (b)	21q22.3	LOW OR NORMAL	REDUCED BENDING ANGLE	YES	NR	YES	<1%
	RSPH3 (b)	6q25.3	LOW	REDUCED BENDING ANGLE	YES	NR	YES	NR
	RSPH4A (b)	6q22.1	LOW OR NORMAL	ROTATIONAL PATTERN, STIFF	YES	NR	NR	<1%
	RSPH9 (b)	6p21.1	LOW OR NORMAL	ROTATIONAL PATTERN, STIFF	YES	NR	YES	NR
	DNAJB13	11q13.4	LOW	REDUCED AMPLITUDE	YES	NR	YES	NR
	CFAP53	18q21.1	NORMAL	NORMAL	YES	YES	NR	NR
	NEK10 (c)	3p24.1	NORMAL	NORMAL	YES	NR	NR	NR
	OFD1 (f)	Xp22.2	LOW	NORMAL, IMMOTILE OR STIFF	YES	YES	NR	NR
	NME5 (e)	5q31.2	NR	NR	YES	NR	NR	NR
	RPGR (f)	Xp11.4	NORMAL	NORMAL OR UNCOORDINATED	YES	NR	NR	NR

a- mutation specific; b- absent or defective radial spokes may be an acquired defect; c- shortened cilia; d- disoriented cilia; e- absent central pair; f- X-Linked transmission; g- autosomal dominant transmission; NR – not reported

In summary, diagnosis of PCD relies on a combination of clinical features and a series of diagnostic tools many of which are expensive and are required to be conducted and interpreted by specialists. Furthermore, genetic diagnosis can help confirm the diagnosis of PCD which in turn can aid family planning, carrier testing, prenatal testing, genetic counselling and possibly in the future individualized treatment (Kim et al, 2014). However, genetic testing and interpretation of results is not straightforward and needs experts who are aware of the complexities of PCD genotyping and follow national and international best practice guidelines.

2.6   Aims and Hypothesis

This research project was conducted with the main intention of providing a clinical molecular diagnosis, to individuals who presented with cilia and respiratory symptoms suggestive of a diagnosis of primary ciliary dyskinesia, an autosomal recessive disorder. Patients included in the project had previously been found to be heterozygous for a variant in a known PCD-associated gene. It was then postulated that a second pathogenic variant may be harboured in the non-coding region of the same gene. Furthermore, the results of this project may support the need and relevance of implementing the assessment of variants in non-coding regions in the diagnostic strategies for PCD. Ultimately, a complete genetic diagnosis will allow carrier testing in family members, guide family planning and enable the inclusion of the patient in future relevant gene therapy procedures or treatment regimes.

2.7   Materials and Methods

2.7.1 Case selection

In this project, cases selected were patients who had previously undergone diagnostic NGS testing at the Royal Brompton and Harefield CGGL due to a clinical diagnosis or suspicion of PCD. Patients who were heterozygous for a single pathogenic or likely pathogenic variant in one of the genes present on the diagnostic PCD NGS panel were included. Additionally, 3 patients with a single variant of unknown significance predicted to alter splicing were also included.

Written informed consent was obtained from all participants or their guardians. Any potentially pathogenic variants identified during this project will be reviewed by a Clinical Scientist and if appropriate, results will be returned to the referring clinicians and patients.

2.7.2 Next Generation Sequencing

2.7.2.1   Panel design and validation

Selection of genes to be included in the project's NGS panel was guided by the patient's initial diagnostic result. Two custom NGS panels containing 17 known PCD- associated genes (Table 3) covering the coding, intronic and untranslated regions of the genes were designed. The NA12878 Genome In A Bottle (GIAB) reference sample (NIST, USA) was included to validate the panel, and a negative control was used to exclude contamination.

Table 3. Custom next generation sequencing PCD gene panels

PANEL 1	PANEL 2
CCDC103	ARMC4
CCDC114	CCDC151
CCDC40	CCDC39
CCNO	DRC1
DNAAF1	DYNC2H1
DNAH11	MCIDAS
DNAH5	HYDIN
DNAH9	DNAI1
SPAG1	

2.7.2.2   DNA extraction, library preparation and sequencing

Previously extracted genomic DNA (gDNA) using the QIAGEN EZ1 Advanced XL or QIAGEN QIAsymphony instrument, following the manufacturer's protocol was quantified using an input of 2µl of DNA in the Qubit 2.0 fluorometer according to manufacturer's instructions.

Genomic DNA was prepared for NGS using a TWIST Bioscience library preparation kit (TWIST Biosciences, USA). Libraries were prepared according to the manufacturer's protocols. Enzymatic fragmentation, end repair and dA-tailing were performed on gDNA followed by the ligation and amplification of TWIST Universal adapters. Individual libraries were then multiplexed and enriched using the TWIST Target Enrichment protocol with customized capture probes to generate libraries enriched with the PCD whole-gene targets. Quality control steps were performed using Agilent TapeStation 2200 D1000 kits (Agilent, USA) following fragmentation and hybridization.

The enriched pools were then diluted and sequenced on the Illumina NextSeq550 sequencer (Illumina, Inc., San Diego, USA) as 149bp paired-end reads using a 300-cycle mid-output v2 cartridge. All samples plus the GIAB were sequenced in parallel.

2.7.2.3   Bioinformatic analysis

Bioinformatic analysis was undertaken by bioinformaticians at the RBH CGGL using the in-house clinical diagnostics bioinformatics pipeline. In summary, the output of the Nextseq550 was demultiplexed using BCL2FASTQ v2.16.0.10 (Illumina Inc., USA) the output sequence reads were then aligned to the reference genome (GRCh38) by BWA-MEM v0.7.17. To improve performance and remove bias, poor quality read ends were trimmed using PRINSEQ v0.20.4 with PCR and optical duplicates flagged using Markduplicates. Variants were called using the GATK4 (Genome Analysis Tool Kit) haplotype caller v4.0.8.1. Following variant calling, variants were annotated by Alamut Batch v1.9 (Interactive Biosoftware, France). Copy number variants (CNVs) were called using an accredited in-house CNV caller which is based on average read depth across predefined regions, compared to those of neighbouring fragments, as well as to a distribution of read depths of the same fragment obtained from other samples with similar read depth patterns. Throughout the bioinformatics pipeline, sequence quality, mapping quality, coverage and target region read depth were assessed using Sequencing Analysis Viewer v2.4.7, FASTQC v0.10.1, GATK PICARD v2.18.7 tools suite and in-house tools. In-house variant prioritisation tools were used to exclude common variants (seen in-house >5 cases or seen in GnomAD v2.1.1 with a MAF >5%). A filter was applied to select the canonical transcripts and further rank variants using the ACMG best practice guidelines.

2.7.3 Manual variant interpretation

Variants that were highlighted as potentially pathogenic by the bioinformatics pipeline were inspected manually. Alamut Visual v2.15 (Interactive Biosoftware, France) was used to assess variants based on allele frequency in control database, including genome aggregation database (gnomAD, 2022), a search of published literature, the Clinical Variant database (ClinVar, 2022) and the Human Gene Mutation database (HGMD, 2022).

2.7.4 *In silico* analysis

Prediction algorithms available in Alamut visual (interactive software) and Splice AI were used for splicing predictions. These included SpliceSiteFinder-like, MaxEntScan, NNSPLICE and GeneSplicer. Variants were considered to have a potential effect on splicing when MaxEntScan predicted a >15% reduction compared to the reference, simultaneously with a SpliceSiteFinder-Like prediction >5% reduction in splice site strength (Ellard et al, 2020) and a delta score higher than 0.2 on Splice AI was obtained (Jaganathan et al, 2019).

2.7.5 Selection of types of samples

Following analyses of expression of the different PCD – related genes in different tissues using the Human Protein Atlas database (Table 4) it was determined that RNA used in this project would have to be sourced from nasal epithelial cells collected by nasal brushing. Transcript abundance is estimated and used to compare to compare genes expression between tissues and it is reported in normalized transcript per million.

Table 4. Protein expression of relevant PCD-related genes in different types of tissue samples

| GENE | NUMBER OF EXONS | NUMBER OF TRANSCRIPTS | SIZE IN BP | RESPIRATORY SYSTEM | MALE REPRODUCTIVE SYSTEM | | | | BREAST AND FEMALE REPRODUCTIVE SYSTEM | | | | | | | BONE MARROW AND LYMPHOID TISSUES | | | | | |
				LUNG	TESTIS	EPIDIDYMIS	SEMINAL VESICLE	PROSTATE	VAGINA	OVARY	FALLOPIAN TUBE	ENDOMETRIUM	CERVIX	PLACENTA	BREAST	APPENDIX	SPLEEN	LYMPH NODE	TONSIL	BONE MARROW	THYMUS
CCDC40	20	16	67982	3.2	8.5	2.6	1.6	3.2	1.2	1.7	15	3.2	2.1	0.6	2.2	0.9	1.3	0.7	0.6	0.3	0.3
CCDC151	13	5	19178	2.3	10.5	0.3	0.5	0.8	0.8	0.6	8.5	0.6	1.2	0.4	0.4	0	2.6	0	0	0	0
DNAH5	79	14	258153	1.7	0.3	0.1	0	1	0	0	3.9	0.1	0.3	0	1.6	0	0	0	0	0	0
DNAH11	82	10	362354	1.2	0.5	0.3	0.1	0.4	0.2	0.3	5.3	0.3	0.3	0.3	0	0	0	0	0	0	0.3
DNAAF1	12	10	36661	3.6	45.7	1.4	0.2	0.2	0	0.4	22.9	0.5	1.3	0.2	0.3	0.1	0.7	0	0	1.3	0
DNAI1	20	9	66238	4.6	33.8	0.2	0.1	0.2	0	0.1	34.2	0.9	1.3	0	0.5	0	0	0	0	0	0
DYNC2H1	90	11	374433	2.8	8	3.8	2.5	2.6	3	5.6	6.5	4.6	4.7	1.4	3.6	1.2	2	0.9	0.7	0.3	1.8
ARMC4	16	13	29520	2.6	41	0.6	1.8	2.5	0.3	1.6	23.4	1	1.1	0	0.2	0.3	0.2	0.1	0	0.7	0.6
DRC1	17	8	58800	3.1	21	0.1	0.2	0.1	0	0	39.6	0.9	0.8	0	0	0	0	0	0	0.7	0.3
HYDIN	86	23	432639	1.4	4.7	2	0.3	0.8	0.4	1.1	5.3	0.9	1	0.1	0.5	0.3	0.4	0.3	0	0.5	0

(Table header band above the columns reads: HUMAN PROTEIN ATLAS)

2.7.6 RNA extraction

After nasal brushing, RNA was extracted from uncultured nasal epithelial cells stored in RNAlater solution using the Maxwell simplyRNA cells kit (Promega, USA). The cells were centrifuged at 5000x g for 3 minutes and medium was removed. Greater centrifugal forces than what was stated on the protocol were required due to the higher density of RNAlater solutions. Samples were homogenized by adding 200 µl of chilled 1-Thioglycerol. 200 µl of lysis buffer was then added to the cell's homogenate and vortexed vigorously. The Maxwell RSC simplyRNA cartridges were prepared and 400 µl of the lysate were transferred to the cartridge. The Agilent TapeStation 2200 High Sensitivity kits (Agilent, USA) was used to quantify the RNA integrity number (RIN).

2.7.7 Reverse transcription

Following RNA extraction, reverse transcription was conducted using the SuperScript IV Reverse Transcriptase (Invitrogen, Themo Fisher Scientific, USA). The reverse transcription (RT) reaction was primed using random hexamers. Template RNA sample was mixed and briefly centrifuged with dNTPs, random hexamers, and nuclease-free water. RNA-primer was heated for 5 minutes at 65ºC and then incubated on ice for at least 1 minute. Meanwhile, a RT reaction mix containing 5xSSIV buffer, DTT, RNaseOUT recombinant RNase inhibitor and SuperScript IV Reverse Transcriptase was prepared. This RT reaction mix was then added to the annealed RNA and incubated at 23ºC for 10 minutes. Samples were incubated for a further 10 minutes at 50ºC and reaction was then terminated at 80ºC for 10 minutes. Afterwards RT reaction samples were stored at -20ºC.

2.7.8 Sanger sequencing

2.7.8.1 Primer design

Primers were designed for variants highlighted by manual and in silico analysis, using the IDT Oligo Entry tool available https://eu.idtdna.com/site/order/oligoentry (IDT, USA). Primers were designed to be 20-25 bp long, with a GC rich between 40-60%, and a melting temperature between 60-65ºC. Primers were checked for SNPs using the Gene Tools SNPCheck V3 available at https://genetools.org/SNPCheck/snpcheck.htm . Primers with SNPs were also checked against preliminary reports generated after NGS.

2.7.8.2 Polymerase chain reaction

Amplification of the cDNA fragments harbouring the variants of interest were performed using 23 µl of the Megamix Blue master mix, 2 µl of the required diluted forward and reverse primers for the respective variants and 1 µl of the respective cDNA sample. The PCR cycling conditions were 95ºC for 3 minutes followed by 30 cycles of 95ºC during 30 seconds for denaturation, 58ºC during 30 seconds for annealing and 72ºC during 45 seconds for elongation. Healthy controls were included for the purposes of comparison and validation of results. A negative control was used to ensure there was no contamination. Furthermore, samples were loaded on the Agilent TapeStation 2200 in order to ensure that the fragments are of the correct size and there was no non-specific amplification nor contamination.

2.7.8.3   PCR cleanup

PCR products were cleaned up using Beckman Coulter sample purification beads (Beckman Coulter, USA). 36 µl of beads were mixed with each sample and incubated for 10 minutes at room temperature. Samples were briefly centrifuged and placed on the plate magnet and the supernatant was removed and discarded. With the plate still on the magnet, beads were washed twice with 70% ethanol. After all the ethanol was removed, the beads were let to air dry for 10 minutes at room temperature. Finally, beads were resuspended in PCR-grade water. After samples were placed on the plate magnet and allowed to clear, 45 µl were transferred to a new 96-well plate.

2.7.8.4   Sequencing

After amplification and purification of the samples, sequencing was performed using the ABI 3500 (Applied Biosystems, USA). Sequencing reaction was performed by mixing 0.75 µl of BigDye V3.1 terminator cycle (Applied Biosystems, USA), 4 µl of 2.5x Buffer, 4.25 µl of water, 1 µl of the cDNA template and 1 µl of either forward or reverse primer at 5pM/ µl. The BigDye thermal cycler conditions were an incubation at 96ºC during a minute, followed by 25 cycles at 96ºC during 10 seconds for denaturation, 50ºC during 5 seconds for annealing and 60ºC for 1 minute and 15 seconds for extension.

Clean-up of sequencing reactions was conducted by mixing 10 µl of CleanSeq beads (Beckman Coulter, USA) to each sample. The beads were washed twice with 85% ethanol whilst the plate was placed on a plate magnet. After all the ethanol was removed, 40 µl of Hi-Di formamide (Applied biosystems, USA) was added to each sample well. The beads were fully resuspended and after centrifuging the plate was placed on the magnet plate until the solution was clear. 35 µl of each sample were transferred to a new labelled 96-well Biosystems Micro-Amp plate (Applied biosystems, USA) and the plate was loaded to the ABI3500 genetic analyser (Applied biosystems, USA).

2.7.8.5   Visualization and data analysis

Sequencing results were visualised and analysed using SeqPatient software (JSI medical systems, Germany). Analysis was conducted by comparison of patient sample variants with sequences of healthy controls.

2.7.9 ddPCR

2.7.9.1   Assay procedure

Samples were initially digested using the restriction enzyme EcoRI. A reaction volume of 20 µl, containing the digested sample, forward and reverse primers, probes and ddPCR supermix for probes (no dUTP) mastermix (Bio-Rad, USA) was loaded into an 8-well cartridge. Droplets were generated using the QX200 droplet generator (Bio-Rad, USA), following manufacturer's instructions. Once droplet generation was completed, 45 µl of droplets were aspirated, transferred to the droplet reader plate, and sealed with a Bio-Rad PX1 plate sealer. Assays were then amplified in a C1000 thermal cycler with the following cycling conditions: an initial activation of the Taq polymerase at 95ºC for 10 minutes, followed by 39 cycles at 94ºC during 30 seconds for denaturation, 60ºC during 1 minute for annealing and extension followed by a final 10-minute enzyme deactivation step at 98ºC. After amplification, droplets were quantified using the QX200 droplet reader (Bio-Rad, USA). A housekeeping gene (*RPP30*) with no known copy number variation covering exon 7, was used a reference gene. Probes targeting the region of interest were fluorescently labeled with FAM, whereas the reference gene/wild type was fluorescently labeled with HEX. A negative control was also used to ensure there was no contamination.

2.7.9.2   Probe and primer design

Probes were designed using the primerquest design tool available at the IDT website https://sg.idtdna.com/site . Probes length, T_m of hydrolysis and GC content was followed in accordance with the instructions of Bio-Rad. Probe sequence was also checked for any possibility of sequence similarity to other areas of the genome using the BLAST function available at www.ensembl.org.

Primers were designed using the Oligo Entry tool at the IDT website and in accordance with Bio-Rad instructions. Primers were checked for SNPs using the Gene Tools SNPCheck V3 available at https://genetools.org/SNPCheck/snpcheck.htm

2.7.9.3   Data acquisition and analysis

Droplet count was stored and analysed using the Quantasoft software (Bio-Rad). Groups were displayed in blue, orange, green and black. Blue is the probe of interest, green is the *RPP30* gene, orange are the droplets with both probes present and black are negative droplets. These groups are automatically assigned into distinct areas by the software. The number of positive and negative droplets read in each channel is then used by the software to calculate the concentration of the sample.

2.8 Results

This project included 23 patients with a clinical diagnosis or suspicion of PCD, who had previously been referred to the CGGL and had been found to possess a single heterozygous variant. Whilst bioinformatic analysis was conducted on the samples of the 23 patients included in this research project, data regarding their cilia diagnostic tests was compiled in order to better understand and guide possible genetic findings. Information available at EPR, regarding cilia ultrastructure, cilia beat pattern and frequency, nitric oxide levels, immunofluorescence staining as well as relevant medical history and symptoms was consolidated for patients included on panel 1 (Table 5) and panel 2 (Table 6). Furthermore, previously identified variants by molecular diagnostic screening of the coding region of the 47 PCD-associated genes and their respective ACMG classification for pathogenicity was also included.

Table 5. Cilia diagnostic features of patients included on panel 1. Information regarding cilia beat frequency and pattern, ultrastructure, nNO and IF was gathered from EPR. Initial heterozygous variant information is also included.

CASE	GENE OF INTEREST	INITIAL HETEROZYGOUS VARIANT	VARIANT CLASSIFICATION	NITRIC OXIDE LEVELS	BEAT FREQUENCY	LIGHT MICROSCOPY	ULTRASTRUCTURE
1	CCDC40	c.1989+1G>A	PATHOGENIC	EXTREMELY LOW	-	STIFF, SLOW AND STATIC IN PLACES	IDA ABSENCE WITH RADIAL SPOKE DISARRANGEMENT
2	DNAH5	c.10815del, p.(Pro3606HisfsTer23)	PATHOGENIC	-	-	-	-
3	DNAH11	c.13380_13383dup, p.(Ala4462LeufsTer22)	LIKELY PATHOGENIC	EXTREMELY LOW	22.4 Hz	HYPERFREQUENT WITH A JITTERY DYSKINETIC PATTERN	NORMAL
4	DNAH11	c.3020T>G, p.(Leu1007Ter)	LIKELY PATHOGENIC	BORDERLINE LOW	12.8-23 Hz	HYPERKINETIC OR STATIC	OUTER TUBULES MISSING
5	DNAH11	c.4669C>T, p.(Arg1557Ter)	LIKELY PATHOGENIC	NO DATA	10.57 Hz	MIXED BEAT - SOME CILIA COORDINATED OTHERS APPEARED STIFF/ STATIC	NORMAL
6	DNAH5	c.5281C>T, p.(Arg1761Ter)	PATHOGENIC	NORMAL	2.75 Hz	SOME STIFF/ STATIC CILIA	ODA DEFECT
7	DNAAF1	c.882G>A, p.(Trp294Ter)	LIKELY PATHOGENIC	NORMAL	11.8 Hz	MIXED BEAT- SOME AREAS NORMAL OTHERS STIFF, REDUCED BEAT AMPLITUDE AND WAVY	ODA DEFECT
8	DNAH11	c.10691+2T>A	PATHOGENIC	EXTREMELY LOW	13.05 Hz	CILIA STATIC, TWITCHING OR HYPERFREQUENT	NORMAL
9	DNAAF1	c.114del, p.(Cys39AlafsTer50)	LIKELY PATHOGENIC	NORMAL	8.12 Hz	CILIA SLIGHTLY STIFF AND DYSKINETIC	OUTER TUBULES MISSING
10	DNAH5	c.4360C>T, p.(Arg1454Ter)	VUS	NORMAL	NO DATA	MAINLY STATIC WITH OCCASIONAL DYSKINETIC CILIA	ODA DEFECT
11	DNAH11	c.6664C>T, p.(Arg2222Ter)	PATHOGENIC	LOW	7.97 Hz	SLIGHTLY STIFF AND JITTERY CILIA	NORMAL
12	DNAH11	c.4011+5G>A	VUS	LOW	MIXED	DYSKINETIC, JITTERY WITH SOME STATIC STRIPS SEEN	NORMAL
13	DNAH11	EXON 32 DELETION	PATHOGENIC	EXTREMELY LOW	8.6 Hz	DYSKINETIC WITH EITHER MOTIONLESS OR FLICKERING WITH AN INCOMPLETE STROKE	SOME EVIDENCE OF SHORTENED ODA
14	CCDC40	c.248del, p.(Ala83ValfsTer84)	PATHOGENIC	EXTREMELY LOW	7.13 Hz	IMMOTILE	MICROTUBULAR DISORGANISATION AND ABSENCE OF IDA

CASE	GENE OF INTEREST	IMMUNOFLUORESCENCE STAINING									SITUS INVERSUS	SYMPTOMS	ADDITIONAL NOTES
		DNAH5	DNAH11	DNALI1	RSPH4A	RSPH1	RSPH9	SPEF2	IFT88	GAS8			
1	CCDC40	-	-	-	-	-	-	-	-	-	YES	DEXTRACARDIA	
2*	DNAH5	-	-	-	-	-	-	-	-	-	NO	BRONCHIECTASIS	
3	DNAH11	-	-	-	-	-	-	-	-	-	NO	BRONCHIECTASIS	
4	DNAH11	FAILED	-	-	FAILED	-	-	-	-	FAILED	NO	BRONCHIECTASIS	
5	DNAH11	PRESENT	-	-	PRESENT	-	-	-	-	PRESENT	NO	BRONCHIECTASIS	
6	DNAH5	PRESENT	-	-	PRESENT	-	-	-	-	PRESENT	NO	WET COUGH AND OTORRHOEA	
7	DNAAF1	PRESENT	-	-	PRESENT	-	-	-	-	PRESENT	NO	BRONCHIECTASIS	
8	DNAH11	PRESENT	-	INCONCLUSIVE	PRESENT	-	-	-	-	-	NO	CHRONIC RHINOSINUSITIS	PATERNAL GRANDMOTHER DEMONSTRATES SYMPTOMS SIMILAR TO PCD PATIENTS
9	DNAAF1	PRESENT	-	-	PRESENT	PRESENT	PRESENT	-	-	PRESENT	NO	BRONCHIECTASIS, ASTHMA AND SEVERE RHINITIS	
10	DNAH5	-	-	-	-	-	-	-	-	-	NO	BRONCHIECTASIS	
11	DNAH11	PRESENT	-	PRESENT	PRESENT	-	-	-	-	-	NO		
12	DNAH11	PRESENT	ABSENT	-	PRESENT	-	-	-	-	-	YES	RHINITIS AND CHRONIC WET COUGH	
13	DNAH11	PRESENT	ABSENT	-	PRESENT	-	-	-	-	INCONCLUSIVE	YES	DEXTRACARDIA	YOUNGER BROTHER DIAGNOSED WITH PCD
14	CCDC40	-	-	ABSENT	-	-	-	-	-	ABSENT	NO	CHRONIC WET COUGH AND BRONCHIECTASIS	GRANDMOTHER APPEARS TO SHOW MILD SYMPTOMS

* No data for patient 2 was available on EPR

Table 6. Cilia diagnostic features of patients included on panel 2. Information regarding cilia beat frequency and pattern, ultrastructure, nNO and IF was gathered from EPR. Initial heterozygous variant information is also included.

CASE	GENE OF INTEREST	INITIAL HETEROZYGOUS VARIANT	VARIANT CLASSIFICATION	NITRIC OXIDE LEVELS	BEAT FREQUENCY	LIGHT MICROSCOPY	ULTRASTRUCTURE
15	DNAI1	c.48+2dup	PATHOGENIC	EXTREMELY LOW	NO DATA	STATIC CILIA	ODA DEFECT
16	DYNC2H1; IFT140; DNAH11	c.11381del; c.508G>A; c.695C>T	VUS	NO DATA	NO DATA	NO DATA	NO DATA
17	ARMC4	c.936+1G>A	LIKELY PATHOGENIC	NO DATA	8.31 Hz	STIFF AND UNSYNCHRONISED	OCCASIONAL LOSS OF ODA
18	DRC1; MCIDAS	c.1868A>C,p.(Asn623Thr); c.218-13_218-6dup	VUS	EXTREMELY LOW	NO DATA	NO DATA	ODA DEFECT
19	CCDC39	c.357+1G>C	PATHOGENIC	LOW	0.00 Hz	COMPLETELY IMMOTILE	IDA AND MICROTUBULAR DISORGANISATION
20	CCDC151	DELETION OF EXONS 1-3	LIKELY PATHOGENIC	NO DATA	10.29 Hz	MIXED BEAT PATTERN – NORMAL/STIFF WITH REDUCED AMPLITUDE	OCCASIONAL LOSS OF ODA
12	HYDIN	c.4548del, p.(Asn1518ThrfsTer17)	LIKELY PATHOGENIC	LOW	MIXED	DYSKINETIC, JITTERY WITH SOME STATIC STRIPS SEEN	NORMAL
21	HYDIN	c.2419_2422del, p.(Val807IlefsTer13)	LIKELY PATHOGENIC	EXTREMELY LOW	5.0 Hz	SLOW, IN SOME INSTANCES ALMOST STATIC AND DYSKINETIC	ODA DEFECT
22	HYDIN	c.7280_7287del, p.(Val2427AlafsTer13)	LIKELY PATHOGENIC	NO DATA	NO DATA	NO DATA	NO DATA
23	HYDIN	c.6669+1G>A	LIKELY PATHOGENIC	LOW	7.0 Hz	STIFF AND DYSKINETIC	ABNORMALITIES OF THE CENTRAL PAIR

CASE	GENE OF INTEREST	IMMUNOFLUORESCENCE STAINING									SITUS INVERSUS	SYMPTOMS	ADDITIONAL NOTES
		DNAH5	DNAH11	DNALI1	RSPH4A	RSPH1	RSPH9	SPEF2	IFT88	GAS8			
15	DNAI1	-	-	-	-	-	-	-	-	-	NO	BRONCHIECTASIS	
16	DYNC2H; IFT140; DNAH11	-	-	-	-	-	-	-	-	-	NO	BRONCHIECTASIS	MOTHER HAS A HISTORY OF BRONCHIECTASIS
17	ARMC4	PRESENT		N/A	PRESENT	PRESENT	PRESENT		PRESENT	PRESENT	NO	INTERMITTENT COUGH	
18	DRC1, MCIDAS	-	-	-	-	-	-	-	-	-	NO	DEXTRACARDIA	PARENTS ARE FIRST DEGREE COUSINS/ 2 SIBLINGS HAVE A CLINICAL DIAGNOSIS OF PCD
19	CCDC39	PRESENT	-	INCONCLUSIVE	-	-	-	-	-	ABSENT	NO		
20	CCDC151	PRESENT	-	-	PRESENT	-	-	-	-	PRESENT	NO	BROCHIOLITIS	FATHER HAS A SINGLE MUTATION IN CCDC151
12	HYDIN	PRESENT	-	-	PRESENT	-	-	-	-	PRESENT	YES	CHRONIC WET COUGH AND PERSISTENT RHINITIS	
21	HYDIN	PRESENT	-	PRESENT	PRESENT	PRESENT	PRESENT	-	-	PRESENT	NO	BRONCHIECTASIS	SISTER DIAGNOSED WITH PCD
22*	HYDIN	-	-	-	-	-	-	-	-	-	NO		
23	HYDIN	-	-	-	-	-	-	ABSENT	-	-	NO	BRONCHIECTASIS	

* No data for patient 22 was available on EPR. Patient 12 was included on both panel 1 and 2.

Following bioinformatic analysis, the NGS sequencing of whole PCD genes containing the coding and non-coding regions, identified hundreds of intronic variants in all the 23 patients included on this research project. Prediction algorithms available in Alamut Visual and Splice AI were used for splice predictions. Intronic variants with relevant or possibly relevant splice predictions are described in table 7. These were selected based on their frequency in control populations databases and splice predictions from Alamut Visual and Splice AI. Additionally, ultrastructure, nNO and cilia beat pattern and frequency were also taken into account, since tight correlations between the implicated gene and genotype have been described. Moreover, CNVs analysis also identified an exon 9 deletion in CCDC39 in patient 19 and a 1678bp deletion of chromosome 16 in patient 21.

Table 7. Pathogenic and potentially pathogenic variants identified by whole gene sequencing, together with Alamut visual and Splice AI splicing predictions

| CASE	GENE OF INTEREST	2ND VARIANT	gnomAD	IN-HOUSE	READ DEPTH/AB	ALAMUT VISUAL			SPLICE EFFECTS	SPLICE AI							
						SSF-L >5%	MES >15%	NNSSPLICE >10%		Δ SCORE				PRE-mRNA POSITION			
										DS_AG	DS_AL	DS_DG	DS_DL	DP_AG	DP_AL	DP_DG	DP_DL
1	CCDC40	c.1441-919G>A	0	2	91, AB=0.37	Y	Y	Y	CRYPTIC DONOR STRONGLY ACTIVATED?	0	0	0,23	0	1	7	1	-48
2	DNAH5	c.11883+523A>G	8/31396	1	78, AB=0.54	N	Y	Y	CRYPTIC ACCEPTOR STRONGLY ACTIVATED?	0,01	0	0	0	-10	-2	-10	-2
4	DNAH11	c.6042-511G>T	0	1	83, AB=0.45	Y	Y	Y		0,08	0	0	0	3	30	-15	-44
5	DNAH11	c.3853-1157A>T	27/25148	1	13, AB=0.23	Y	Y	Y	CRYPTIC ACCEPTOR STRONGLY ACTIVATED?	0	0	0,04	0				5
7	DNAAF1	c.2065+705G>A	10 / 31396	2	69, AB=0.35	N	N	N		0,03	0	0	0	8	-5	7	-50
8	DNAH11	c.8511-7T>G	1/194452	2	47, AB=0,47	N	Y	N	POSSIBLE EFFECT AT NEAREST SPLICE SITE	0	0,3	0	0	41	7	7	19
9	DNAAF1	c.864-2147C>T	1 / 31340	1	53, AB=0.38	N	Y	N	CRYPTIC ACCEPTOR STRONGLY ACTIVATED?	0	0	0	0	-44	-4	20	-46
12	DNAH11	c.8155-29A>G	0	1	116, AB=0.52	N	N	N		0	0,21	0	0	1	29	28	2
13	DNAH11	c.1152T>A	–	–	–	–	–	–		0	0	0,55	0,47	-7	43	-3	42
14	CCDC40	c.1441-919G>A	0	2	93, AB=0.45	Y	Y	Y	CRYPTIC DONOR STRONGLY ACTIVATED?	0	0	0,23	0	1	7	1	-48
15	DNAI1	c.513+85G>A	4 / 31368	1	72, AB=0.51	N	N	N		0.26	0	0.22	0.05	-46	0	11	-20
16	DYNC2H1	c.4611+250G>A	10 / 31384	1	65, AB=0.54	N	Y	Y	CRYPTIC ACCEPTOR STRONGLY ACTIVATED?	0.03	0	0	0.06	7	0	0	-47
19	CCDC39	HETEROZYGOUS EXON 9 DELETION	–	–	–	–	–	–		–	–	–	–	–	–	–	–
21	HYDIN	1,678 bp DELETION OF CHROMOSOME 16	–	–	–	–	–	–		–	–	–	–	–	–	–	–

AB = Allelic Balance; SSF-L = SpliceSiteFinder- Like; DS = Delta Score; DP = Delta Position; AG = Acceptor Gain; AL= Acceptor Loss; DG= Donor Gain; DL = Donor Loss

Delta scores range from 0 to 1 and can be interpreted as the probability that the variant affects splicing at any position within a window around it (+/- 50bp by default).

Pre mRNA-position – for each variant, Splice AI looks within a window (+/- 50bp by default) to see how the variant affects the probability of different positions in the pre-mRNA being splice acceptor or donor. Negative values are upstream (5') of the variant, and positive are downstream (3') of the variant.

## 2.8.1 Patient 4 *DNAH11* c.6042-511G>T

Patient 4 is a 48-year-old woman with bronchiectasis, borderline low nasal nitric oxide levels, normal ciliary axoneme structure on electron microscopy, hyperkinetic cilia beat pattern on light microscopy and presents absence of staining for *DNAH11* on a recently conducted immunofluorescence antibody staining technique. These findings are highly consistent and suggestive of *DNAH11* as a genetic cause for primary ciliary dyskinesia. This patient had previously been found to possess a heterozygous nonsense likely pathogenic variant in DNAH11: c.3020T>G, p.(Leu1007Ter). Following whole gene sequencing, the patient was also found to be heterozygous for an intronic variant in *DNAH11*: c.6042-511G>T. This variant is predicted, by three splice site prediction algorithms in Alamut Visual plus (Fig. 14 and 15) and Splice AI to introduce a new splice acceptor site in intron 35, 3 base pairs downstream of the variant. This variant has never been detected in control populations and has also not been reported previously.
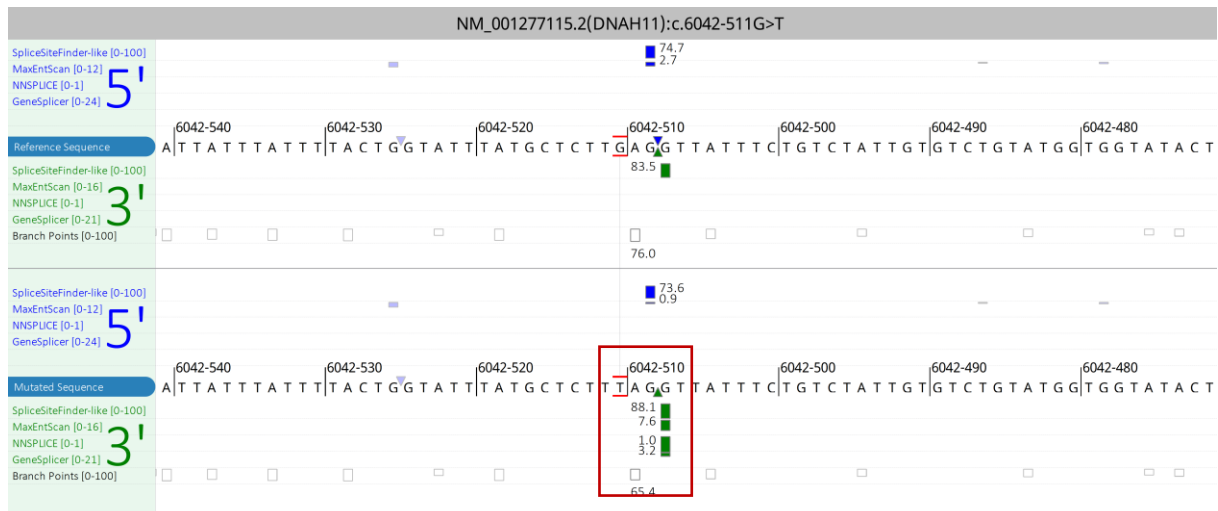
Figure 14. Heterozygous intronic DNAH11: c.6042-511G>T variant identified on patient 4. Screenshot from Alamut Visual splice predictions which foresee the introduction of a new acceptor splice site (red box) 3 base pairs downstream of the variant in intron 35. Wildtype sequence is shown in the upper window and variant sequence in the lower window.

A new cryptic donor present 177 base pairs downstream of intronic *DNAH11* variant is also expected to be introduced. This in turn was expected to create a pseudoexon within intron 35.
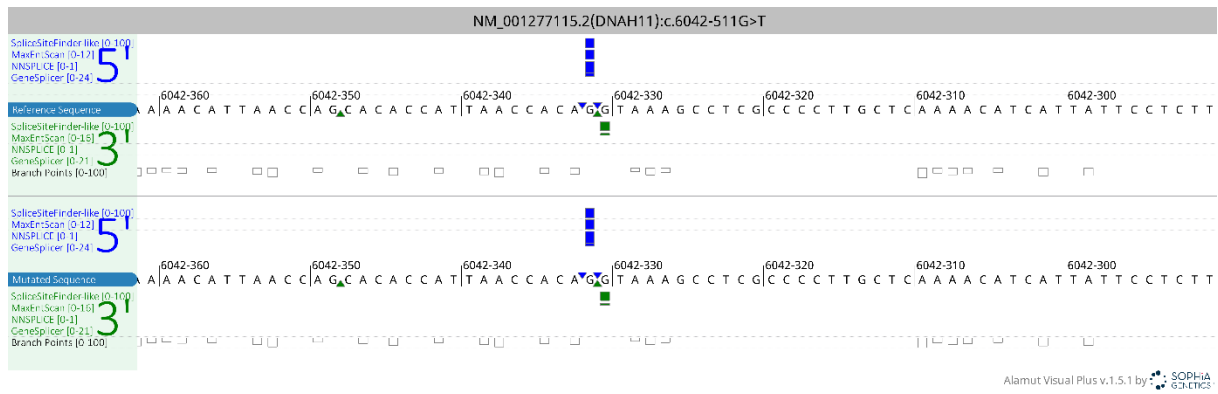
Figure 15. Screenshot from Alamut Visual splice predictions which foresee the introduction of a new cryptic donor, 177 base pairs downstream of the DNAH11 variant on intron 35.

In order to confirm in silico splicing predictions, RNA studies were conducted. Following RNA extraction and reverse transcription, cDNA was amplified and purified. Primers for exon 32 and exon 37 were used. Products of amplification and purification were then placed on the Agilent TapeStation and fragment sizes were calculated (Fig. 16). Afterwards, samples were sequenced using the ABI 3500 and compared to the healthy control sample (Fig. 17).
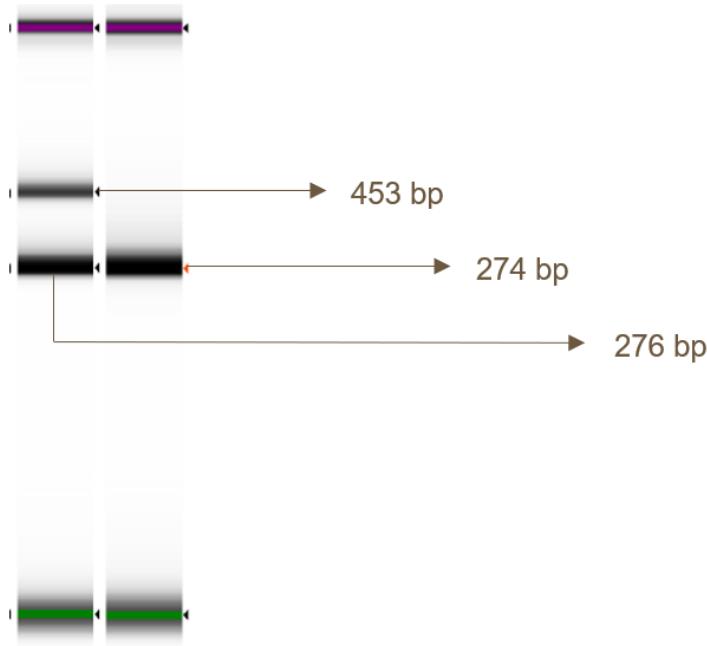
Figure 16. Agilent TapeStation of the amplified and fragmented DNAH11: c.6042-511G>T variant. On the left is patient 4 sample, whereas on the right is the healthy control sample. Patient sample shows two distinct amplicons separated by a 177 bp pair difference between the top fragment and bottom fragment.
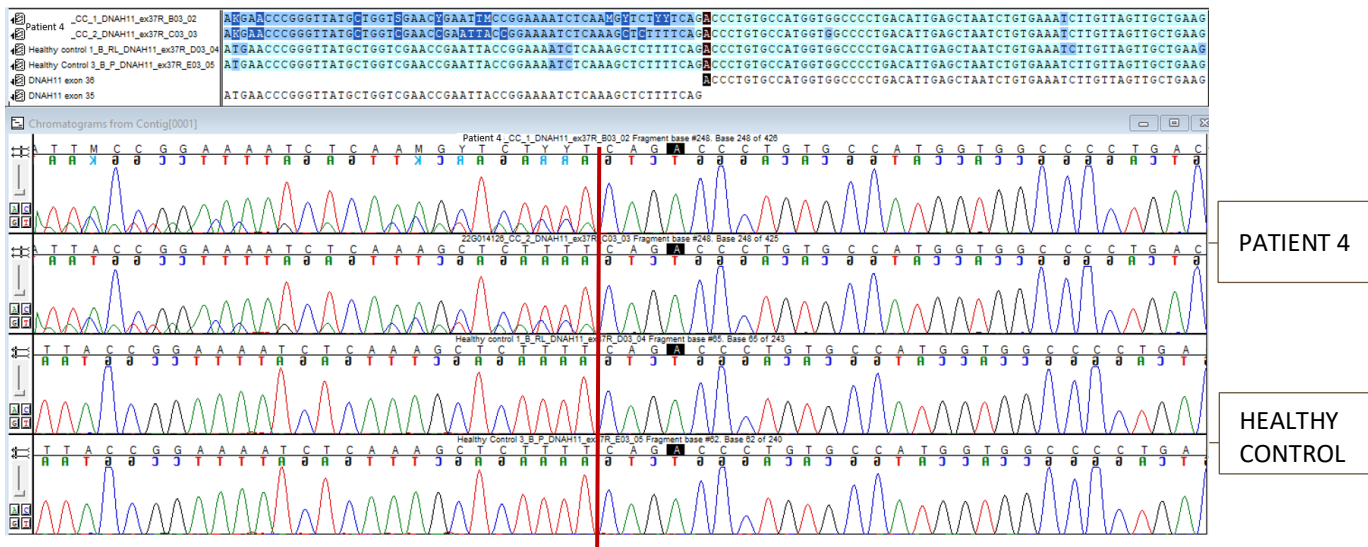

Figure 17. SeqPatient sequencing analysis of cDNA of the healthy control sample and patient 4 sample which contains a DNAH11 intronic variant. On top are the reverse sequences of patient 4 and at the bottom are the reverse sequences of the healthy control. The red line marks the 3' beginning of the 177 base pair insertion on intron 35. The complete 177bp insertion is as follows:
5'-
GTTATTTCTGTCTATTGTGTCTGTATGGTGGTATACTTCTCAGCTCCACATTTAGTGATATTGTGTTGGTATCTT GGAATCAGCCATGGTGGAAGAATTTATTCCATGGAAATAAGCAAGTGCTATTTTCTAGACAGCCGGTTGTGAA AACATTAACCAGCACACCATTAACCACAG - 3'

The existence of different nucleotide peaks on the chromatogram can be explained by the low splicing levels, which in turn result on the pseudoexon not being fully expressed.

2.8.2 Patient 13 *DNAH11* c.1152T>A

Patient 13 is a 31-year-old woman with dextrocardia, *situs inversus totalis*, extremely low nasal nitric oxide levels, a mostly normal ciliary ultrastructure with some evidence of shortened outer dynein arms on electron microscopy, hyperkinetic cilia beat pattern with twitching at the tips on light microscopy and presents absence of staining for *DNAH11* on immunofluorescence antibody staining technique. Furthermore, the patient also presents with relevant family history since her brother has been diagnosed with PCD. These findings are highly consistent and suggestive of *DNAH11* as a genetic cause for primary ciliary dyskinesia.

This patient had previously been found to be heterozygous for a pathogenic exon 32 deletion in *DNAH11*. Following whole gene sequencing, and in silico analysis using the Alamut Visual prediction tools and Splice AI, this patient was also found to contain a synonymous heterozygous variant in *DNAH11*: c.1152T>A. This variant is absent from controls such as the Exome Sequencing Project, 1000 Genomes Project or Exome Aggregation Consortium. ClinVar has one previously reported case, entered as likely benign, however considered the year and the lab who reported the variant, it most likely refers to the same patient. This variant is predicted by Splice AI to create a new splice donor gain therefore truncating the exon and leading to a 45bp loss.
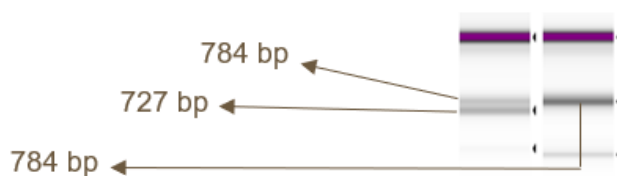

Figure 18. Agilent TapeStation of the amplified and fragmented DNAH11 c.1152T>A variant. On the left is patient 13 samples, whereas on the right is the healthy control sample. Patient sample shows two distinct amplicons.

In order to confirm in silico splicing predictions, RNA studies were conducted. Following RNA extraction and reverse transcription, cDNA was amplified and purified. Primers for exon 4 and exon 8 were used. Products of amplification and purification were then placed on the Agilent TapeStation and fragment sizes were calculated (Fig. 18). Afterwards, samples were sequenced using the ABI 3500 and compared to the healthy control sample (Fig. 19).

Figure 19. A -. SeqPatient analysis of cDNA of the healthy control sample. Background noise was generated by the amplicons highlighted on figure 12B; C – SeqPatient analysis of cDNA of patient 13 sample. Background noise was generated by the wildtype allele. Due to the proximity of the amplicon fragments, the cut out of the bands was sub-optimal therefore generating the background noise. Nevertheless, visualization of the introduction of a new splice donor site can be seen 3 bp upstream of the synonymous variant. This then resulted on the truncation of exon 6 leading to a 45 bp loss. D – Impact of the new splice donor gain on the translated protein. A new splice donor is introduced 3 bp upstream of the synonymous variant (shown in red), leading to a 45bp/ 15 amino acid truncation of exon 6 (shown in black).

2.8.3 Patient 14 *CCDC40* c.1441-919G>A

Patient 14 is a 13-year-old boy with bronchiectasis, chronic wet cough, and ear symptoms. Patient 14 presented with very low nasal nitric oxide levels, weak residual movement/ immotile cilia on light microscopy and occasional absence of the inner dynein arms and microtubular disorganisation on electron microscopy. Furthermore, using immunofluorescent antibody staining techniques it was possible to confirm the expected absence of staining of *DNALI1* and *GAS8*, both of which stain for the inner dynein arms and dynein regulatory complex proteins. Previous genetic testing had detected, a single nonsense heterozygous variant on *CCDC40* c.248del, which fits well with the above beat pattern and ultrastructural abnormalities.

Following whole gene sequencing, the patient was also found to contain a heterozygous intronic variant in *CCDC40*: c.1441-919G>A. This variant is predicted, by three splice site prediction algorithms in Alamut Visual (Fig. 20) and Splice AI to introduce a new splice donor site on intron 9, 1 base pair downstream of the variant. This variant has never been detected in control populations, has not been reported previously either and additionally is also present in patient 1 of this research project.
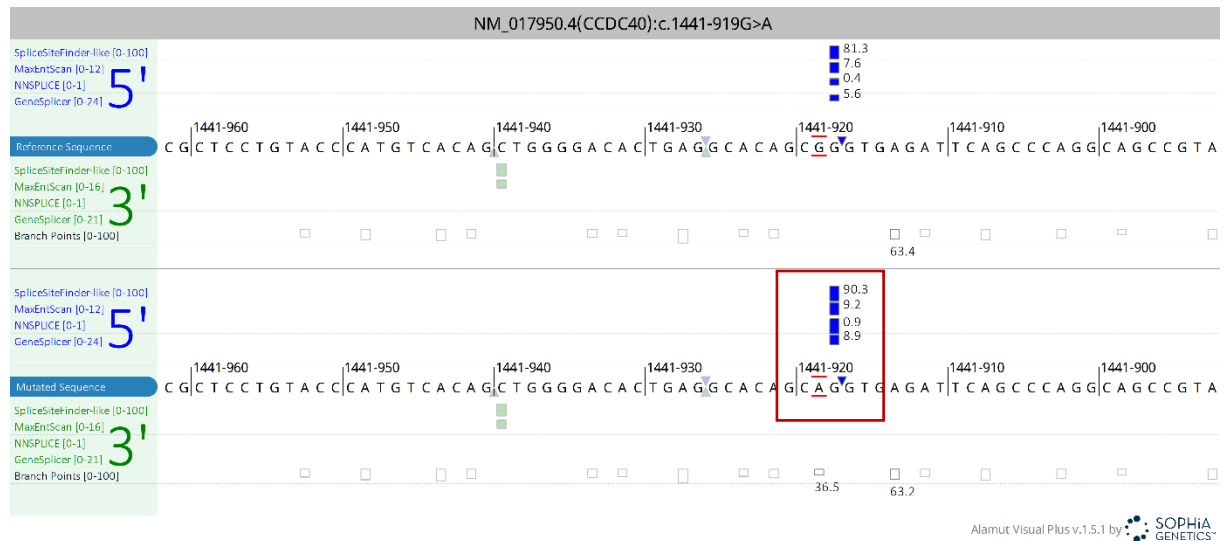

Figure 20. Heterozygous intronic CCDC40: c.1441-919G>A variant identified on patient 14. Screenshot from Alamut Visual splice predictions which foresee the introduction of a new donor splice site (red box), 1 bp downstream of the variant on intron 9. Wildtype sequence is shown in the upper window and variant sequence is shown on the lower window.

This variant is predicted to introduce a novel 23bp pseudoexon. In order to confirm in silico splicing predictions, RNA studies were performed. Following RNA extraction and reverse transcription, cDNA was amplified and purified. Primers for exon 8 and exon 11 were used. Products of amplification and purification were then placed on the Agilent TapeStation and fragment sizes were calculated (Fig. 21). Afterwards, samples were sequenced using the ABI 3500 and compared to the healthy control sample (Fig. 22).
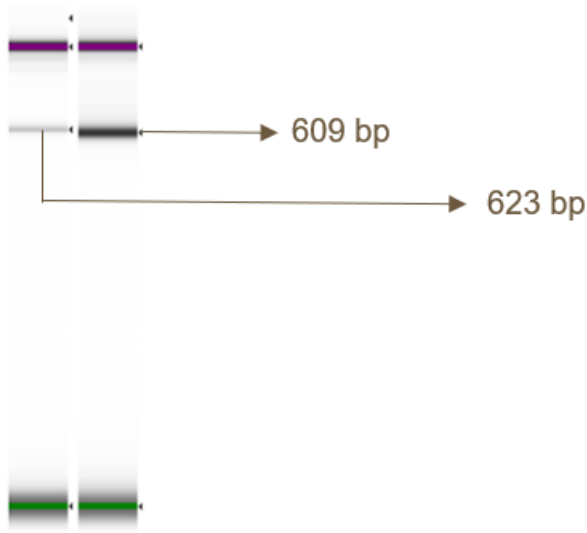
Figure 21. Agilent TapeStation of the amplified and fragmented CCDC40: c.1441-919G>A variant. On the left is patient 14 sample, whereas on the right is the healthy control sample. Patient sample shows an amplicon with a 24bp larger fragment when compared to the healthy control sample.

Figure 22. Sanger sequencing chromatogram of patient 14 and healthy control sample. Patient sample demonstrates equal nucleotide peaks on both reverse and forward sequencing, therefore not proving the expected 23bp pseudoexon insertion.

2.8.4 Patient 19 – *CCDC39* exon 9 deletion
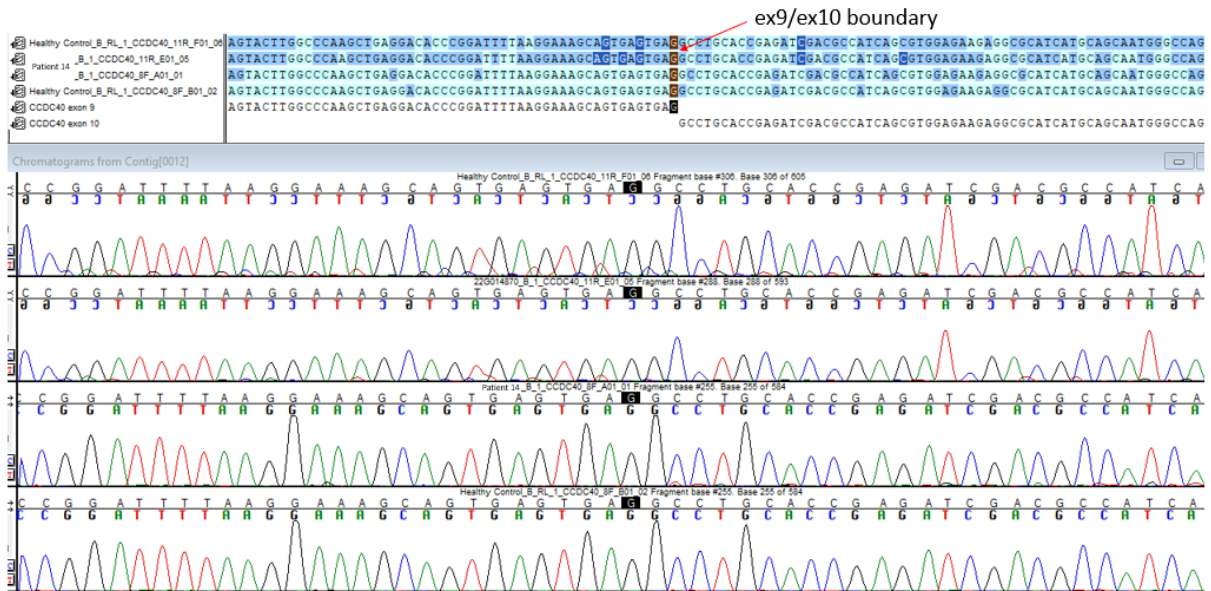
Patient 19 is a 6-year-old boy, with low nasal nitric oxide levels, completely immotile cilia, and inner dynein arm defect and microtubular disorganisation detected on electron microscopy. Following an initial referral for genetic molecular testing, patient 19 was found to carry a heterozygous pathogenic splice site variant *CCDC39* c.357+1G>C. This variant is predicted to abolish an essential splice site leading to loss of function.

A custom research panel of 8 whole genes (coding and non-coding regions) associated with PCD, then identified an exon 9 *CCDC39* deletion. This deletion has not been detected in control populations and a deletion of this exon has been reported once previously by another clinical laboratory in association with PCD. This result was then confirmed using ddPCR (Fig.24).
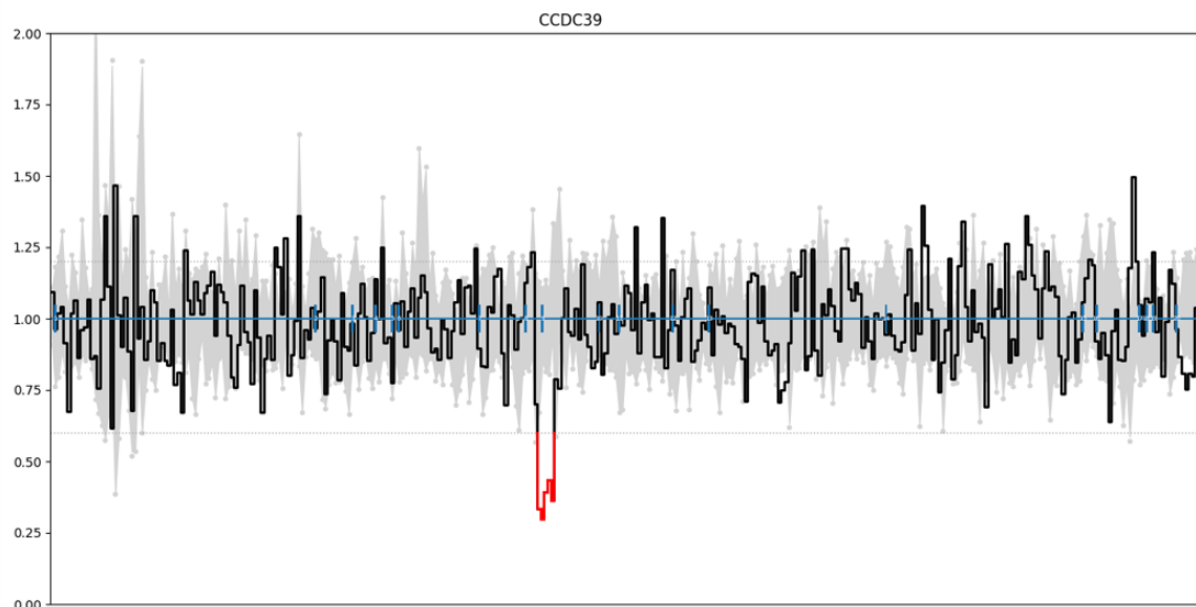
Figure 23. Graph depicting relative copy number of exons and introns of CCDC39 in patient 19 identified with whole gene sequencing and bioinformatic analysis. Blue rectangles are exons and blue lines are non-coding sequences. Grey areas are other copy numbers of other samples run alongside these samples. Red lines show significant copy number variation. A heterozygous deletion of exon 9 is shown.

Figure 24. ddPCR confirmation assay of CCDC39 exon 9 deletion on patient 19. The assay measures copy number variation in a single exon within the deleted region (exon 9) compared to a gene known to have a normal copy number (RRP30) and to a negative. The ratio for a normal copy number in this region was approximately 0.99 and for a deletion present on the patient sample was 0.447.

2.8.5 Patient 21 – 1678bp deletion on chromosome 16

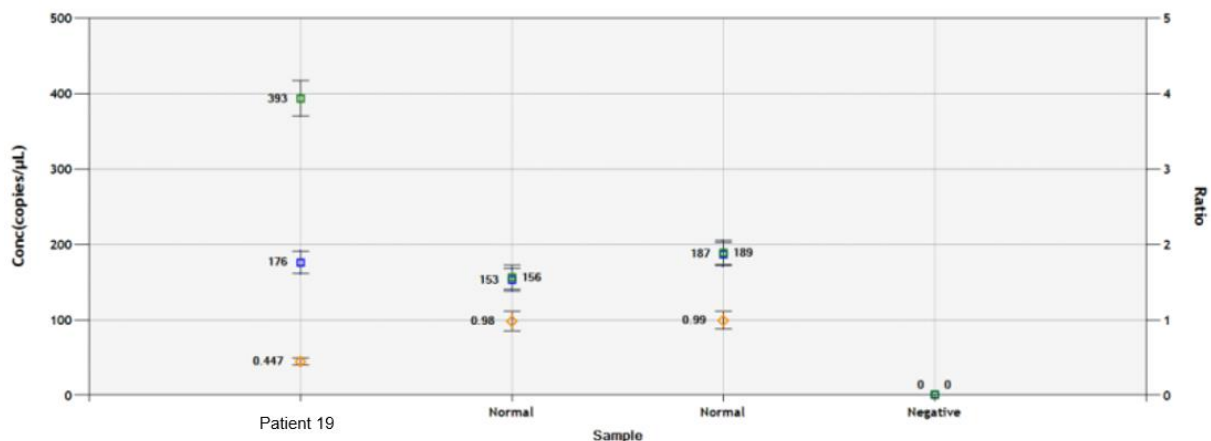Patient 21 is a 44-year-old male, with bronchiectasis, extremely low nasal nitric oxide levels, slow and in some instances almost static and dyskinetic cilia on light microscopy and presents outer dynein arm defects on electron microscopy. Furthermore, patient 21 also has a sister previously diagnosed with PCD. Following an initial referral for genetic molecular testing, patient 21 was found to carry a heterozygous likely pathogenic novel frameshift variant *HYDIN* c. 2419_2422del p.(Val807IlefsTer13). This variant is predicted to cause premature protein truncation, leading to loss of function.

Long read sequencing was then performed on patient 21 sample Oxford Nanopore Technology by a US collaborator of the CGGL. A likely pathogenic 1678bp deletion of chromosome 16 encompassing exon 37 of *HYDIN* was identified. This result was then confirmed using ddPCR (Fig. 25).
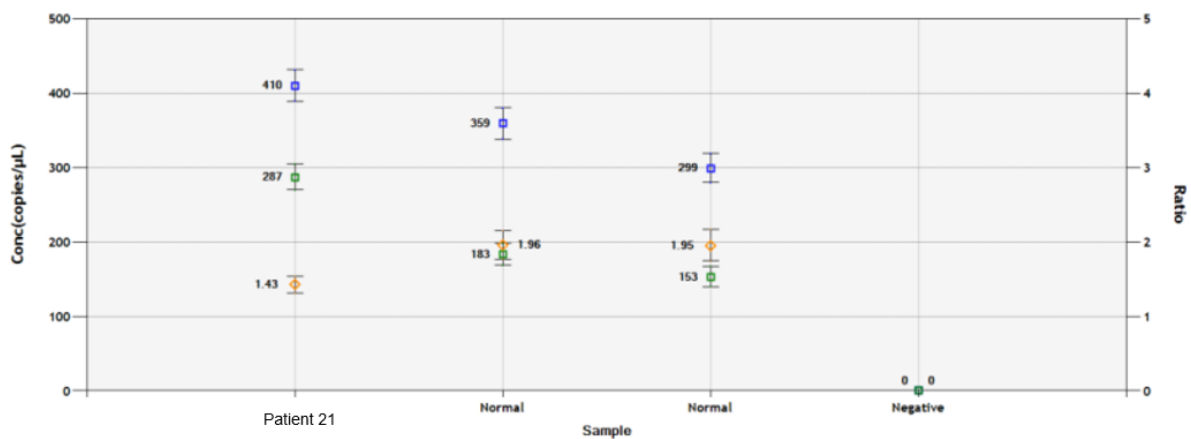
Figure 25. ddPCR confirmation assay of HYDIN exon 37 deletion on patient 21. The assay measures copy number variation in a single exon within the deleted region (exon 9) compared to a gene known to have a normal copy number (RRP30) and to a negative. The ratio for a normal copy number in this region was approximately 1.96 and for a deletion present on the patient sample was 1.43. Ratio values for HYDIN are different than the usual values due to the existence of the pseudogene HYDIN 2.

2.8.6 Other intronic variants highlighted

Patient 8 presented with extremely low nNO levels, normal ultrastructure on EM and a mixture of static or twitching and hyper frequent cilia on light microscopy. Following, whole-gene sequencing this patient was found to be heterozygous for a *DNAH11* c.8511-7T>G. This variant is present at a very low frequency on control populations. Alamut visual predicts a possible effect at the nearest splice site and Splice AI predicts an acceptor loss with a delta score of 0.3, 7bp downstream of the variant.

Patient 12, presented with low nNO levels, normal ultrastructure on EM and a mixture of dyskinetic or jittery cilia with some static strips on light microscopy. *DNAH11* staining was also absent on IF. Following, whole-gene sequencing this patient was found to be heterozygous for a *DNAH11* c.8155-29A>G variant. This variant is absent on control populations. Alamut visual predicts a possible effect at the nearest splice site and Splice AI predicts an acceptor loss with a delta score of 0.21, 29bp downstream of the variant. Both *DNAH11* intronic variants found on patients 8 and 12 are predicted to weaken or compete with the original splice site (on exon 52 and 50 respectively) and therefore altering splicing.

Patient 15 presented with extremely low nNO levels, ODA defects on electron microscopy and static cilia on light microscopy. Following whole gene sequencing this patient was found to be heterozygous for an intronic variant (*DNAI1* c.513+85G>A). This variant has been encountered in population controls and is predicted by Splice AI to activate a donor splice site 46 bp upstream of the variant (Δ score = 0.26) and an acceptor splice site 11bp downstream of the variant (Δ score = 0.22).

2.9 Discussion of Results

At present, nearly one-third of the patients who demonstrate symptoms and clinical features suggestive of PCD, remain without a confirmed genetic diagnosis. PCD genetic diagnosis requires that two pathogenic/ likely pathogenic variants must be found in a single known PCD gene. However, certain patients such as the ones included in this research, have been found to contain only a single heterozygous pathogenic or likely pathogenic variant.

Currently, targeted sequencing is the preferred method for the molecular diagnosis of PCD. However, this current focus on only the coding region of the PCD-causative genes may lead to an oversight of copy number variations or single nucleotide variants located in non-coding regions. By taking a "whole gene sequencing" approach using a custom next-generation sequencing panel including the coding and non-coding regions of 17 known PCD-genes, we aimed to investigate whether there were any undetected potentially pathogenic intronic variants or copy number variations.

This research project was able to identify previously unreported pathogenic variants in 4 out of the 23 patients. Additionally, another 5 patients were identified with possible pathogenic intronic variants, although complementary molecular testing (RNA analysis) is required. These findings support the initial hypothesis, that a second variant harboured in the non-coding region may be causative of PCD.

2.9.1 Detection of deep intronic variants

Over the past decades the search for disease causing variants has been focusing exclusively on the coding genome. However, the functional and therefore disease relevant of the rest of the genome is often overlooked (Spielmann and Mundlos, 2016). With advancements in technology and reduction of costs, the performance of genome-wide association studies and whole-genome sequencing have become a possibility. These have revealed thousands of sequence variants associated with different human disorders. A large majority of the variants identified are present outside the coding sequences and some of them have been shown to perturb binding sites of transcription factors, local chromatin structure or co-factor recruitment, ultimately resulting in changes of transcriptional output of the target genes (Gao et al, 2018).

Genetic and functional studies have shown that non-coding regions have an impact on conditions such as neurodegenerative brain diseases (Frysdas et al, 2021); psychiatric disorders (Xiao, Chang and Li, 2017) and cystic fibrosis (Vecchio-Pagán et al, 2016). Recently,

non-coding variants have also been shown to play an important role in primary cliary dyskinesia pathology (Ellingford et al, 2018).

Intronic sequences account for a significant portion of the human genome and contain essential splicing codes such as cis-acting elements, namely exonic/intronic splicing enhancers/silencers and trans-acting factors such as serine arginine-rich proteins and heterogenous nuclear ribonucleoproteins, transcription, chromatin modification and RNA secondary structure (Tang et al, 2020). These elements are recognised by the spliceosome which is mediated by donor and acceptor splice sites located at exon-intron junctions, to promote a precise excision of introns (Jung, Lee and Choi, 2021).

RNA splicing is a fundamental post-transcriptional mechanism which generates multiple functional RNAs from a single transcript. Throughout this mechanism, introns are removed from pre-mRNA in order to generate mature mRNA ready to be translated (Hoogenhof, Pinto and Creemers, 2016). Pseudoexons are nonfunctional intronic sequences which can be activated by aberrant splicing leading to its inclusion in mRNA, which in turn may cause disease (Metherell et al, 2001).

Conventional sequencing analysis is restricted to exons and exon-intron boundaries. Therefore, deep intronic variants such as the one found on patient 4 (*DNAH11* c.6042-511G>T), would never be reported. Deep intronic variants may result in altered gene expression either through cryptic splicing or disruption of transcription regulatory motifs (Vaz-Drago, Custódio and Carmo-Fonseca, 2017).

Patient 4 was found to be heterozygous for a nonsense *DNAH11* c.3020T>G variant in exon 16 and a deep intronic variant on intron 35 highlighted by *in silico* splicing prediction tools in *DNAH11* c.6042-511G>T. New splice sites created by intronic variants can decrease the specificity or fidelity of splice site selection or activate cryptic acceptor/donor splice sites that are normally not used and lead to the inclusion of pseudoexons in mature RNA (Almeida et al, 2017). cDNA sequencing of nasal epithelial cells of patient 4 has shown the insertion of a 177bp pseudoexon, therefore proving the impact of this deep intronic variant at the transcriptome level.

Although, the *DNAH11* c.6042-511G>T variant was highlighted by 3 prediction tools on Alamut Visual, however the variant did not meet the recommended 0.2 threshold by Splice AI. Splice AI scores can be attenuated by aberrant transcripts impacted by nonsense mediated decay. Additionally, smaller effect sizes result in lower-scoring cryptic splice variants across both gain and loss events (Jaganathan et al, 2019). Low splicing levels may have led to the

pseudoexon not being fully expressed. This is turn may have resulted in the low delta score predicted by Splice Ai.

Following whole gene sequencing, patients 1 and 4 were also found to contain a heterozygous deep intronic variant in *CCDC40*: c.1441-919G>A. This variant is predicted to introduce a 23bp insertion of a pseudoexon. The impact of this variant at the transcriptome level was not confirmed since cDNA sequencing of individual 14, showed that both the patient and healthy control demonstrated equal nucleotide peaks on both reverse and forward sequencing. Although, the additional pseudoexon was represented in the size of cDNA amplification, its products were too weak to be detected by Sanger sequencing. Further experiments are required to be performed to confirm this. Techniques such as ddPCR which presents a higher sensitivity may be conducted instead of Sanger sequencing.

Patient 1 initial genetic testing identified a pathogenic variant on *CCDC40* c.1989+1G>A, whereas patient 14 genetic testing detected a single nonsense heterozygous variant on *CCDC40* c.248del. Considered nasal nitric oxide levels, beat pattern, ultrastructural abnormalities and *in silico* prediction tools it would be expected that the intronic variant (*CCDC40*: c.1441-919G>A) would have an impact at the transcriptome level. However, this could be explained by technical manual errors. Due to the proximity of the expected amplicons (only a 23bp difference), cut out of the visualised bands proved itself difficult.

The use of NuSieve™ 3:1 agarose gel instead of a 1% agarose gel may be an effective strategy to overcome the issues of cutting out the visualised bands. Since NuSieve gel yields strong gels for fine resolution of small DNA, RNA and PCR products ≤ 1 kb, it may be a suitable approach to obtain accurate sequencing products and with less background noise. Furthermore, primers to cover all of the *CCDC40* gene have also been designed in an attempt to demonstrate the effect of the deep intronic variant at the transcriptome level and to provide the patients with a genetic diagnosis.

Pathogenic variants in the *CCDC40* gene are associated with misplacement of the central pair of microtubules and defective assembly of inner dynein arms and dynein regulatory complexes (Becker-Heck et al, 2011). Moreover, PCD caused by pathogenic variants in *CCDC40* is associated with worse pulmonary function, significantly more lobes of consolidation and impaired growth parameters in comparison to patients with PCD due to other genes (Ghandourah and Dell, 2018).

Management of PCD symptoms focuses on delaying disease and improving patients' quality of life and its strategy is based on expert consensus, the clinician's personal experience and data extrapolated from other respiratory conditions (Goutaki and Shoemark, 2022).

Nevertheless, achieving a genetic diagnosis can be of extreme importance not only to guide clinical management, elucidate disease progression and prognosis but also to identify suitable gene therapy and clinical trials that may be available in the near future.

Patients 8 and 12 presented with *DNAH11* intronic variants predicted to weaken or compete with the original splice site (on exon 52 and 50 respectively). In spite of the *in silico* splice predictions and the ultrastructural and clinical findings being corroborative of a diagnosis of *DNAH11* as the PCD- causative gene, cDNA sequencing of the patients would be required to confirm changes at the transcriptome level.

Patient 15 presented with an *DNAI1* intronic variant. The aforementioned variant is suspected to lead to the activation of the donor and acceptor splice sites on intron 6 and is conjectured to result on a 57bp pseudoexon insertion. However, as previously mentioned for patients 8 and 12, cDNA sequencing of the patient would be required to confirm changes at the transcriptome level.

Throughout this research project obtaining RNA from nasal epithelial cells has proven extremely difficult, not only due to the age of the patients, the fact that some patients were unwilling to undertake the procedure but also due to the quality of the sample. PCD affected individuals often present with mucus retention. RNases present on the mucus may promote degradation of RNA present into smaller components and therefore impede RNA extraction (Kuang et al, 2018). Extraction of RNA from blood samples in PAX tubes may be feasible by taking an approach using more sensitive techniques. Moreover, the use of nested-PCR may also be attempted to amplify genes such as *DNAH11* which are present in low abundance.

Pathogenic variants can have diverse impacts on RNA processing, such as exon skipping, cryptic splicing, intron inclusion, leaky splicing or less frequently the inclusion of pseudoexons into mature mRNA as previously mentioned (Caminsky, Mucaki and Rogan, 2015). The introduction of whole gene/ genome sequencing and cDNA/transcriptome sequencing will become a key diagnostic tool for uncovering mutations undetected by targeted panel and whole exome sequencing. Moreover, it will also play an important aspect on deciphering the impact that pathogenic variants have on splice effects and/or length of the transcript.

2.9.2 Splice AI predicts the splice effect of a synonymous variant

Synonymous single nucleotide variants (SSNVs) are a product of the degeneracy of the genetic code, where the same amino acid may be encoded by more than one codon. The effects of SSNVs on molecular functionality, were once thought to be irrelevant. However, earlier studies have argued that synonymous variants are as likely to be pathogenic as non-synonymous variants (Zeng and Bromberg, 2019). SSNVs can disrupt transcription factor regulation, splicing regulation, co-translational folding, protein synthesis, mRNA stability, and cause a plethora of other functionally relevant changes (Shi et al, 2019).

Patient 13 was originally found to possess a heterozygous pathogenic exon 32 deletion in *DNAH11*. A second synonymous heterozygous variant in *DNAH11*: c.1152T>A was then highlighted by Splice AI, following an initial classification as a likely benign variant by Alamut visual and the in-house clinical diagnostics bioinformatics pipeline. This variant was therefore disregarded as likely benign. Furthermore, multiple lines of computational suggested no impact on gene or gene product and another reputable source had recently reported the evidence as likely benign, although considered the date and the laboratory who reported it, it most likely refers to the same patient.

Taken into account the phenotype, cilia ultrastructure and the absence of *DNAH11* staining on IF, which fitted extremely well with a *DNAH11* variant has a causative source, splice predictions by Splice AI were further explored. The synonymous variant was given a 55% probability of being responsible for the creation of a new splice donor gain 3 base pairs upstream of the intronic variant. Subsequential, cDNA sequencing confirmed protein truncation and a loss of 45bp due to the synonymous variant. At present, the patient is undergoing further molecular genetic testing alongside her parents and her brother in order to determine whether the two variants were inherited in *cis* or *trans*. If the variants are proven to have been inherited in *trans,* then a molecular genetic diagnosis of PCD can be confirmed.

2.9.3 Improved CNV detection in whole gene NGS

Advances in CNV calling have enabled coding and non-coding variants and CNVs to be identified using NGS technologies. An in-house CNV-calling pipeline compared the average read depth of regions with those of other samples and neighbouring regions within the same sample. In this project, CNVs that were not detected in initial diagnostic testing, were identified using the whole gene sequencing panel. The TWIST Bioscience technology was used whereas diagnostic sequencing uses the Agilent SureSelect QXT library preparation. The use

of a technology with a more uniform coverage may account for the reason as to why previous testing was unable to detect the CNV.

The genomics field has been revolutionised in the last few years due to the wide implementation of NGS technologies. However, despite short-read sequencing being cost-effective, accurate and supported by a wide range of analysis tools and pipelines it also presents with its limitations (Amarasinghe et al, 2020). Short read lengths pose a limitation for the identification of structural variants, sequencing repetitive regions, phasing of alleles and distinguishing highly homologous genomic regions (Mantere, Kersten and Hoischen, 2019).

The introduction of long-read sequencing technologies has been shown to reduce sequencing time, to reduce or eliminate sequencing biases introduced by PCR amplification and to increase the read length from tens of bases to tens of thousands of bases per read (Lu, Giordano and Ning, 2016). Therefore, long-read sequencing can be used to identify variants such as the one found on patient 21, which with short-read sequencing would not be identified either due to errors in calling variant calling or due to the inability to capture certain genomic regions.

2.9.4 Patients with no relevant variants identified

Despite the extension of sequencing to the non-coding regions in PCD-related genes, this research project was unable to identify significant relevant variants in 14 out of the 23 patients enrolled in this research project. Some of the variants present on these patients were present at high frequencies in control population databases, however since PCD is a recessive condition this does not entirely exclude the probability of their pathogenicity. Furthermore, some of the variants highlighted by *in silico* splicing predictions did not meet the recommended threshold however, only RNA analysis would be able to completely exclude their impact on splicing. Lastly, patients with *HYDIN* variants became incredibly difficult to assess due to the extreme large number of variants highlighted by the bioinformatics pipeline and the existence of the pseudogene *HYDIN2.*

Although significant progress has been achieved regarding the mechanisms and genes linked to PCD, the current rate at which new ciliopathy genes are identified suggests that many remain undiscovered (Dam et al, 2019). With dozens of new potential PCD-causative genes, it may be speculated that patients without relevant variants may attribute their phenotype to PCD-causative genes who have yet to be identified.

2.10 Conclusion

The results of this project support the relevance and usefulness of the inclusion of the assessment of non-coding variants in the clinical setting, in the future. At present, the implementation of WGS in the diagnostic setting would prove itself inefficacious and extremely challenging, due to the large amount of data generated and the substantial number of variants of unknown significance. However, research projects such as this one and the exchange of variant data between diagnostic laboratories may get us a step closer to implement WGS and perhaps provide a genetic diagnosis to patients such as the ones included in this study, who would not be able to acquire a diagnosis if an approach beyond the coding regions was not pursued.

Furthermore, a complete genetic diagnosis will allow carrier testing in family members, guide family planning and enable the inclusion of the patient in future relevant gene therapy procedures.

References

1.Abboud, R., Ford, G., and Chapman, K. (2005). Emphysema in alpha1-antitrypsin deficiency: does replacement therapy affect outcome? Treat Respir Med, 4, 1-8

2.Abel, H., and Duncavage, E. (2013). Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. Cancer genetics, 206(12), 432-440

3.Abramowicz, A., and Gos, M. (2018). Splicing mutations in human genetic disorders: examples, detection and confirmation. Journal of applied genetics, 59,253—268

4.Afzelius, B. (1976). A human syndrome caused by immotile cilia. Science, 193(4250), 317-319

5.Almeida, R., et al. (2017). Whole gene sequencing identifies deep-intronic variants with potential functional impact in patients with hypertrophic cardiomyopathy. Plus one, 12(8), e0182946. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5552324/ [Accessed 8th June 2022]

6.Al-Turkmani, M., Deharvengt, S., and Lefferst, J. (2020). Molecular assessment of human diseases in the clinical laboratory. Essential concepts in molecular pathology, 2nd ed: Academic Press, 563-578

7.Amarasinghe, S., et al. (2020). Opportunities and challenges in long-read sequencing data analysis. Genome Biology, 21, 30

8.Andelfinger, G., Loyes, B., and Dietz, H. (2015). A decade of discovery in the genetic understanding of thoracic aortic disease. Canadian Journal of Cardiology, 32, 13-25

9.Austin, E., and Loyd, J. (2014). The genetics of pulmonary arterial hypertension. Cir Res, 115, 189-202

10.Barbato, A., et al. (2009). Primary ciliary dyskinesia: a consensus statement on diagnostic and treatment approaches in children, Eur Respir J, 34(6), 1264-1276

11.Barbitoff, Y., et al. (2020). Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage. Scientific Reports, 10, 2057

12.Baz-Redón, N., et al. (2019). Role of immunofluorescence and molecular diagnosis in the characterization of primary ciliary dyskinesia. Scientific letters/ Arch Bronconeumol, 55(8), 436-449

13.Baz-Redón, N., et al. (2020). Immunofluorescence analysis as a diagnostic tool in a Spanish cohort of patients with suspected primary ciliary dyskinesia. Journal of Clinical Medicine,9, 3603

14.Becker- Heck, A., et al. (2011). The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis formation. Nature genetics, 43, 79-84

15.Behan, L., et al. (2016). PICADAR: a diagnostic predictive tool for primary ciliary dyskinesia. European Respiratory Journal, 16, 1103-1112

16.Behan, L., et al. (2017). Validation of a health-related quality of life instrument for primary ciliary dyskinesia (QOL-PCD). Thorax, 72, 832-839

17.Belhassan, K., and Granadillo, J. (2021). Current approaches to genetic testing in pediatric disease. Biochemical and molecular basis of pediatric disease, 5th ed, Academic Press, 15-36

18.Bewicke-Copley, F., et al. (2019). Applications and analysis of targeted genomic sequencing in cancer studies. Computational and structural biotechnology journal, 17, 1348-1359

19. Bhatt, R., and Hogg, C. (2020). Primary ciliary Dyskinesia: a major player in a bigger game. Breathe, 16, 200047

20. Bio-Rad. (2015). ddPCR copy number variation assays. Available from: https://www.researchgate.net/profile/Yuan-Yeu-Yau/post/Droplet-digital-PCR-ddPCR/attachment/59d6496379197b80779a3e73/AS%3A470810663100421%401489261417697/download/ddPCR+for+Copy+Number+Variation+Assays.pdf [Accessed 23rd May 2022]

21. Bio-Rad. (2022). Droplet digital PCR (ddPCR) technology. Available from: https://www.bio-rad.com/en-uk/life-science/learning-center/introduction-to-digital-pcr/what-is-droplet-digital-pcr?ID=MDV31M4VY [Accessed 23rd May 2022]

22. Bio-Rad. (2022). Planning droplet digital PCR experiments. Available from: https://www.bio-rad.com/en-uk/life-science/learning-center/introduction-to-digital-pcr/planning-ddpcr-experiments?ID=MDV33OKG4 [Accessed 23rd May 2022]

23. Blanch, B., et al. (2017). Routinely collected health data to study inherited heart disease: a systematic review (2000-2016). Open Heart, 4, e000686 Available from: doi: 10.1136/openhrt-2017-000686 [Accessed 19th April 2022]

24. Boaretto, F., et al. (2016). Diagnosis of primary ciliary dyskinesia by a targeted next-generation sequencing panel: molecular and clinical findings in Italian patients. The Journal of Molecular Diagnostics, 18(6), 912-922

25. Börekçi, Ş., and Müsellim, B. (2021). Decreasing rate of unknown bronchiectasis etiology: evaluation of 319 adult patients with bronchiectasis. Turkish Thoracic Journal, 22, 18-23

26. Boycott, K., et al. (2015). The clinical application of genome-wide sequencing for monogenic diseases in Canada: position statement of the Canadian college of medical genetics. J Med Genet, 52, 431-437

27. Brennan, S., Ferkol, T., and Davis, S. (2021). Emerging genotype-phenotype relationships in primary ciliary dyskinesia. International Journal of Molecular Sciences, 22, 8272

28. Brieler, J., Breeden, M., and Tucker, J. (2017). Cardiomyopathy: an overview. Am Fam Physician, 96(10), 640-646

29. Buermans, H., and Dunnen, J. (2014). Next generation sequencing technology: advances and applications. Biochimica et biophysica acta (BBA) – molecular basis of disease, 1842(10), 1932-1941

30. Bush, A., et al. (2007). Primary ciliary dyskinesia: current state of the art. Arch Dis Child, 92, 1136-1140

31. Cadrin- Tourigny, J., and Tadros, R. (2022). Predicting sudden cardiac death in genetic heart disease. Canadian Journal of Cardiology, 38(4), 479-490

32. Caminsky, N., Mucaki, E., and Rogan, P. (2015). Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. F1000 Research, 3, 282

33. Castiblanco, J. (2013). A primer on current and common sequencing technologies. Autoimmunity: from bench to bedside. Bogota: El Rosario University Press

34.Ceulemans, S., Van der Vem, K., and Del-Favero. (2011). Targeted screening and validation of copy number variations. Genomic structural variants. Methods in Molecular biology (methods and protocols). Springer, New York, vol 838, 311-328

35.Chilvers, M., Rutman, A., and O'Callaghan, C. (2003). Ciliary beat pattern is associated with specific ultrastructural defects in primary ciliary dyskinesia. J Allergy Clin Immunol, 112(3), 518-524

36.Chopra, M., Tendolkar, M., and Vardhan, V. (2019). Case series of pulmonary alveolar microlithiasis from India. BMJ Case Reports, 12, 227406 Available from: http://orcid.org/0000-0002-8324-8348 [Accessed 3rd May 3, 2022]

37.Cinetto, F., et al. (2018). The broad spectrum of lung diseases in primary antibody deficiencies. European respiratory review, 27(149), 1-17

38.Cirino, A., et al. (2017). Role of genetic testing in inherited cardiovascular disease – A review. JAMA Cardiology, 2(10), 1153-1160

39.ClinVar. (2022). National library of medicine – national center for biotechnology information. Available from: https://www.ncbi.nlm.nih.gov/clinvar/ [Accessed 1st July 2022]

40.Coll, M., et al. (2018). Incomplete penetrance and variable expressivity: Hallmarks in channelopathies associated with sudden cardiac death. Biology (Basel), 7, 3. Available from: 10.3390/biology7010003 [Accessed 20th April 2022]

41.Collins, S., et al. (2014). Nasal nitric oxide screening for primary ciliary dyskinesia: systematic review and meta-analysis. Eur respir J, 44, 1589-1599

42.Collins, S., Walker, W., and Lucas, J. (2014). Genetic testing in the diagnosis of primary ciliary dyskinesia: state-of-the-art and future perspectives. Journal of Clinical Medicine, 3(2), 491-503

43.Dam, T., et al. (2019). CiliaCarta: An integrated and validated compendium of ciliary gene. PlosOne, 14(5), e0216705. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0216705 [Accessed 13th June 2022]

44.Damseh, N., et al. (2017). Primary ciliary dyskinesia: mechanisms and management. The application of clinical genetics, 10, 67-74

45.Davis, S., et al. (2015). Clinical features of childhood primary ciliary dyskinesia by genotype and ultrastructural phenotype. American Journal of respiratory and critical care medicine, 191 (3), 316-324

46.Davis, S., et al. (2018). Primary ciliary dyskinesia: longitudinal study of lung disease by ultrastructure defect and genotype. American Journal of respiratory and critical care medicine, 199(2), 190-198

47.Deharvengt, S., et al. (2020). Nucleic acid analysis in the clinical laboratory. Contemporary practice in clinical chemistry, 4th ed, Academic Press, 215-234

48.Derbyshire, E., and Calder, P. (2021). Bronchiectasis – could immunonutrition have a role to play in future management? Frontiers in Nutrition, 98, 652410

49.Devine, M., and Garcia, C. (2012). Genetic interstitial lung disease. Clin Chest Med, 33, 95-110

50.Di Resta, C., et al. (2018). Next-generation sequencing approach for the diagnosis of human diseases: open challenges and new opportunities. EJIFCC, 29, 4-14

51.Dixon, M., and Shoemark, A. (2017). Secondary defects detected by transmission electron microscopy in primary ciliary dyskinesia diagnostics. Ultrastructural Pathology, 41(6), 390-398

52. Djakow, J., et al. (2016). An effective combination of sanger and next generation sequencing in diagnostics of primary ciliary dyskinesia. Pediatric Pulmonology, 51, 498-509

53. Dutcher, S., and Brody, S. (2020). HY-DIN' in the cilia: Discovery of central-pair related mutations in primary ciliary dyskinesia. American Journal of Respiratory Cell and Molecular Biology, 62(3), 281-282

54. Eijk-Van Os, P., and Schouten, J. (2011). Multiplex ligation-dependent probe amplification (MLPA) for the detection of copy number variation on genomic sequences. Methods Mol Biol, 688, 97-126

55. Eksi, D., et al. (2022). Novel gene variants associated with primary ciliary dyskinesia. Indian Journal of Pediatrics. Available from: https://doi.org/10.1007/s12098-022-04098-z [Accessed 14th April 2022]

56. Ellard, S., et al. (2020). ACGS best practice guidelines for variant classification in rare disease. ACGS. Available from: https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf [Accessed 1st July 2022]

57. Ellingford, J., et al. (2018). Whole genome sequencing enables definitive diagnosis of cystic fibrosis and primary ciliary dyskinesia. bioRxiv Available from: https://doi.org/10.1101/438838 [Accessed 7th April 2020]

58. European Lung. (2022). Genetic susceptibility. European Lung white book. Available from: https://www.erswhitebook.org/chapters/genetic-susceptibility/ [Accessed 28th April 2022]

59. Fassad, M., et al. (2020). Clinical utility of NGS diagnosis and disease stratification in a multiethnic ciliary dyskinesia cohort. J Med Genet, 57, 322-330

60. Fitzgerald, J., et al. (2020). A deep intronic variant activates a pseudoexon in the MTM1 gene in a family with X-linked myotubular myopathy. Molecular syndromology, 11, 264-270

61. Fletcher, A., et al. (2020). Inherited thoracic aortic disease – new insights and translational targets. Circulation, 141(19), 1570-1587

62. Fliegauf, M., et al. (2005). Mislocalization of DNAH5 and DNAH9 in respiratory cells from patients with primary ciliary dyskinesia. Am J Respir Crit Care Med, 171(12), 1343-1349

63. Flume, P., Chalmers, J., and Olivier, K. (2018). Advances in bronchiectasis: endotyping, genetics, microbiome and disease heterogeneity. Lancet, 392 (10150), 880-890

64. Focșa, I., Budișteanu, M., and Bălgrădean, M. (2021). Clinical and genetic heterogeneity of primary ciliopathies (review). International Journal of Molecular Medicine, 48(3), 1107-3756

65. Franklin, W., et al. (2014). Pathology, biomarkers, and molecular diagnostics. Abeloff's clinical oncology, 5th ed. Churchill Livingstone, 226-252

66. Fretzayas, A., and Moustaki, M. (2016). Clinical spectrum of primary ciliary dyskinesia in childhood. World Journal of Clinical Pediatrics, 5, 57-62

67. Frydas, A., et al. (2021). Uncovering the impact of noncoding variants in neurodegenerative brain disease. Trends in Genetics, 38, 3, 258-272

68. Fu, D. Cardiac arrhythmias: diagnosis, symptoms and treatments. Cell Biochem Biophys, 73, 291-296

69. Gao, L., et al. (2018). Identifying noncoding risk variants using disease-relevant gene regulatory networks. Nature communications, 9, 702

70. Garcia, C. (2004). Inherited interstitial lung disease. Clinics in chest medicine, 25, 421-433

71. Ghandourah, H., and Dell, S. (2018). Severe disease due to CCDC40 gene variants and the perils of late diagnosis in primary ciliary dyskinesia. BMJ Case Reports. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6144183/ [Accessed 12th June 2022]

72. Gileles-Hillel, A., et al. (2020). Whole-exome sequencing accuracy in the diagnosis of primary ciliary dyskinesia. ERJ Open Res, 6(4)

73. Girolami, F., et al. (2018). Contemporary genetic testing in inherited cardiac disease tools, ethical issues, and clinical applications. Journal of Cardiovascular Medicine, 19, 1-11

74. gnomAD Browser. (2022). gnomAD – Genome Aggregation Database. Available from: https://gnomad.broadinstitute.org/ [Accessed 1st July 2022]

75. Goel, K., et al. (2021). Updates in the management of alpha-1 antitrypsin deficiency lung disease. Touch respiratory, 6, 26-30

76. Gomes, A., and Korf, B. (2018). Genetic testing techniques. Pediatric cancer genetics, 47-64

77. Goodwin, S., McPherson, J., and McCombie, R. (2016). Applications of next-generation sequencing – coming of age: ten years of next-generation sequencing technologies. Nature reviews genetics, 17, 333-351

78. Gould, C., Freeman, A., and Olivier, K. (2012). Genetic causes of bronchiectasis. Clin Chest Med, 33, 249-263

79. Goutaki, M., and Shoemark, A. (2022). Diagnosis of primary ciliary dyskinesia. Clinics in Chest Medicine, 43, 127-140

80. Gravesande, K., and Omran, H. (2005). Primary ciliary dyskinesia: clinical presentation, diagnosis, and genetics. Annals of Medicine, 37, 439-449

81. Greenstone, M., et al. (1984). Hydrocephalus and primary ciliary dyskinesia. Archives of Disease in Childhood, 59(5), 481-482

82. Grumbt, B., et al. (2013). Diagnostic applications of next generation sequencing in immunogenetics and molecular oncology. Transfusion Medicine and Hemotherapy, 40(3), 196-206

83. Gui, L., et al. (2019). Scaffold subunits support associated subunit assembly in the Chlamydomonas ciliary nexin-dynein regulatory complex. PNAS, 116 (46), 23152-23162

84. Gui, M., et al. (2021). Structures of radial spokes and associated complexes important for ciliary motility. Nature structural and molecular biology, 28, 29-37

85. Guo, T., et al. (2017). An effective combination of whole-exome sequencing and runs of homozygosity for the diagnosis of primary ciliary dyskinesia in consanguineous families. Scientific reports, 7, 7905

86. Gupta, N., and Verma, V. (2019). Next-generation sequencing and its application: empowering in public health beyond reality. Microbial technology for the Welfare of Society. Microorganisms for sustainability, vol 17, Springer, Singapore

87. Haworth, A., Savage, H., and Lench, N. (2016). Diagnostic Genomics and clinical bioinformatics. Medical and health genomics: Academic Press, 37-50

88. Heather, J., and Chain B. (2016). The sequence of sequencers: the history of sequencing DNA. Genomics, 107, 1-8

89.HGMD. (2022). The human gene mutation database. Available from: http://www.hgmd.cf.ac.uk/ac/index.php [Accessed 1st July 2022]

90.Hildebrandt, F., Benzing, T., and Katsanis, N. (2013). Ciliopathies. The New England Journal of Medicine, 364(16), 1533-1543

91.Hogg, C., and Bush, A. (2021). CON: Primary ciliary dyskinesia diagnosis: genes are all you need! Paediatric Respiratory Reviews, 37, 34-36

92.Holkeri, A., et al. (2020). Predicting sudden cardiac death in a general population using an electrocardiographic risk score.Heart, 106, 427-433

93.Hoogenhof, M., Pinto, Y., and Creemers, E. (2016). RNA Splicing, Regulation and Dysregulation in the Heart. Circulation Research, 118(3), 454-468

94.Horani, A., and Ferkol, T. (2021). Understanding primary ciliary dyskinesia and other ciliopathies. J Pediatr, 230, 15-22

95.Horani, A., Brody, S., and Ferkol, T. (2014). Picking up speed: advances in the genetics of primary ciliary dyskinesia. Pediatric research, 75, 158-164

96.Horani, A., et al. (2012). Whole-exome capture and sequencing identifies HEATR2 mutation as a cause of primary ciliary dyskinesia. The American Journal of Human Genetics, 91, 685-693

97.Horani, A., et al. (2015). Genetics and biology of primary dyskinesia. Paediatric respiratory reviews, 18, 18-24

98.Hyland, R., and Brody, S. (2022). Impact of motile ciliopathies on human development and clinical consequences in the newborn. Cells, 11, 125

99.Hyland, R., and Brody, S. (2022). Impact of motile ciliopathies on human development and clinical consequences in the newborn. Cells, 11, 125 [Image]

100.Ingles, J., et al. (2020). Genetic testing in inherited heart diseases. Heart, Lung and Circulation, 29, 505-511

101. International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. Nature, 409, 860-921

102.Ishige, T., Itoga, S., and Matsushita, K. (2018). Locked nucleic acid technology for highly sensitive detection of somatic mutations in cancer. Advances in clinical chemistry, 83, 53-72

103.Jackson, C., et al. (2016). Accuracy of diagnostic testing in primary ciliary dyskinesia. Eur Respir J, 47, 837-848

104.Jaganathan, K., et al. (2019). Predicting splicing from primary sequence with deep learning. Cell, 176, 535-548

105.Jaganathan, K., et al. (2019). Predicting splicing from primary sequence with deep learning. Cell, 176(3), 535-548

106.Jamuar, S., D'Gama, A., and Walsh, C. (2016). Somatic mosaicism and neurological diseases. Genomics, circuits and pathways in clinical neuropsychiatry. Academic Press, 179-199

107.Jayasena, C., and Sironen, A. (2021). Diagnostics and Management of Male Infertility in Primary Ciliary Dyskinesia, 11(9), 1550

108.Jung, H., Lee, K., and Choi, J. (2021). Comprehensive characterization of intronic mis-splicing mutations in human cancers. Oncogene, 40, 1347-1361

109. Kanzi, A., et al. (2020). Next generation sequencing and bioinformatics analysis of family genetic inheritance. Frontiers in genetics, 11, 544162

110. Kartagener, M. (1933). Zur pathogenese der bronkiectasien. Bronkiectasien bei situs viscerum inversus. *Beitr Klin Tuberk Spezif Tuberkuloseforsch*. 1933, (83), 489–501

111. Kelly, M., and Semsarian, C. (2009). Multiple mutations in genetic cardiovascular disease – A marker of disease severity? Circulation: Cardiovascular Genetics, 2(2), 182-190

112. Kennedy, M., et al. (2007). Congenital Heart Disease and Other Heterotaxic Defects in a Large Cohort of Patients With Primary Ciliary Dyskinesia. Circulation, 115(22), 2814-2821

113. Kerkhof, J., et al. (2017). Clinical validation of copy number variant detection from targeted next-generation sequencing panels. J Mol Diagn, 19, 905-920

114. Khosla, A., et al. (2016). Systemic vasculopathies: imaging and management. Radiologic clinics in North America, 54(3), 613-628

115. Kim, R., et al. (2014). The role of molecular genetic analysis in the diagnosis of primary ciliary dyskinesia. Ann Am Throrac Soc, 11(3), 351-359

116. Kim, S., and Dynlacht, B. (2013). Assembling a primary cilium. Current opinion in cell biology, 25(4), 506-511

117. Knowles, M., et al. (2013). Primary Ciliary Dyskinesia. Recent Advances in Diagnostics, Genetics, and Characterization of Clinical Disease. Am J Respir Crit Care Med, 188(8), 913-922

118. Knowles, M., et al. (2014). Mutations in RSPH1 cause primary ciliary dyskinesia with a unique clinical and ciliary phenotype. Am J Respir Crit Care Med, 189(6), 707–717

119. Knowles, M., Zariwala, M., and Leigh, M. (2016). Primary Ciliary Dyskinesia. Clin Chest Med, 37(3), 449-461

120. Kojabad, A., et al. (2021). Droplet digital PCR of viral DNA/RNA, current progress, challenges, and future perspectives. Journal of Medical Virology, 93(7), 4182-4197

121. Kreicher, K., et al. (2018). Hearing loss in children with primary ciliary dyskinesia. International Journal of Pediatric Otorhinolaryngology, 104, 161-165

122. Kristof, A., et al. (2017). An official American thoracic society wokshop report: translational research in rare respiratory diseases. Ann Am Thorac Soc, 14(8), 1239-1247

123. Kuang, J., et al. (2018). An overview of technical considerations when using quantitative real-time PCR analysis of gene expression in human exercise research. PloS One, 13(5), e0196438 Available from : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5944930/ [Accessed 17th June 2022]

124. Kuehni, C., and Lucas, J. (2017). Diagnosis of primary ciliary dyskinesia: summary of the ERS Task Force Report. Breathe, 13, 166-178

125. Kuehni, C., et al. (2010). Factors influencing age at diagnosis of primary ciliary dyskinesia in European children. European Respiratory Journal, 36, 1248-1258

126. Kuhlmann, K., Cieselski, M., and Schumann, J. (2021). Relative versus absolute RNA quantification: a comparative analysis based on the example of endothelial expressin of vasoactive receptors. Biological Procedures Online, 23,6 Available from: https://biologicalproceduresonline.biomedcentral.com/articles/10.1186/s12575-021-00144-w#ref-CR4 [Accessed 20th May 2022]

127.Kurkowiak, M., Ziętkiewicz, W., and Witt, M. (2009). Recent advances in primary ciliary dyskinesia genetics. J Med Genet, 52, 1-9

128.Lee, C., et al. (2013). Common applications of next-generation sequencing technologies in genomic research. Translational cancer research, 2, 33-45

129.Lee, J., and Gleeson, J. (2011). A systems-biology approach to understanding the ciliopathy disorders. Genome Medicine, 3, 59

130.Lee, L., and Ostrowski, L. (2021). Motile cilia genetics and cell biology: big results from little mice. Cellular and molecular life sciences, 78(3), 769-797

131.Legendre, M., Zaragosi, L., and Mitchison, H. (2021). Motile cilia and airway disease. Seminars in Cell and Developmental Biology, 110, 19-33 [Images]

132.Leigh, M., et al. (2009). Clinical and genetic aspects of primary ciliary dyskinesia/ Kartagener syndrome. Genetics in Medicine, 11(7), 473-487

133.Leigh, M., et al. (2013). Clinical and Genetic Aspects of Primary Ciliary Dyskinesia/ Kartagener Syndrome. Genet Med, 11(7), 473-487

134.Leigh, M., et al. (2019). Primary Ciliary Dyskinesia (PCD): A genetic disorder of motile cilia. Transl Sci Rare Dis, 4(1-2), 51-75

135.Leigh, M., O'Callaghan, C., and Knowles, M. (2011). The challenges of diagnosing primary ciliary dyskinesia. Proceedings of the American Thoracic Society, 8(5), 434-437

136.Li, Q., et al. (2021). Unraveling synonymous and deep intronic variants causing aberrant splicing in two genetically undiagnosed epilepsy families. BMC Medical Genomics,14, 152

137.Li, Y., et al. (2016). DNAH6 and its interactions with PCD genes in heterotaxy and primary ciliary dyskinesia. PLOS GENETICS, e1005821

138.Lindeman, N., Fletcher, J., and Longtine, J. (2021). Application of modern techniques. Diagnostic histopathology of tumors, 5th ed. Elsevier, 2288-2310

139.Liners, J., Médart, L., and Collignon, L. (2019). Primary pulmonary hypoplasia. J Belg Soc Radiol, 103, 7

140.Lobo, J., Zairwala, M., and Noone, P. (2015). Primary Ciliary Dyskinesia. Semin Respir Crit Care Med, 36(2), 169-179

141.Lu, H., Giordano, F., and Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics, Proteomics & Bioinformatics, 14(5), 265-279

142.Lucas, J., and Leigh, M. (2014). Diagnosis of primary ciliary dyskinesia: searching for a gold standard. European Respiratory Journal, 44, 1418-1422

143.Lucas, J., et al. (2014). Diagnosis and management of primary ciliary dyskinesia. Archives of disease in childhood,99(9),850-856

144.Lucas, J., et al. (2016). Diagnostic methods in primary ciliary dyskinesia. Paediatric Respiratory Reviews, 18, 8-17

145.Lucas, J., et al. (2017). European Respiratory Society Guidelines for the diagnosis of primary ciliary dyskinesia. European Respiratory Journal, 49, 1601090

146.Lucas, J., et al. (2020). Primary ciliary dyskinesia in the genomics age. The Lancet Respiratory Medicine, 8(2), 202-216

147.Lucas, J., et al. (2020). Primary ciliary dyskinesia in the genomics age. Lancet Respir Med, 8(2), 202-216

148.Macfarlane, L., et al. (2021). Diagnosis and management of non-cystic fibrosis bronchiectasis. Clinical Medical Journal, 21(6), 571-577

149.Mandras, S., Mehta, H., and Vaidya, A. (2020). Pulmonary hypertension: a brief guide for clinicians. Mayo Clin Proc, 95(9), 1978-1988

150.Mantere, T., Kersten, S., and Hoischen, A. (2019). Long-read sequencing emerging in medical genetics. Frontiers in Genetics, 10, 426

151.Mardis, E. (2008). Next-generation sequencing platforms. Annu. Rev. Genomics Hum. Genet., 9, 387-402 [Image]

152.Mardis, E. (2013). Next-generation sequencing platforms. Annu. Rev. Anal. Chem., 6, 287-303

153.Maron, B., et al. (2006). Contemporary definitions and classification of the cardiomyopathies. Circulation, 113(14), 1807-1816

154.Marshall, C., et al. (2015). Whole-exome sequencing and targeted copy number analysis in primary ciliary dyskinesia. G3 (Bethesda), 5(8), 1775-1781

155.McKenna, W., and Elliott, P. (2020). Diseases of the myocardium and endocardium. Goldman-Cecil Medicine, 26th ed, 54, 297-314

156.McLoud, T., and Boiselle, P. (2010). Chronic obstructive pulmonary disease and asthma. Thoracic Radiology 2nd ed, 242-252

157.Metherell, L., et al. (2001). Pseudoexon activation as a novel mechanism for disease resulting in atypical growth-hormone insensitivity. American Journal of Human Genetics, 69(3), 641-646

158.Mianné, J., et al. (2018). Induced pluripotent stem cells for primary ciliary dyskinesia modelling and personalized medicine. American Journal of Respiratory Cell and Molecular Biology, 59(6), 672-683 [Image]

159.Mirra, V., Werner, C., and Santamaria, F. (2017). Primary ciliary dyskinesia: An update on clinical aspects, genetics, diagnosis, and future treatment strategies. Frontiers in pediatrics,5, 135

160.Mitchinson, H., and Valente, E. (2017). Motile and non-motile cilia in human pathology: from function to phenotypes. Journal of Pathology, 241, 294-309

161.Modarage, K., Malik, S., and Goggolidou, P. (2022). Molecular diagnostics of ciliopathies and insights into novel developments in diagnosing rare diseases. British Journal of Biomedical Science, 79,10221

162.Montavon, T., Thevenet, L., and Duboule, D. (2012). Impact of copy number variations (CNVs) on long-range gene regulation at the HoxD locus. PNAS, 109(50), 20204-20211

163.MRC Holland. (2022). SALSA MLPA Available from: https://www.mrcholland.com/technology [Accessed April 2022]

164.Mullowney, T., et al. (2014). Primary Ciliary Dyskinesia and Neonatal Respiratory Distress. Pediatrics, 134(6), 1160-1166

165.Murphy-Ryan, M., Psychogios, A., and Lindor, N. (2010). Hereditary disorders of connective tissue: a guide to the emerging differential diagnosis. Genetics in Medicine, 12(6), 344-354

166. Musunuru, K., et al. (2020). Genetic testing for inherited cardiovascular diseases: A scientific statement from the American Heart Association. Circulation: Genomic and Precision Medicine, 13(4), 373-385 [Table]

167. Narayan, D., et al. (1994). Unusual inheritance of primary ciliary dyskinesia (Kartagener's syndrome). Journal of Medical Genetics, 31(6), 493-496

168. Nikiforova, M., and Nikiforov, Y. (2019). Molecular anatomic pathology: principles, techniques, and application to immunohistologic diagnosis. Diagnostic Immunohistochemistry, 5th ed. Elsevier, 47-62

169. Noone, P., et al. (1999). Discordant organ laterality in monozygotic twins with primary ciliary dyskinesia. Am J Med Genet, 82(2), 155-160

170. O' Connor, M., Horani, A., and Shapiro, A. (2021). Progress in diagnosing primary ciliary dyskinesia: the North America perspective. Diagnostics, 11, 1278

171. Olmedillas-López, S., García-Arranz, M., and García-Olmo, D. (2017). Current and emerging applications of droplet digital PCR in oncology. Molecular Diagnosis & Therapy, 21, 493-510

172. Omran, H., et al. (2000). 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. American Journal of Respiratory Cell and Molecular Biology, 23(5)

173. Ostrowski, L, Dutcher, S., and Lo, C. (2011). Cilia and models for studying structure and function. Proceedings of the American Thoracic Society, 8, 423-429

174. Otto, C., Savla, J., and Hisama, F. (2020). Cardiogenetics: a primer for the clinical cardiologist. Heart, 106, 938-947

175. Patir, A., et al. (2020). The transcriptional signature associated with human motile cilia. Scientific Reports, 10, 10814 Available from: https://doi.org/10.1038/s41598-020-66453-4 [Accessed March 29, 2022]

176. Pereira, R., Oliveira, J., and Sousa, M. (2020). Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. J. Clin. Med, 9, 132

177. Popatia, R., Haver, K., and Casey, A. (2014). Primary Ciliary Dyskinesia: An Update on New Diagnostic Modalities and Review of the Literature. Pediatric Allergy, Immunology and Pulmonology, 27(2), 51-59

178. Pös, O., et al. (2021). DNA copy number variation: main characteristics, evolutionary significance, and pathological aspects. Biomedical Journal, 44(5), 548-559

179. Primorac, D., et al. (2021). Sudden cardiac death – a new insight into potentially fatal genetic markers. Front Med, 8, 647412

180. Prulière-Escabasse, V., Coste, A., and Chauvin, P. (2010). Otologic Features in Children With Primary Ciliary Dyskinesia. Arch Otolaryngol Head Neck Surg, 136(11), 1121-1126

181. Pua, C., et al. (2016). Development of a comprehensive sequencing assay for inherited cardiac condition genes. Journal of Cardiovascular Translational Research, 9, 3-11

182. Qian, X., et al. (2021). Identification of deep intronic splice mutations in a large cohort of patients with inherited retinal diseases. Frontiers in genetics, 12, 647400

183. Qin, D. (2019). Next- generation sequencing and its clinical application. Cancer biology & medicine, 16, 4-10

184.Raidt, J. et al. (2014). Ciliary beat pattern and frequency in genetic variants of primary ciliary dyskinesia. Eur Respir J, 44, 1579-1588

185. RB&HH. (2021). History Available from: https://www.rbhh-specialistcare.co.uk/about-us/history [Accessed 15 September 2021]

186.Reeskamp, L., et al. (2018). A deep intronic variant in LDLR in familial hypercholesterolemia. Circulation: Genomic and Precision Medicine, 11(12)

187.Reiter, J., and Leroux, M. (2017). Genes and molecular pathways underpinning ciliopathies. Nature reviews molecular cell biology, 18, 533-547

188.Renner, S., et al. (2019). Next-generation sequencing of 32 genes associated with hereditary aortopathies and related disorders of connective tissue in a cohort of 199 patients. Genetics in Medicine, 21, 1832-1841

189.Rosenfeld, M., Ostrowski, L. and Zariwala, M. (2018). Primary ciliary dyskinesia: keep it on your radar. Thorax, 73(2),101-102

190.Roy, S., et al. (2018). Standards and guidelines for validating next-generation sequencing bioinformatics pipelines: a joint recommendation of the association for molecular pathology and the college of American pathologists. The journal of molecular diagnostics, 20, 4-27

191.Royal Brompton and Harefield hospitals. (2021). Clinical Genetics and Genomics Laboratory Available from: https://www.rbht.nhs.uk/our-services/clinical_support/laboratories/clinical-genetics-and-genomics-laboratory [Accessed 15 September 2021

192.Royal Brompton and Harefield hospitals. (2021). Royal Brompton Hospital Available from: https://www.rbht.nhs.uk/our-hospitals/royal-brompton-hospital [Accessed 15 September 2021]

193.Rubbo, B., and Lucas, J. (2017). Clinical care for primary ciliary dyskinesia: current challenges and future directions. European Respiratory Review, 26, 170023

194.Rubbo, B., et al. (2019). Accuracy of high-speed video analysis to diagnose primary ciliary dyskinesia. CHEST,18, 2329

195.Rubbo, B., et al. (2020). Clinical features and management of children with primary ciliary dyskinesia in England. Arch Dis Child, 105, 724-729

196.Sakamoto, K., et al. (2021). Ependymal ciliary motion and their role in congenital hydrocephalus. Child's Nervous System, 37, 3355-3364

197.Sanger, F., Nicklen, S., and Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci, 74(12), 5463-5467

198.Schoentgen, F. (2020). Insight on the role of RKIP in cancer through key protein partners and cellular protusions. Prognostic and Therapeutic Applications of RKIP in Cancer: Academic Press, 3-35

199.Schwartz, P., et al. (2020). Inherited cardiac arrhythmias. Nature Reviews Disease Primers, 6, 58

200.Shaikh, T. (2018). Copy number variation disorders. Curr Genet Med Rep, 5(4), 183-190

201.Shapiro, A., et al. (2015). Diagnosis, monitoring, and treatment of primary ciliary dyskinesia: PCD foundation consensus recommendations based on state of the art review, 51(2), 115-132

202.Shapiro, A., et al. (2018). Diagnosis of Primary Ciliary Dyskinesia; An Official American Thoracic Society Clinical Practice Guideline. American Journal of Respiratory and Critical Care Medicine, 197(12), 24-39

203.Shapiro, A., et al. (2020). High-speed video microscopy Analysis presents limitations in diagnosis of primary ciliary dyskinesia. Am J Respir Crit Care Med, 201, 122-123

204.Shapiro, A., et al. (2020). Limitations of nasal nitric oxide testing in primary ciliary dyskinesia. Am J Respir Crit Care Med,202(3), 476-477

205.Shi, F., et al. (2019). Computational identification of deleterious synonymous variants in human genomes using a feature-based approach. BMC Medical Genomics, 12, 12

206.Shoemark, A., et al (2017). Accuracy of immunofluorescence in the diagnosis of primary ciliary dyskinesia. American Journal of Respiratory and Critical Care Medicine, 196, 94-101

207.Shoemark, A., et al. (2020). International consensus guideline for reporting transmission electron microscopy results in the diagnosis of primary ciliary dyskinesia (BEAT PCD TEM criteria). Eur Respir J. Available from: https://doi.org/10.1183/13993003.00725-201 [Accessed 26th March 2022]

208.Shukla, S., et al. (2020). Chronic respiratory diseases: An introduction and need for novel drug delivery approaches. Targetting chronic inflammatory lung diseases using advanced drug delivery systems, 1-31. Available from: 10.1016/B978-0-12-820658-4.00001-7 [Accessed 24th April 2022]

209.Simonneau, G., et al. (2013). Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol, 62(25), 34-41

210.Smith, E., and Yang, P. (2004). The radial spokes and central apparatus: mechano-chemical transducers that regulate flagellar motility. Cell motility and the cytoskeleton, 57, 8-17

211.Soler-Palacín, P., et al. (2018). Primary immunodeficiency diseases in lung disease: warning signs, diagnosis and management. Respiratory Research, 19, 219

212.Spielmann, M., and Mundlos, S. (2016). Looking beyond the genes: the role of non-coding variants in human disease. Human Molecular Genetics, 25(2), 157-165

213.Stewart, J., Manmathan, G., and Wilkinson, P. (2017). Primary prevention of cardiovascular disease: a review of contemporary guidance and literature. Available from: 10.1177/2048004016687211 [Accessed 18th April 2022]

214.Stratton, M. (2008). Genome resequencing and genetic variation. Nature biotechnology, 26, 65-66

215.Stuppia, L., et al. (2012). Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. Int J Mol Sci, 13(3), 3245-3276

216.Suwinski, P., et al. (2019). Advancing personalized medicine through the application of whole exome sequencing and big data analytics. Frontiers in Genetics,10, 49

217.Tang, S., et al. (2020). Cis- and trans-regulations of pre-mRNA splicing by RNA editing enzymes influence cancer development. Nature communications, 11, 799

218.Townsend, N., et al. (2022). Epidemiology of cardiovascular disease in Europe. Nature reviews cardiology, 19, 133-143

219.Van den Heuvel, L., et al. (2019). Informing relatives at risk of inherited cardiac conditions: experiences and attitudes of healthcare professionals and counselees. European Journal of Human Genetics, 27, 1341-1350

220.Vanaken, G., et al. (2017). Infertility in an adult cohort with primary ciliary dyskinesis: phenotype-genotype association. European Respiratory Journal, 50, 1700314

221.Varghese, M. (2014). Familial hypercholesterolemia: a review. Ann Pediatr Cardiol, 7(2), 107-117

222.Vaz-Drago, R., Custódio, N., and Carmo-Fonseca, M. (2017). Deep intronic mutations and human disease. Hum Genet, 136, 1093-1111

223.Vecchio, F., et al. (2017). Next-generation sequencing: recent applications to the analysis of colorectal cancer. Journal of translational medicine, 15, 246

224.Vecchio-Pagán, B., et al. (2016). Deep resequencing of CFTR in 762 F508del homozygotes reveals clusters of non-coding variants associated with cystic fibrosis disease traits

225.Vessies, D., et al. (2021). An automated correction algorithm (ALPACA) for ddPCR data using adaptive limit of blank and correction of false positive events improves specificity of mutation detection. Clinical Chemistry, 67(7), 959-967

226.Villamil, C., et al. (2020). Validation of droplet digital polymerase chain reaction for salmonella spp. Quantification. Front. Microbiol, 11, 1512

227.Vrablik, M., et al. (2020). Genetics of familial hypercholesterolemia: new insights. Frontiers in Genetics, 11, 574474

228.Walker, W., et al. (2012). Nitric oxide in primary ciliary dyskinesia. Eur Respir J, 40, 1024-1032

229.Walker, W., et al. (2012). Nitric oxide in primary ciliary dyskinesia. European Respiratory Journal, 40, 1024-1032

230.Walker, W., et al. (2013). Upper and lower airway nitric oxide levels in primary ciliary dyskinesia, cystic fibrosis and asthma. Respiratory medicine, 107(3), 380-386

231.Werner, C., Onnebrink, J., and Omran, H. (2015). Diagnosis and management of primary ciliary dyskinesia, 4(2)

232.Wheway, G., et al. (2021). Whole genome sequencing in the diagnosis of primary ciliary dyskinesia. BMC Medical Genomics, 14, 234

233.Wilde, A., and Behr, E. (2013). Genetic testing for inherited cardiac disease. Nature Reviews Cardiology, 10, 571-583

234.Willis, A., Van den Veyver, I., and Eng, C. (2012). Multiplex ligation-dependent probe amplification (MLPA) and prenatal diagnosis. Prenatal diagnosis, 32(4), 315-320

235.Wirschell, M., et al. (2011). Regulation of ciliary motility: conserved protein kinases and phosphatases are targeted and anchored in the ciliary axoneme. Archives of biochemistry and biophysics, 510(2), 93-100

236.Wong, A., Ryerson, C., and Guler, S. (2020). Progression of fibrosing interstitial lung disease. Respiratory Research, 21, 32

237.Xiao, X., Chang, H., and Li, M. (2017). Molecular mechanisms underlying noncoding risk variations in psychiatric genetic studies. Molecular psychiatry, 22, 497-577

238.Xuan, J., et al. (2013). Next-generation sequencing in the clinic: promises and challenges. Cancer letters, 340(2), 284-295

239.Yao, Y., and Shen, K. (2017). Monogenic diseases in respiratory medicine: clinical perspectives. Pediatric Investigation, 1, 27-31

240.Yonker, L., et al. (2020). Recognizing genetic disease: a key aspect of pediatric pulmonary care. Pediatr Pulmonol, 55(7), 1794-1809

241. Zariwala, M., Omran, H., and Ferkol, T. (2011). The emerging genetics of primary ciliary dyskinesia. Proceedings of the American Thoracic Society, 8(5), 430-433

242. Zeng, Z., and Bromberg, Y. (2019). Predicting functional effects of synonymous variants: a systemic review and perspectives. Frontiers in Genetics, 10, 914 Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6791167/ [Accessed 10th June 2022]

243. Zhang, F., et al. (2009). Copy number variation in human health, disease, and evolution. Annual review of genomics and human genetics, 10, 451-481

244. Zhang, P., and Fernandes, S. (2014). Other post-PCR detection technologies. Pathobiology of human disease: A dynamic encyclopedia of disease mechanisms. Elsevier, The Netherlards, 4074-4088 [Image]

245. Zhao, M., et al. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics, 14, S1

246. Zhao, X., et al. (2021). Clinical characteristics and genetic spectrum of 26 individuals of Chinese origin with primary ciliary dyskinesia. Orphanet Journal of Rare Diseases, 16, 293