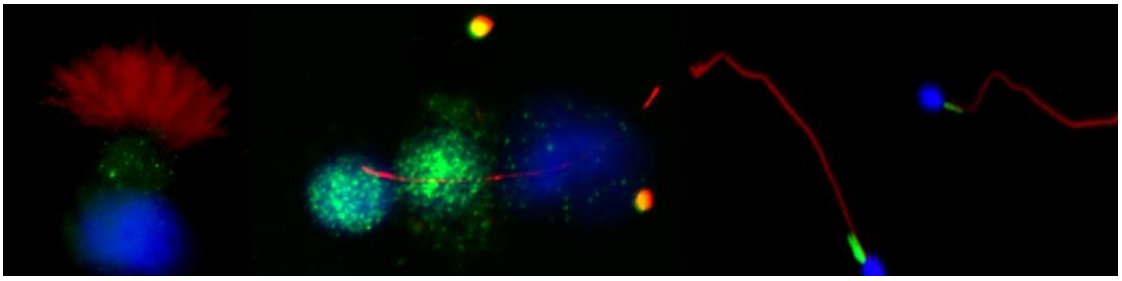
UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR TÉCNICO



The genetics of Primary Ciliary Dyskinesia:
from the flagellum to the cilium

Rute Ribeiro Pereira

Supervisor Doctor Mário Manuel da Silva Leite de Sousa

Co-Supervisor Doctor Joaquim Manuel Sampaio Cabral

Thesis approved in public session to obtain the PhD Degree in Bioengineering

Jury final classification: Pass with Distinction

2019

UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR TÉCNICO

The genetics of Primary Ciliary Dyskinesia:
from the flagellum to the cilium

Rute Ribeiro Pereira

Supervisor Doctor Mário Manuel da Silva Leite de Sousa

Co-Supervisor Doctor Joaquim Manuel Sampaio Cabral

Thesis approved in public session to obtain the PhD Degree in Bioengineering

Jury final classification: Pass with Distinction

Jury

Chairperson: Doctor Duarte Miguel de França Teixeira dos Prazeres, Instituto Superior Técnico, Universidade de Lisboa

Members of the Committee:

Doctor Mário Manuel da Silva Leite de Sousa, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto.
Doctor Alberto António Moreira Caldas Afonso, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto
Doctor Maria de Lourdes Gomes Pereira, Universidade de Aveiro
Doctor Gabriel António Amaro Monteiro, Instituto Superior Técnico, Universidade de Lisboa
Doctor Cláudia Alexandra Martins Lobato da Silva, Instituto Superior Técnico, Universidade de Lisboa

Funding Institutions

Fundação para a Ciência e a Tecnologia, I. P. (FCT)

Grant PD/BD/105767/2014

2019

Resumo

A discinesia ciliar primária (PCD) é uma doença rara autossómica recessiva cujas alterações génicas cursam com anomalias do axonema, o motor ciliar. O axonema é um cilindro composto por 9 dupletos de microtúbulos periféricos, que incluem dois braços de dineína (DA), que se encontram ligados entre si pelo complexo regulador das dineínas (DRC) e conectados a um par central de microtúbulos por projeções radiais. Os pacientes com PCD apresentam infeções respiratórias crónicas, podendo associar-se (em cerca de 50% dos casos) a *situs-inversus-totalis* (clinicamente designado por síndrome de Kartagener (KS)) e à infertilidade.

A presente tese de doutoramento teve como principal questão perceber que variantes genéticas conduzem à KS e ausência dos DA no axonema. Para tal, pacientes com essas características foram selecionados e realizada a completa do exoma (WES). A segunda questão aqui levantada, foi de que modo as variantes encontradas afetam a expressão génica e proteica, no sentido de melhor compreender a relação genótipo-fenótipo.

Esta tese iniciou-se com a análise genética por sequenciação de Sanger de pacientes com total imotilidade dos espermatozoides. Porém, num dos pacientes, que apresentava KS e ausência dos DA no axonema dos espermatozoides, o estudo genético foi inconclusivo, tendo-se procedido à sua avaliação por WES, de que resultou na identificação de uma nova variante no gene *CCDC103*. Este gene foi descrito como um fator importante para o "*assemble*" dos DA, mas encontra-se pouco caracterizado nos humanos. Assim, neste trabalho, foi evidenciado que a estratégia de sequenciação gene-a-gene não é efetiva para a PCD e despoletou interesse para o estudo mais detalhado dos perfis de expressão do gene *CCDC103*. Assim, 2 pacientes com KS e portadores de variantes patogénicas neste gene foram estudados. Os resultados sugerem que essas variantes causam ausência dos DA, e afetam a expressão do gene e da proteína. Adicionalmente, verificamos que o gene *CCDC103* é expresso em diferentes tipos de células, sendo que os espermatozoides apresentam uma maior expressão do gene, e a proteína uma localização celular distinta. A presença de isoformas proteicas observada,

sugeriu modificações pós-translacionais, sugerindo que CCDC103 pode desempenhar uma função regulatória.

Seguidamente, o exoma de 3 crianças, com anomalias nos DA e com KS, foi analisado por WES e realizada a correlação genótipo-fenótipo nas crianças e respetivas famílias. Numa das crianças encontraram-se variantes patogénicas em dois genes, o *DNAH5* e o *DNAH7.* Observou-se no paciente a redução da expressão do mRNA e uma alteração localização das proteínas em ambos os genes. Noutra criança, o WES não foi suficiente para deteção de variantes patogénicas, tendo-se procedido à análise do número de cópias. Encontraram-se 3 novas variações potencialmente causadoras da doença em genes não relacionados previamente com a PCD. Deste trabalho, dois aspetos principais devem ser destacados. Primeiramente, salientamos a possibilidade de um mecanismo de interação entre dois genes (*DNAH5* e *DNAH7*) na produção do fenótipo. Seguidamente, os dados permitiram-nos salientar as limitações do WES.

Na última parte desta tese, estudos preliminares, salientam a importância, da aplicação de ferramentas bioinformáticas, nomeadamente as regiões em homozigotia a partir de dados de WES. Estas permitem adjuvar na filtragem das variantes e assim a melhorar o diagnostico.

De modo geral, os resultados aqui obtidos contribuem para melhorar os conhecimentos sobre os fatores genéticos envolvidos na patofisiologia da PCD/KS e imotilidade do flagelo. Estes são de extrema importância para auxiliar o diagnóstico genético e no desenvolvimento de futuras terapias.

Palavras-Chave:

Axonema; cílios móveis; Discinesia ciliar primária; sequenciação de exoma; síndrome de Kartagener.

Abstract

Primary ciliary dyskinesia (PCD) is a rare ciliopathy caused by dysfunction of motile cilia. The axoneme, the main motor of cilia, is composed of 9 peripheral microtubule doublets, each possessing two dynein arms (DA). The doublets are linked by the dynein regulatory complex (DRC) and connected to a single pair of central microtubules, by radial spokes (RS). Defects in motile cilia impairing its proper function, cause ineffective mucociliary clearance, organ laterality defects (known as *situs inversus*) and infertility. Kartagener's syndrome (KS) occurs in half of PCD patients and is characterized by a triad of *situs inversus*, chronic sinusitis, and bronchiectasis.

The main aim of the present doctoral thesis was to study the genetic background of patients presenting KS with an absence of DA in its axoneme. To achieve this goal, KS patients with DA defects were selected and whole exome sequencing (WES) was performed. The second question raised here was how the variants found affect gene and protein expression, to better understand the genotype-phenotype correlation.

This thesis initiated with the genetic analysis by Sanger's sequencing of patients with total sperm immotility. In one of the patients, who had KS and absence of DA in the sperm axoneme, the genetic study was inconclusive and was evaluated by WES, which resulted in the discovery of a new variant in the *CCDC103* gene, that codes for a small protein shown to be essential for DA assembly, but which is weakly characterized in humans. Thus, it was evidenced that the gene-to-gene sequencing strategy is not effective for PCD and prompted interest to study *CCDC103* gene expression profiles. Consequently, 2 patients with KS and carriers of pathogenic variants in this gene were studied. The results suggest that those variants cause the absence of DA and affect gene and protein expression. In addition, we found that the *CCDC103* gene is expressed in different cell types, with the spermatozoa having higher expression of the gene, and a different cellular location of the protein. The presence of protein isoforms observed, suggested post-translational modifications, suggesting that CCDC103 may play a regulatory role.

Subsequently, the exome of 3 children, with abnormalities in DA and KS, was analysed by WES and performed genotype-phenotype in children and their families. In one of the children, pathogenic variants were observed in two genes, *DNAH5* and *DNAH7*. Reduction of mRNA expression and a mislocalization of the proteins in both genes were observed. In another child, the typical analysis by WES do not found pathogenic variants. Further analyses by different approaches (including copy number variation, (CNV)) finally revealed possible novel candidates for PCD. From this work, two aspects should be highlighted. Firstly, the possibility of a mechanism of interaction between two genes (*DNAH5* and *DNAH7*) in the production of the phenotype is highlighted. From this analysis two main conclusions are highlighted. Firstly, we hypothesized a model in which multiple variants in distinct genes, may have combined effects in PCD phenotype. Second, the data from the other children, emphasis the WES limitations as with the implementation of the typical workflow no disease-causing variants were found. Further analyses by different approaches (including copy number variation, (CNV)) finally revealed possible novel candidates for PCD.

Finally, preliminary data highlighted the importance of application of 'runs' of homozygosity from WES data as an effective tool to assist variant filtering and aiming to improve the diagnostic yield.

Overall, the results obtained during this research thesis contribute to increasing the understanding of the genetic factors involved in pathophysiology of PCD/KS and sperm motility, which is of paramount importance to assist the current diagnosis, and for the development of novel therapies in the future.

Key words: Axoneme, Motile cilia, Primary ciliary dyskinesia Dynein arms; Whole exome sequencing, Kartagener syndrome

Agradecimentos

A realização desta tese só foi possível devido ao apoio incondicional de pessoas que vão ficar para sempre marcadas na minha vida e para as quais dedico algumas palavras.

Em primeiro lugar, quero agradecer ao meu Orientador, o Professor Dr. Mário Sousa, que apoiou as minhas decisões pessoais e académicas e abraçou esta aventura de braços abertos mesmo sabendo que seria uma jornada difícil. Agradeço-lhe por todo apoio, disponibilidade e dedicação que me deu ao longo deste projeto.

Um Muito Obrigada também ao Jorge Oliveira, que foi sem dúvida um elemento fundamental neste projeto, sem as suas orientações, ideias, correções este projeto não teria sido possível. Sei que o seu tempo era escasso, que tinha uma vida já sobrecarregada com os seus projetos pessoais e profissionais, mas ainda assim, conseguiu encontrar tempo para me dedicar, e por isso estou-lhe grata.

Não posso deixar de agradecer ao Prof. Joaquim Cabral, que compreendeu as razões que me levaram a trocar de projeto, assim possibilitando a realização desta tese.

Quero agradecer também à Dra. Rosário Santos, por ter colaborado neste projeto, permitido que fossem utilizadas as instalações e ferramentas da Unidade de Genética Molecular do Centro de Genética Médica Dr. Jacinto de Magalhães/ Centro Hospitalar do Porto, e por toda simpatia e disponibilidade de apoio que sempre me demonstrou. Um Muito obrigado também à Márcia Oliveira, por todo o apoio e ensinamentos que me deu. Quero também agradecer à Ana Rita e Catarina Monteiro por toda ajuda dada nas sequenciações e apoio geral. Estou também grata a toda equipa da Unidade de Genética Molecular que desde do primeiro dia me recebeu de braços abertos. Agradeço a todos por estarem sempre disponíveis, mesmo estando *super* atarefados.

Quero agradecer também toda a equipa do ICBAS. Muito Obrigado Elsa Oliveira e Ângela Alves, por toda a disponibilidade, apoio e simpatia demonstrados e pela grande ajuda particularmente na parte da microscopia. Agradeço também ao Prof. Pedro Oliveira, Prof. Marco Alves e a todo o grupo, pela simpática receção que demostraram aquando do início dos meus trabalhos no ICBAS. Particularmente, agradeço à Tânia Dias, Raquel Bernardino, Susana Almeida e Ana Maria por toda ajuda quer na cedência de amostras, quer na explicação das metodologias e sugestões de melhorias. Obrigado

também à Cláudia, Fernanda e Célia por se demonstrarem sempre prontas a ajudar caso precisasse. Agradeço também ao ICBAS como instituto por ter permitido que realiza-se nas suas instalações os trabalhos conducentes a esta tese de doutoramento.

Com todas estas pessoas que acima indique, aprendi muito não só a nível profissional como também pessoal e por isso um Muito Obrigado.

Quero também agradecer à Dra. Telma Barbosa por todo apoio clínico prestado. Agradeço também aos centros médicos que colaboraram com este trabalho tanto na cedência de amostras como de pacientes particularmente o Centro Materno Infantil do Centro Hospitalar do Porto; Centro Hospitalar de Entre o Douro e Vouga E.P.E.; Centro de Genética da Reprodução Prof. Alberto Barros e Centro Hospitalar de Vila Nova de Gaia/Espinho. Agradeço também aos Pacientes e famílias que sem eles estes trabalhos não seriam possíveis.

Não posso também deixar de agradecer o financiamento, sem o qual este trabalho não poderia ter existido. Particularmente à FCT por me financiar com uma bolsa de doutoramento e pelos apoios dados à UMIB que por sua vez financiou os meus trabalhos.

Aos meus amigos, muito obrigada por estarem sempre lá, por me ouvirem e também pelas sugestões que me deram. Peço-vos desculpa pelas vezes que não estive disponível para vocês, mas sei que me compreendem. Obrigada a todos!

Em último, mas mais importante, **agradeço à minha família**, vocês foram o meu pilar. Obrigado!!!

Ao Hélder, agradeço a paciência, amor e carinho. Desculpa os fins de semana que passei em frente ao computador e não contigo. Obrigado por me ouvires reclamar quando as coisas corriam pior, mesmo não percebendo totalmente o que estava a dizer! Obrigada!

Obrigada ao meu pai, que sempre me apoiou e acarinhou.

Um Muito Obrigada à minha Mãe! Não há palavras para explicar o que sei que fizeste e como estou agradecida por tudo. Obrigada!

Index

List of Abbreviations

Abbreviation	Full name
AD	Autosomal dominant
AR	Autosomal recessive
BAM	Binary Alignment Map
Bp	Base Pair
CBF	Ciliary beat frequency
CFTR	Cystic fibrosis transmembrane receptor
CNV	Copy number variation
CPC	Central pair complex
DA	Dynein arms
DFS	Dysplasia of the fibrous sheath
DRC	Dynein regulatory complex
ERS	European Respiratory Society
HC	Heavy chain
HM	Homozygosity mapping
HMM	Hidden Markov model
HSVMA	High-speed video microscopy analysis
IBD	Identical by descent
IBS	Identical by state
IC	Intermediate chains
IDA	Inner dynein arms
IFT	Intraflagellar Transport
KS	Kartagener syndrome
LC	Light chain
LD	Linkage disequilibrium
LOD	Logarithm of odds
LRD	Left-right dynein
Mbs	Megabase

Abbreviation	Full name
MTBD	Microtubule-binding domain
NGS	Next-Generation sequence
NO	Nitric oxide
ODA	Outer dynein arms
ODA-DC	Outer dynein arm docking complex
PCD	Primary ciliary dyskinesia
PCR	Polymerase chain reaction
ROH	Runs of homozygosity
RS	Radial spokes
SAM	Sequence Alignment Map
SNP	Single nucleotide Polymorphism
SNV	Single nucleotide variants
SV	Structural variants
TEM	Transmission electron microscopy
VCF	Variant call format
VUS	Variant of uncertain significance
WBC	White blood cells
WES	Whole exome sequence
WGS	Whole genome sequence

The aims and outline of the thesis

The main aim of this doctoral thesis was to study the genetic background of patients presenting Kartagener syndrome (KS) with the absence of axoneme dynein arms. KS is a clinical condition characterized by *situs inversus totalis* and primary ciliary dyskinesia (PCD). The search for gene variants was mainly performed using next generation sequencing (NGS) technologies, namely through whole-exome sequencing (WES).

The secondary aim of the present thesis was to analyse the genotype-phenotype correlations in these patients. For this, the pathogenic variants found in these patients were analysed regarding mRNA expression, protein expression and protein cellular localization.

The identification of genetic factors involved in the pathophysiology of PCD and KS, as well as, regarding sperm motility (that is often a consequence of PCD) will be of utmost interest in the future development of newer therapies. The complex journey of human induced pluripotent stem cells differentiation into airway epithelium as well as the efforts to restore protein function using viral vectors begun just recently. Maybe just a few years' time span, hopefully, we will see therapy for PCD and the emerge of novel forms of assisted reproductive technology. Nevertheless, having a genetic cause identified is for now of great value for personalized medicine to assist the diagnosis, which enables the selection of the best current possible treatment or patient care.

The present Thesis is composed of five main chapters, starting with a **general introduction** to cilium structure and function.

In the **second part**, a compilation of the methodologies used during this PhD to achieve the aims referred above.

Follows a **third part** which includes the results obtained during the thesis and its discussion. The results were further divided into 5 five sub-chapters:

The **first sub-chapter** is dedicated to the genetic causes of sperm total immotility research in a group of five infertile patients, including one with KS in which WES analysis was performed.

Subsequently, in **the second sub-chapter**, we studied the genotype-phenotype relationships of the previously characterized male patient and added a further female KS patient, also showing the absence of dynein arms in axoneme of nasal cilia. WES was also performed in the latter revealing a homozygous variant in the same *CCDC103* gene. This gene was described as an important dynein arm assembly factor, with pathogenic variants previously found in other patients with PCD and KS. However, only a few reports concerning this gene have been published, and thus information is reduced. Additionally, variants found in this gene have never been analysed before in terms of gene expression. Therefore, in this sub-chapter we have unfolded the expression profiles of these novel *CCDC103* variants.

Later, in **third and fourth sub-chapters**, it is described the genetic background of three children also with KS and absence of dynein arms in the axoneme of nasal cilia. The third is dedicated to the comprehensive clinical, ultrastructural and genetic study, by WES, of two of these children. In the fourth sub-chapter **4**, we explore the genetic aetiology of the third child. This case is independently presented as after WES analyses it was not possible to find obvious disease-causing variants. Further analyses by different methodologies were applied.

In the last sub-chapter (**fifth sub-chapter**), we show the work performed using bioinformatic methodologies to assist WES analysis.

This thesis ends with a **fourth part** in which the main conclusions and future research perspectives are highlighted.

Articles published under the scope of this thesis:

R. Pereira, M. Sousa, J. Oliveira (2019) Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. Journal of Clinical Medicine (submitted in November 2019)

Pereira, R., Barbosa, T., Alves, Â., Santos, R., Oliveira, J., Sousa, M., 2019. Unveiling the genetic etiology of primary ciliary dyskinesia: when standard genetic approach is not enough. Advances in Medical Sciences. https://doi.org/10.1016/j.advms.2019.10.003

Pereira, R., Barbosa, T., Gales, L., Oliveira, E., Santos, R., Oliveira, J., Sousa, M., 2019. Clinical and Genetic Analysis of Children with Kartagener Syndrome. Cells 8(8). https://doi.org/10.3390/cells8080900

Pereira, R., Oliveira, M.E., Santos, R., Oliveira, E., Barbosa, T., Santos, T., Gonçalves, P., Ferraz, L., Pinto, S., Barros, A., Oliveira, J., Sousa, M., 2019 Characterization of CCDC103 expression profiles: further insights in primary ciliary dyskinesia and in human reproduction. Journal of Assisted Reproduction and Genetics https://doi.org/10.1007/s10815-019-01509-7

Oliveira J*., Pereira R*., Santos R., Sousa M. (2018) Evaluating Runs of Homozygosity in Exome Sequencing Data - Utility in Disease Inheritance Model Selection and Variant Filtering. In: Peixoto N., Silveira M., Ali H., Maciel C., van den Broek E. (eds) Biomedical Engineering Systems and Technologies. BIOSTEC 2017. Communications in Computer and Information Science, vol 881. Springer, Cham. https://doi.org/10.1007/978-3-319-94806-5_15 (*-equally contributing first authors)

Oliveira, J*., Pereira, R*., Santos, R. and Sousa, M. (2017). Homozygosity Mapping using Whole-Exome Sequencing: A Valuable Approach for Pathogenic Variant Identification in Genetic Diseases.In Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 3: BIOINFORMATICS, (BIOSTEC 2017), pages 210-216. https://doi.org/10.5220/0006248502100216 (*-equally contributing first authors)

Pereira R., Sá R., Barros A. and Sousa M. (2017) Major regulatory mechanisms involved in sperm motility. Asian journal of andrology. 19(1):5-14. https://doi.org/10.4103/1008-682X.167716.

Pereira R. & Sousa M. (2015). NGS and Male Infertility: Biomarkers Wanted. Annual Research & Reviews in Biology, 8(5): 1-4, https://doi.org/10.9734/ARRB/2015/20263

Pereira, R., Oliveira, J., Ferraz, L., Barros, A., Santos, R., & Sousa, M. (2015). Mutation analysis in patients with total sperm immotility. Journal of Assisted Reproduction and Genetics, 32(6), 893-902. http://doi.org/10.1007/s10815-015-0474-6

Pereira, R., Oliveira, J., Sousa, M.,(2014). A molecular approach to sperm immotility in humans: A review. Medicina Reproductiva y Embriología Clínica 01, 15-25. https://doi.org/10.1016/S2340-9320(15)30004-9

- New variants added to ClinVar database (see Attachment 1).

Chapter 1

General introduction

Contents

1. Introduction

Rare diseases are life-threatening conditions, often chronic and progressive, that although individually rare, collectively they affect about 65 million individuals worldwide. Most are complex and multisystemic genetic disorders, affecting quality and life expectancy of patients, representing a considerable burden to the affected families and to the society. Most of the rare diseases have no suitable cure, although some innovative orphan drugs are being successfully tested. Nonetheless, upon an early diagnosis and optimally managed from a clinical perspective, the disease prognostics could be largely improved.

Primary ciliary dyskinesia (PCD, MIM #244400) is a rare autosomal recessive ciliopathy that causes multiple respiratory problems, organ laterality defects (as *situs inversus*) and reproductive complications leading to infertility in both women and men. There is no "gold standard" diagnostic for PCD, thus it relies on a combination of several tests. The genetic study is an important final proof of PCD, frequently requiring high-throughput genotyping techniques such as NGS. Despite the unquestionable importance of genetics in PCD, due to its extraordinary genetic complexity and heterogeneity, the knowledge of PCD is still largely incomplete. Consequently, they are some unmet medical needs that require continuous investigations to fulfil the deficit of medical and scientific knowledge.

Cilia was firstly observed by Antony Van Leeuwenhoek in 1675, using a simple mounted lens. Since then, great findings were achieved (Satir, 1995, 2017). The increasing knowledge about the structure and function of cilia and flagella during the past years is the result of continued improvements in microscopy techniques and in the advances and developments of high-throughput techniques, particularly in the improvements of -omics technologies. These are of upmost importance for unveiling the complexity of cilia biology and consequently in the understanding the pathophysiology of PCD. However, the cilia puzzle is far from finished and several pieces are still missing.

1.1. Cilia in health and disease

1.1.1. Cilium types

Cilia are microtubule-based cell organelles that outgrowths from the plasma membrane of the most eukaryotic cells. Cilia are present in most eukaryotic cells and can be divided into two major groups (Fig. 1.1-1): **motile and non-motile** (Fliegauf *et al.*, 2007). **Motile cilia** beat rhythmically and are critical for cell locomotion and to drive fluid transport over epithelia. Motile cilia can be further subdivided into **motile monocilia** and **multiple motile cilia** (Choksi *et al.*, 2014).

Two distinctive cells belong to **motile monocilia** subgroup: the well-known **sperm flagellum** (Fig. 1.1-2a), that generates a symmetrical, lower amplitude waveform that drives the sperm in a relatively straight line through the female reproductive tract (Turner, 2003); and **nodal cilia cells**, which are located in the embryo node (Hensen's node) . The nodal cilium moves in a clockwise rotation and is critical to creating a leftward flow across the embryo ventral nodes (ventral surface of Hensen's node). It is involved in ciliary signalling functions during organ embryogenesis that rule the determination of left–right asymmetric body axis organ arrangement (Okada *et al.*, 2005; Shinohara and Hamada, 2017).

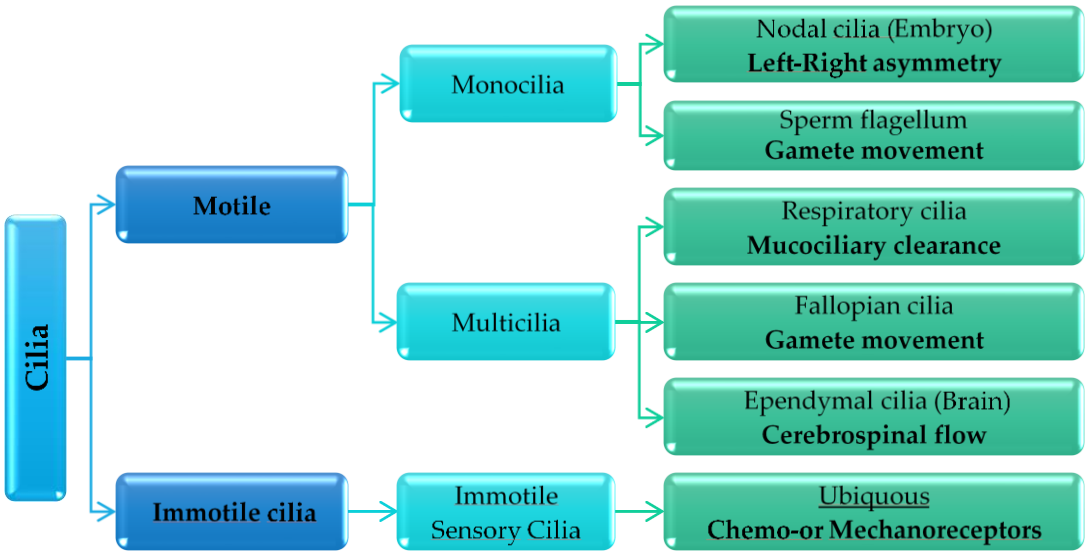


Figure 1.1-1. Hierarchical scheme of different types of cilia and it's major function in mammalian.

Multiple motile cilia, are responsible for the transport of extracellular fluids along the epithelial surface, and are located on **epithelial cells of the fallopian tubes** (Fig. 1.1-2b), **brain ependymal cells** (Fig. 1.1-2c), and **respiratory airways** (Fig. 1.1-2d,e,f) (Fliegauf *et al.*, 2007). The flow generated by the multiciliated epithelium is critical for several tissue-specific functions, namely for **airway mucociliary clearance**. In **Eustachian tubes and middle ear**, the mucociliary clearance drains mucus away from the middle ear into the nasopharynx, thus preventing infection from ascending to the middle ear (Ohashi *et al.*, 1985; Van Der Baan, 1991; Li *et al.*, 2014) The **oral cavity and the nasal cavity** are the main external openings of the respiratory system. Due to the presence of cilia that captures inhaled particles and microorganisms, it is the nasal cavity who mainly filters the air of unwanted contaminants (Bustamante-Marin and Ostrowski, 2017). In the **trachea-bronchial tree** (Fig. 1.1-2 d,e) cilia play a critical mechanism for lung defence. This **mucociliary clearance** protects the lungs, as cilia movement allow to remove mucus, containing inhaled pathogens and pollutants, towards the mouth (Bustamante-Marin and Ostrowski, 2017).

The **non-motile cilium**, also known as primary cilium, is typically one single cilium that is present on most vertebrate cell types and has important roles as critical regulators of signal transduction during development and homeostasis (Goetz and Anderson, 2010).

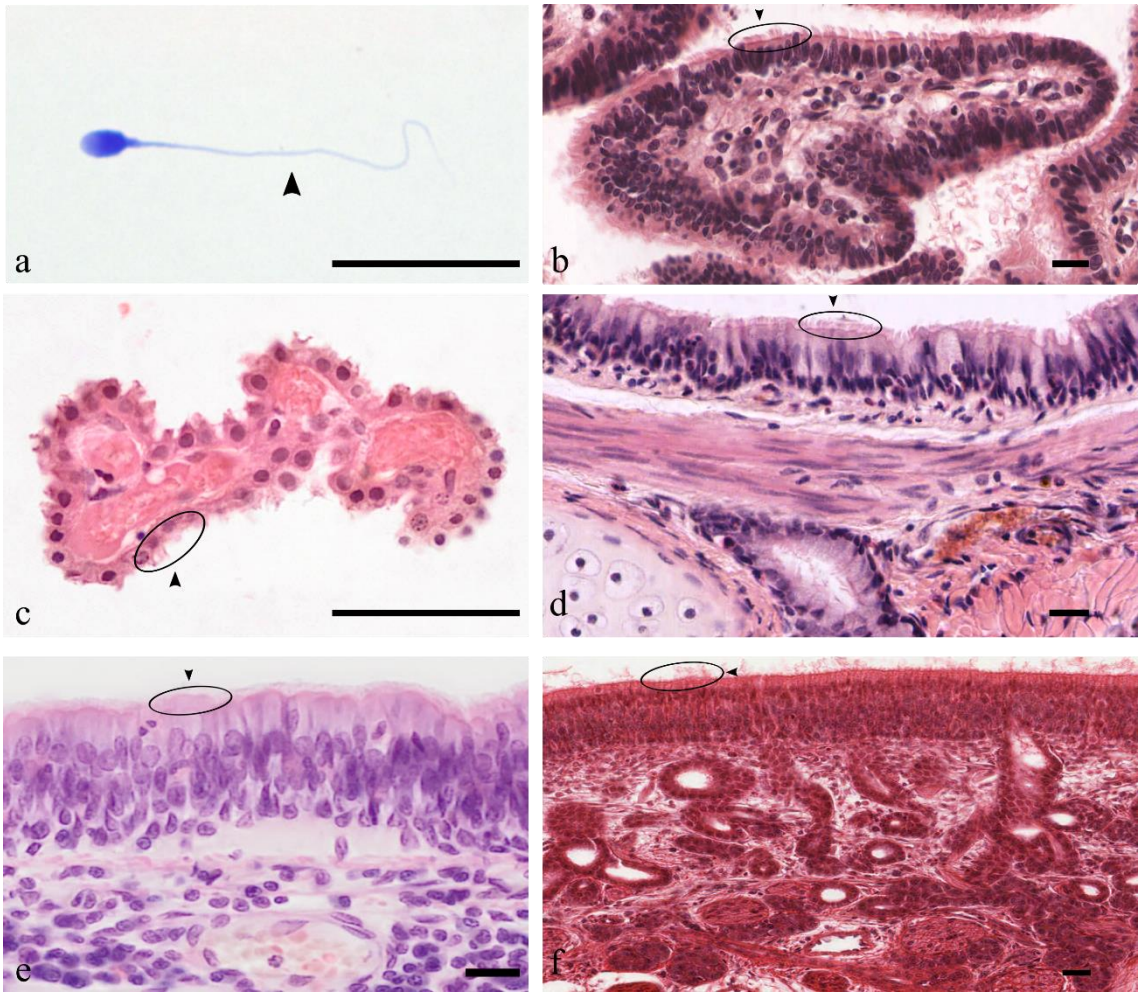


Figure 1.1-2. Light micrograph showing the sperm cell (**a**) and the different locations of motile ciliated cells, namely epithelial cells of the infundibulum of uterine tube (fallopian tubes, **b**); choroid plexus (brain ependymal cells, **c**) and cells from respiratory system [bronchi **d**), trachea **e**) and nasal mucosa **f**)]. Hematoxylin and eosin stain. Scale: 20μm.

Images b) to f) kindly provided by Prof. Eduardo Rocha, full Professor from Instituto de Ciências Biomédicas de Abel Salazar (ICBAS-UP), Department of Microscopy.

1.1.2. Cilium structure

Although the different types of cilia exert various distinctive tissue-specific functions during development, tissue morphogenesis and homeostasis, their basic structure is quite similar. Cilia can be structurally organized into sub-structures (Fig. 1.1-3), which include the ciliary tip, axoneme, ciliary membrane, transition zone and basal body (Fliegauf *et al.*, 2007; Ishikawa and Marshall, 2011; Mirvis *et al.*, 2018).

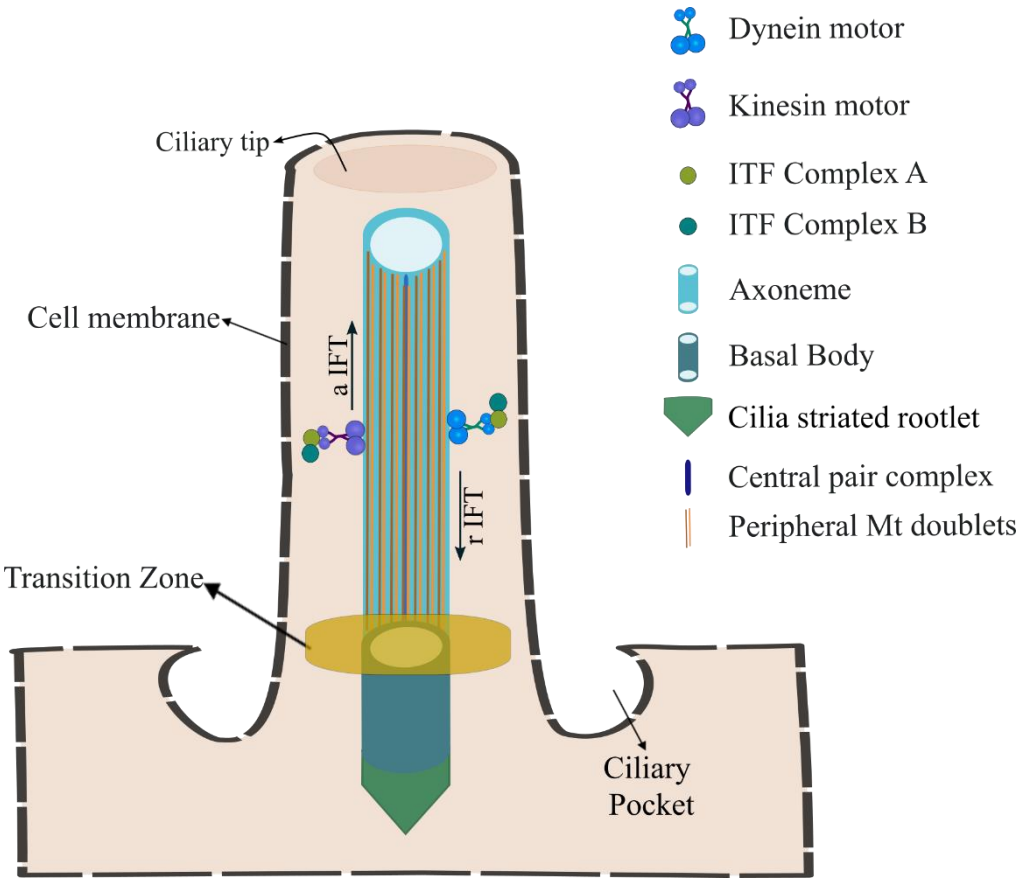


Figure 1.1-3. Simplified view of the cilia main constituents and the anterograde and retrograde intraflagelar transport (aIFT and rITF, respectively). Mt: microtubule. Adapted from: (Rosenbaum and Witman, 2002).

a. Cilium and its main constituents

The Axoneme

The **axoneme**, the structural core of the cilium, is originated from the basal body, a modified centriole derived from the mother centriole that associates with the apical cell membrane at the onset of ciliogenesis (Mirvis *et al.*, 2018). The axoneme is a highly conserved structure. It is a hollow cylinder similar to the centriole and basal body. However, its wall is composed of 9 doublets of microtubules instead of the 9 triplets of microtubules of centrioles. Each microtubule doublet consists of an internal complete microtubule-A onto which is attached a second external and incomplete microtubule-B (Nicastro *et al.*, 2006; Ishikawa, 2017). Microtubule-A is composed of 13-protofilaments, whereas microtubule-B contains 10-protofilaments (Afzelius *et al.*, 1995). Protofilaments are made of heterodimers of α-tubulin and β-tubulin molecules (Oakley, 2000).

Microtubule-A contains a pair of projections named dynein arms (DA). Dynein arms interact with the B-tubule of the adjacent doublet in an ATP-dependent manner, driving doublet sliding and motility (King, 2016). Dynein arms are designated by their position as inner (IDA) and outer (ODA) dynein arms. Each microtubule doublet is connected to the neighbour doublet through the nexin link/bridge. Nexin is a dynein regulatory complex (DRC). The DRC mediate signals, either mechanical, enzymatic or both, between different axoneme components (Heuser *et al.*, 2009). The axoneme of motile cilia contains two central microtubules linked (central link/bridge) by a series of regularly spaced linkages (central projections), and are surrounded by a fibrillar central sheath, constituting the central pair complex (CPC or central apparatus), forming the pattern 9d+2s (Goodenough and Heuser, 1985; Ishikawa, 2017). The connection between the peripheral doublets and the CPC is made through radial projections called radial spokes (RS). The dynein regulatory complex functions as a dynein regulator by mediating structural interactions between the dynein arms, the A-tubule and the radial spokes (Piperno *et al.*, 1994; Heuser *et al.*, 2009); and the RS and CPC are essential for the regulation of DA activity, signal from the CPC are transmitted to DA, driving doublet slipping and axoneme binding (Omoto *et al.*, 1999; Smith and Yang, 2004; Yang *et al.*,

2006). The nine doublets are numbered 1 to 9 in a clockwise direction with number 1 being the one situated on the perpendicular plane to the two central microtubules.

Ultrastructural differences were observed in the axoneme structure of the different cilia types (Fig. 1.1-4). Regarding **motile cilia,** two distinct axoneme conformations were observed. The typical 9d+2s arrangement, found in the axoneme of respiratory epithelial cells (Fig. 1.1-5 a, b), brain ependymal cells and epithelial cells of the fallopian tubes, beats synchronously to create a directional fluid-flow at the tissue level (Fliegauf *et al.*, 2007). The **sperm flagellum** has the same 9d+2s microtubular arrangement but is a distinct type motile cilium (Fig. 1.1-5 c, d) (Inaba, 2007). An exception is the cilia from cells located in the embryonic region node (Hensen's node) and known as **nodal cilia**. **Motile monocilia** cause a different waveform, beating in a unidirectional manner and its axoneme (9d+0) do not include RS neither CPC but have DA and DRC (Okada *et al.*, 2005). However, this arrangement remains to be elucidated as some studies had identified different types of axonemal arrangements, namely the 9d+2s and a novel arrangement consisting of 9 outer doublets and 4 inner singlets (9d+4s) (Feistel and Blum, 2006; Caspary *et al.*, 2007).

In contrast, **primary cilia (non-motile cilia)** axonemes lack the two central microtubules, RS and DA, although maintaining DRC (9d+0). An exception is the kinocilium found in the inner ear, which is considered a non-motile cilium but was observed a 9d+2s arrangement without IDA (Spoon and Grant, 2011).

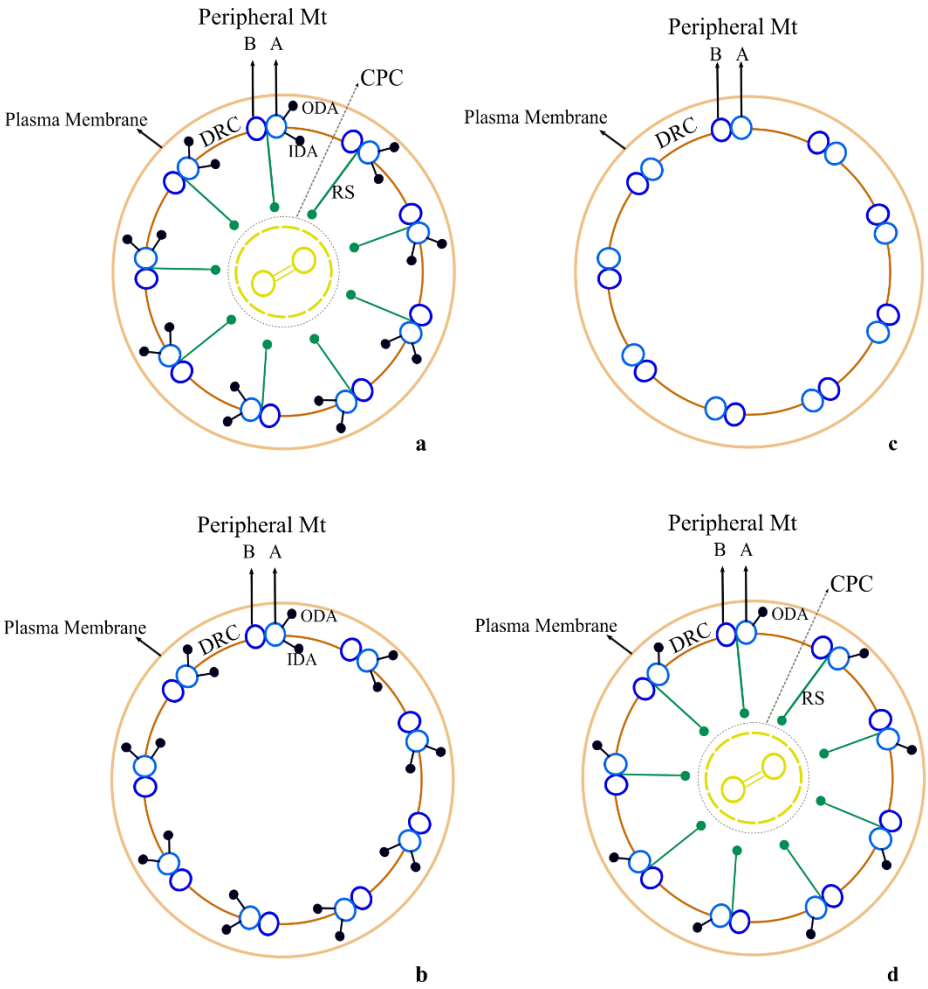


Figure 1.1-4. Schematic representation of the axoneme structure from different types of cilia. A. Representation of motile 9+2 cilia; The axoneme is composed by 9 peripheral doublets of microtubules (Mt) linked by the dynein regulatory complex (DRC) and connected by radial spokes (RS) to a single pair of central Mt (central tubules represented by light green), which are surrounded by a fibrillar central sheath, constituting the central pair complex (CPC). Each pair of peripheral Mt has an inner (A) and an outer (B) Mt. From each A-Mt arise two dynein arms: the outer (ODA) and inner (IDA) dynein arms. **B. Representation of motile 9+0 cilia (nodal cilia).** The Nodal monocilia is also motile, but the RS and CPC are absent. **C. Representation of non-motile 9+0 cilia.** This is also known as sensorial cilia, and lacks dynein arms, RS and CPC. **D. Representation of non-motile 9+2 cilia** (kinocilium), which the structure is identical to motile cilia, but the IDA are absent.

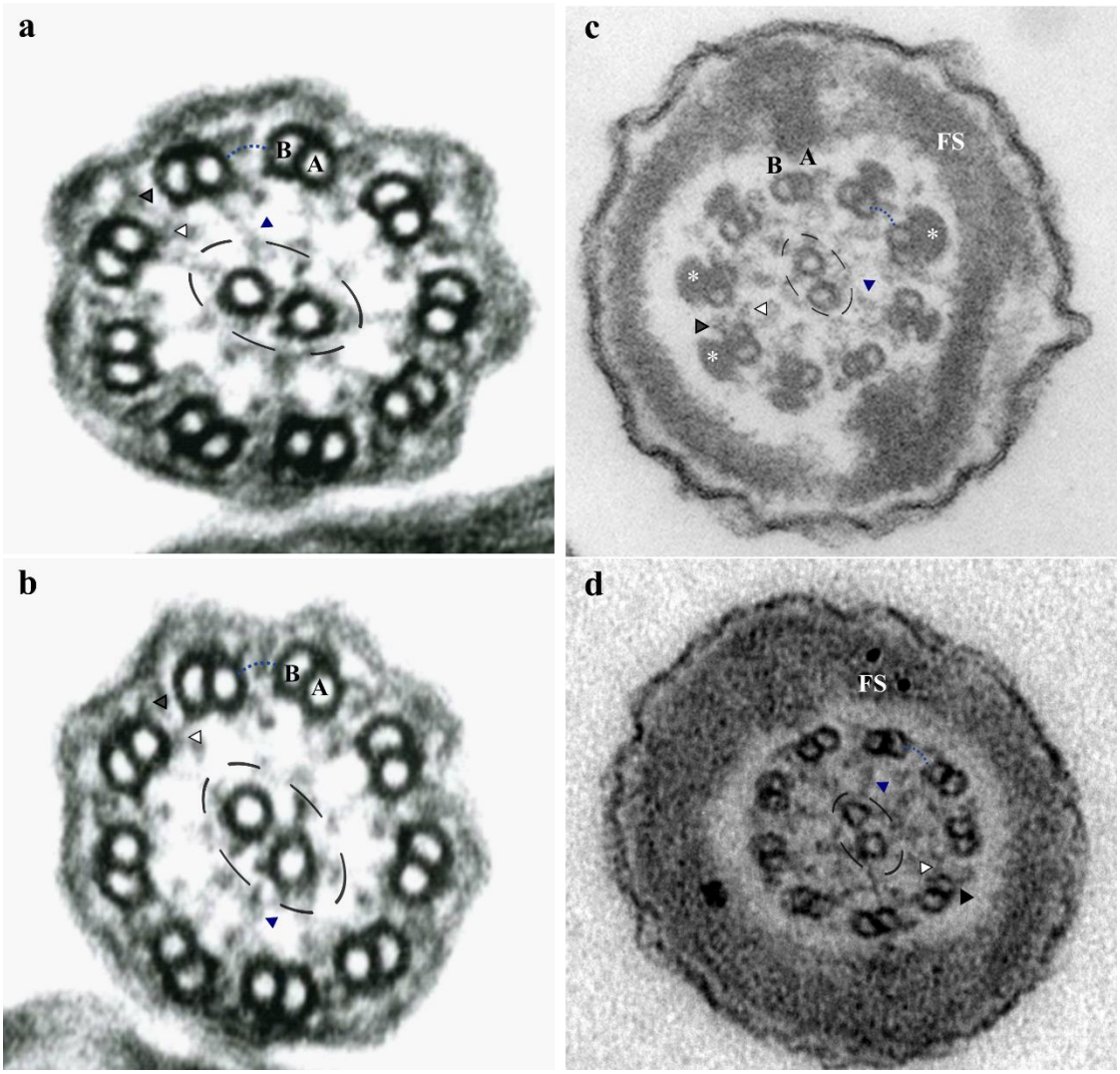


Figure 1.1-5. **Ultrastructure of the axoneme from respiratory epithelial cells (a, b) and sperm cells (c, d),** with nine peripheral pairs of microtubules and two single central microtubules, thus forming the pattern 9 +2. The inner peripheral microtubule is indicated by an **A**, while the outer is indicated by an **B.** The DRC that links the microtubule doublets are represented by dot blue line. The radial projections, that bind the doublets to the two central microtubules are indicated by blue arrowhead. The microtubule A of each doublet has two dynein arms: an outer (black arrowhead) and an inner (white arrowhead). The central pair complex, with two central microtubules joined by the central bridge and wrapped in a fibrillar sheath, are marked by a dashed circle. Regarding the axoneme of sperm cells (**c** and **d**), additionally possess the outer dense fibres (*) adjacent to the axoneme and since the principal piece up to the end piece there are rings of the fibrous sheath (FS).

Image provided by Mário Sousa, MD, PhD, Cathedratic Professor, from ICBAS-UP, Department of Microscopy.

The ciliary tip

The ciliary tip is localized at the top of the cilium and contains the microtubule plus ends. At this region, new tubulin subunits (αβ-tubulin heterodimers) are added to the axonemal microtubules during cilia growth. The ciliary tip also contains signalling molecules and can undergo morphological changes in response to signalling processes (Sloboda, 2005). The **intraflagellar transport** (IFT) is a bidirectional process by which axonemal/ciliary precursors and sensory proteins, known as **IFT particles**, are constantly delivered from their site of synthesis in the cell body to the axonemal assembly site at the ciliary tip, a transport that then convey cargo back (Rosenbaum and Witman, 2002). The IFT particles (molecules and/or proteins that are moved by IFT) contain IFT-A and IFT-B subcomplexes (Cole *et al.*, 1998), and assemble into **IFT trains** (multiple IFT particles connected in an array). Cilia and flagella microtubules suffer a constant **turn over** at the plus end (Marshall and Rosenbaum, 2001). This is a bidirectional process that relies on two large super-families of **microtubule motors**: kinesins and dyneins (Further reviewed elsewhere by King, 2000; Sweeney and Holzbaur, 2018). Consequently, IFT is divided into two types: **anterograde IFT** and **retrograde IFT** (Fig. 1.1-3). In anterograde IFT, kinesins, namely Kinesin-II, are the motors that move IFT trains from the cilia base to the tip. At the cilia tip, the cargo transported by kinesins is dissociated from kinesin motors and is mixed with complexes from other IFT trains. Then, IFT trains are reorganized by dyneins and retrograde IFT begins, with transport of IFT particles from the tip to the ciliary base, though the action of dyneins, namely cytoplasmic dynein 1b and 2 (Pedersen and Rosenbaum, 2008; Chien *et al.*, 2017). IFT is thus essential for the formation and maintenance of all cilia and flagella (Pedersen and Rosenbaum, 2008; Taschner and Lorentzen, 2016; Jensen *et al.*, 2018).

Basal body

The basal body is a hallow barrel whose walls are formed by 9 microtubule triplets (A, B, C). The basal body is derived from centrioles and serves as the template from which the axoneme extends (Marshall, 2008). In this process, microtubule C is caped, and only microtubules A and B expand, originating the 9 doublets of the axoneme. In multi-ciliated epithelia, ciliogenesis begins with centriole multiplication, and then the mother centriole migrates to the apical cell surface, acquires accessory structures, and become

linked to the cell membrane (**transition fibers**). Once docked to the cell membrane, it is known as the basal body. Finally, the axoneme is extended to form the cilium (Bettencourt-Dias and Glover, 2007; Dawe *et al.*, 2007; Ishikawa and Marshall, 2011).

The transition zone and the ciliary necklace

The basal region of the cilia, above the basal body, is called the **transition zone** (cilia neck). This region is made by Y-shaped protein linkers that connect microtubule doublets to the cilia membrane. The membrane of the transition zone is called the **ciliary necklace**. The ciliary necklace is a specialized membrane domain that consists of circumferential strands (rows) of intramembranous particles around the basal region of the axoneme. Their number is species and cell-specific. The transition zone compartmentalizes the cilium, functioning as a gate that controls communication between the cytosol and the ciliary compartment (Deane *et al.*, 2001). The free access of cellular components to the cilium is thus blocked by the basal body and specialized structures within the transition zone (Garcia *et al.*, 2018). The transition zone thus acts as a barrier to regulate the selective access of proteins and lipids to the axoneme and cilia membrane. Some pieces of evidence point to a functional interaction between the transition zone and the IFT machinery in the control of entry and exit of proteins into cilia (Czarnecki and Shah, 2012; Reiter *et al.*, 2012; Gonçalves and Pelletier, 2017). In cilia with a 9d+2s pattern, the central pair is polymerized in the transition zone by a microtubule organizer center (Inaba and Mizuno, 2016; Werner *et al.*, 2017)

The ciliary membrane

The ciliary membrane is essential for the function of cilia as it contains proteins with sensory functions that localize to specific membrane regions (Pazour and Bloodgood, 2008). As the ciliary membrane is continuous with the plasma membrane, much of the communication between cytoplasmic signals and cilia occur through the ciliary membrane, which behaves as sensor for specific cell events (Rohatgi and Snell, 2010; Garcia *et al.*, 2018). Moreover, the ciliary membrane plays an important role in the regulation of the ciliary beat pattern due to the presence of ion channels and pumps specifically located within the ciliary membrane (Salathe, 2007). Well-known examples are calcium channels and calcium pumps (Schmid and Salathe, 2011). In particular situations, the ciliary membrane may also participate in cell adhesion. For instance,

adhesive properties of the cilia membrane are thought to play important roles in mammalian fertilization, such as in mediating sperm flagellum membrane binding to the ciliary membrane of oviduct epithelial cells (Lefebvre *et al.*, 1995) and by promoting adhesive bonds between the membrane tip of the sperm flagellum and elements of the cumulus mass (Norwood *et al.*, 1978; Talbot *et al.*, 2003).

1.1.3. Motile Cilium: a singular structure

a. Mechanisms to motility

Ciliary and flagellar motility are critical to several tissue-specific functions as discussed previously. This elegant process remains quite conserved from more primitive organisms, such as the green alga *Chlamydomonas reinhardtii* to more complex organisms like humans. Even though it might seem a simple process, it is highly regulated, and tissue-specific dependent on several metabolic pathways and regulatory mechanisms (Shingyoji, 2012; Choksi *et al.*, 2014; Pereira *et al.*, 2017).

Cilia and flagella are, as already mentioned, showed specific and important differences, particularly regarding the motility, making these two distinctive structures (Dawe *et al.*, 2007; Inaba, 2007). The typical ciliary movement is a sweeping movement beating with a forward power stroke and a backward recovery stroke cilia, also known as effective and recovery strokes (Fig. 1.1-6c), respectively (Chilvers and O'Callaghan, 2000; Shingyoji, 2012) whereas flagellar motility is a propeller-like motion characterized by a rhythmic and asymmetric movement that usually consists of successive waves originating at either the base or the tip of the flagellum (Fig. 1.1-6a) (Luconi *et al.*, 2006). Likewise, the ciliary movement, the flagellar beat also contains a forward power stroke and a backward recovery stroke stage (Bayly *et al.*, 2011). Both in cilia and flagellum, the power stroke produces the propulsive force, whereas the recovery stroke corresponds to periods of slower rotation and translation. The recovery stage may be described as a stage where the source of potential energy is being refilled to be available to be further converted into kinetic energy during the next forward power stroke (Shingyoji, 2012).

In contrast to cilia, sperm cells need to have two types of motility to accomplish properly its function: the activated motility, a symmetrical lower amplitude waveform, characteristic from freshly ejaculated sperm; and the hyperactivated motility, in an asymmetrical and higher amplitude waveform, that is seen in sperm at the site and time of fertilization in mammals (Suarez and Pacey, 2006; Suarez, 2008; Suarez, 2010). Notwithstanding those critical differences between cilia and flagella, both have a common origin and share common pathways and mechanisms regarding its formation, maintenance and function.

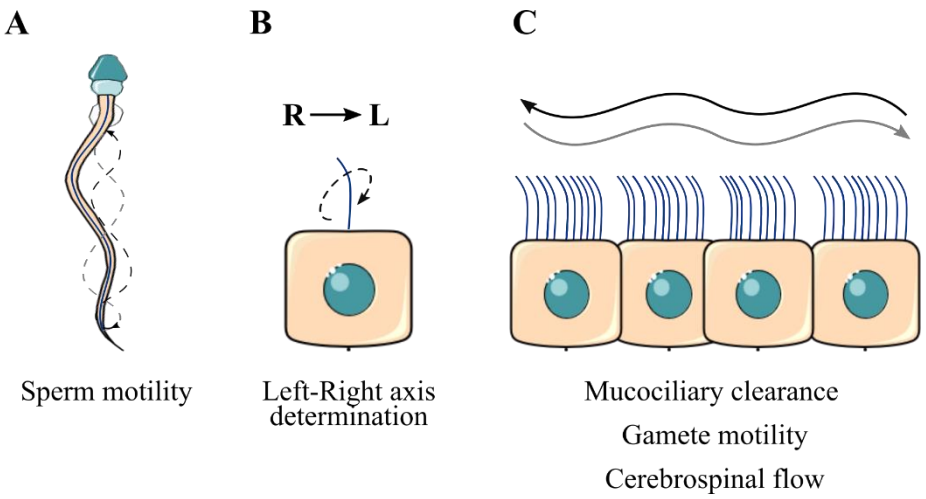


Figure 1.1-6. Schematic representation of the different types of cilia motility and the main biological function to which the movement is associated. A. Representation of the flagellar motion. It is characterized by a rhythmic and asymmetric movement, that contains forward stroke (black dot line) and recovery stroke (grey dot line) **B** Node and their clockwise rotation, which generates a leftward effective stroke and rightward recovery stroke vital to define mammalian left-right axis. **C** Drawing of ciliary sweeping movement with a forward power stroke (black arrow) and a backward recovery stroke (grey arrow), which is critical for mucociliary clearance, gamete motility in female reproductive system and to cerebrospinal flow.

Dyneins

Dynein arms are motor proteins that convert the chemical energy contained in ATP into the mechanical energy of movement. Dynein arms are regularly placed along the doublet microtubules. Outer dynein arms are organised at intervals of 24 nm and are believed to be responsible to regulate the ciliary beat frequency (CBF). The ODA are elegantly controlled by external factors such as phosphorylation, variations in Ca^{2+} concentration and flow as well as alterations in redox state (King, 2018b). In contrast, IDA are docked to microtubule precisely in a 96-nm axonemal repeat and was shown to control of ciliary/flagellar bending (Hwang *et al.*, 2018; King, 2018b).

The DA consist of protein complexes forming the heavy (HC), light (LC) and the intermediate (IC) chains (Fig. 1.1-7). In mammalians, ODA is comprised of two HC of about 500 kDa; three to five IC of about 120–60 kDa; and six LC with approximately 30–80 kDa each (King, 2018). In contrast to the ODA, which is quite well defined structurally and shows a homogeneous pattern regarding its composition and organization, the IDA has been presented as structurally heterogeneous and is thought to contain at least seven heterodimeric and monomeric heavy chain isoforms.

The HC is the main responsible for the dynein motor activity as comprising the motor machinery and the sites of both ATP hydrolysis and ATP-sensitive microtubule binding. The HC is thus responsible for transducing chemical energy into directed mechanical force applied to the microtubule surface (Roberts *et al.*, 2013; Schmidt and Carter, 2018). HC is generally organized into five functional elements: (1) a **"tail" domain** located at N-terminal domain and that is required for dimerization and cargo binding; (2) "**motor domain**"(or head), at C-terminal region, that folds into ring containing six ATPase domains associated with diverse cellular activities (AAA for: ATPases associated diverse cellular activities)(Kon *et al.*, 2012; Schmidt *et al.*, 2012); (3) a **"linker"**, that is composed by a series of subdomains (mostly α-helices) connecting the motor domain and tail, essential for motility (Roberts *et al.*, 2009); (4) the "**stalk**", that consists of a pair of helices (CC1 and CC2) wrapped around each other to form an antiparallel coiled-coil that emerge from the forth AAA region (Gibbons *et al.*, 2005) and (5) the **microtubule-binding domain** (MTBD), is a small α-helical domain at the end of the stalk responsible for binding the dynein to microtubules with the MTBD at its end, that is supported by a coiled-coil structure located in the AAA5, called the "**buttress"** (Gee *et al.*, 1997; Redwine *et al.*, 2012; King, 2016).

Regarding the AAA motifs, the first four possess P-loop motifs that are well conserved in all dynein sequences and can bind ATP (i.e. are nucleotide-binding sites of the dynein motor). The first AAA (AAA1), is critical for movement as is only one that hydrolyses ATP efficiently and mediates the power stroke (Burgess *et al.*, 2003). The AAA modules 2 to 4 are non-hydrolytic and possess intact P-loops that possibly are capable of binding nucleotides and may interact with AAA1 to generate the movement. The two last AAA domains lack P-loops and are believed that consequently do not bind nucleotides and their functions are not yet totally defined (Kon *et al.*, 2012; Roberts *et al.*, 2013; Schmidt and Carter, 2018). Between the 4 and 5 AAA modules is located the MTBD at the tip of an antiparallel coiled-coil stalk. which is the functional contact between dynein and the microtubule (Roberts *et al.*, 2013; Schmidt and Carter, 2018). The interzone between each AAA domain appears to be important for the conformational change during mechanochemical cycles, which presumably exerts the power stroke (Roberts *et al.*, 2013).

Regarding the IC and LC they play important roles in motor assembly, namely, it helps to specify the intracellular location of the dynein (Wirschell *et al.*, 2013) and regulation of motor activity (Hwang *et al.*, 2018; King, 2018a). Further functions have been recognised the LC, such as redox-sensitive vicinal dithiols, Ca2+ binding (King, 2018a) and intraflagellar transport (Pazour *et al.*, 1998).

The variety of structure and function of these chains highlights its great complexity, with many regulatory mechanisms involved and more to be identified.

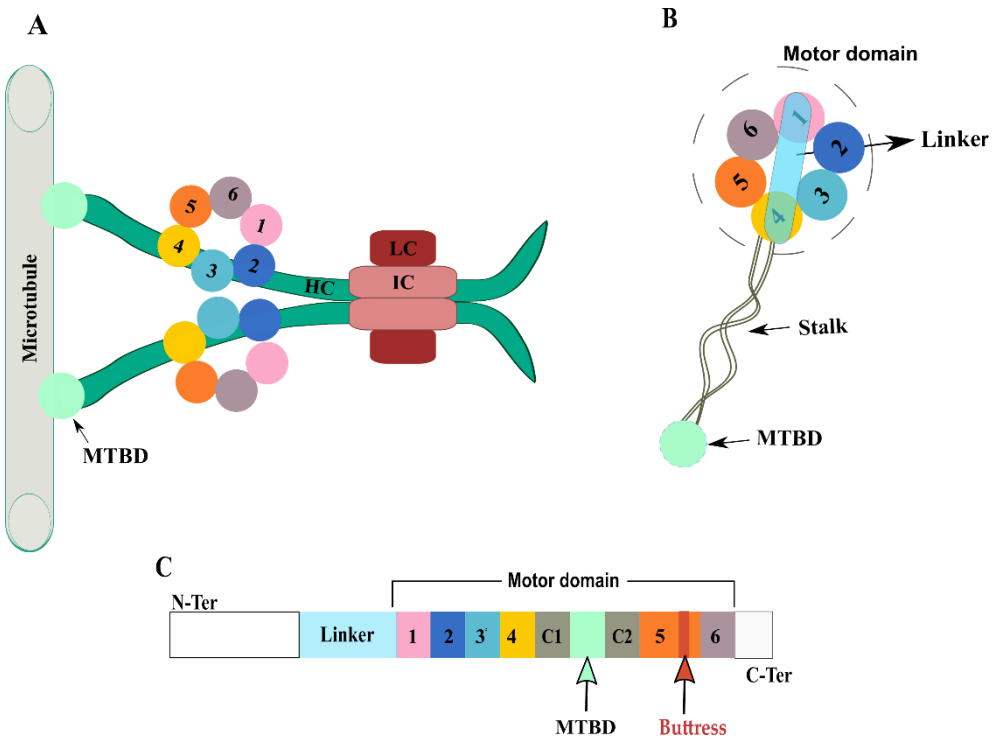


Figure 1.1-7. Architecture of the dynein motor. (A) Schematic representation of a dynein motor complex. Dynein motor is a dimer containing two identical heavy chain (HC) subunits, with its motor domain and microtubule binding domain (MTBD), and intermediated and light chains, respectively IC and IL. **(B)** Diagram representing dynein heavy chain domain structure: the dynein motor domain with its ring of six AAA domains (numbers 1 to 6 represent, correspondingly the AAA1 to AAA6); the linker domain, that stretches across a ring of six AAA domains; the stalk, which emerges from the small domain of AAA4 and end with the MTBD domain. **(C)** The illustration of the linear map of a general dynein HC composition, including the N-terminal tail, the linker and the six AAA domains (labelled 1–6). The position of the stalk is represented in dark green, by the two predicted coiled coil structures CC1 (C1) and CC2 (C2) that are located between AAA4 and AAA5. MTBD and buttress are also indicated, as well as the C-terminal domain. Adapted from Schmidt *et al.* (2012); King (2016)

Dynein Docking Complex and Dynein Regulatory Complex

The outer dynein arm docking complex (ODA-DC) is a structure that interacts directly with the DA being responsible for the assembly and binding of ODA at microtubules doublet in regular intervals. The ODA-DC contains three polypeptides (DC1-DC3) (Takada *et al.*, 2002; Casey *et al.*, 2003). The DC1 and DC2 polypeptides are suggested to potentially determine the spacing of the ODA (Takada *et al.*, 2002), while the DC3 has four motifs that possibly bind Ca^{2+} thus playing a role in calcium-regulated

ODA activity, as well as, other important roles in the regulation of the ODA (Casey *et al.*, 2003).

The DRC functions are of upmost importance within the dynein regulation and limitation of doublet sliding (Heuser *et al.*, 2009; Lin *et al.*, 2011; Porter, 2018). At least DRC proteins were identified: DRC1 and -2 that are located in the base plate; DRC3, -4, and -7 found in the linker domain (in the n-terminal region of HC), and DRC5 and -6 are localized to the distal lobe of the linker of the N-DRC (Lin *et al.*, 2011). Studies using DRC mutants showed that besides their regulatory functions, DRC plays a critical role in mediating structural interactions between the DA, the A-tubule of the outer doublet, and the RS (Piperno *et al.*, 1992; Gardner *et al.*, 1994; Piperno *et al.*, 1994). For instance, DRC components 3, 4, and 7 possibly stabilize the binding of IDA isoforms, whereas components 1 and 2 may form part of the binding site for both RS and certain inner arm subspecies (Porter, 2018).

Radial spokes and central pair complex

The CPC is composed of two structurally asymmetric and biochemically distinct central microtubules, linked by central projections and surrounded by a fibrillar sheath. It was suggested that mechanical interactions between the CP and RS are converted into a biochemical signalling pathway that may alter the phosphorylation state of the dynein arms, which then regulates microtubule sliding (Smith and Yang, 2004; Porter, 2018). An example of the involvement of CP/RS in motility was given by Smith and co-workers (Smith and Lefebvre, 1997; Smith, 2002; Smith and Yang, 2004). They proposed that RS and CPC may be involved in converting simple symmetric bends into the asymmetric waveforms required for forward swimming and in the release of ATP inhibition in a controlled manner. Moreover, central pair may function as a distributor to provide a local signal to the RS that selectively activates subsets of DA (Omoto *et al.*, 1999; DiPetrillo and Smith, 2010).

The RS in Chlamydomonas flagella contains at least 23 distinct polypeptides, termed RS protein (RSP)1 to RSP23 with a combined molecular mass of approximately 1200 kDa (Yang *et al.*, 2006).

Several remain poorly characterized, while others key function has been attributed. For instance, RSP3 anchors the RS to the outer microtubule doublet and contains an AKAP (for A-kinase anchoring protein) domain and binds in vitro the cyclic AMP (cAMP)-dependent protein kinase regulatory subunit (Gaillard *et al.*, 2001). The AKAP kinases are able to bind to protein kinase A (PKA)substrates, which enables the activation of the PKA catalytic subunit (C) and the raise of the cAMP. This process allows phosphorylation (by PKA). In the opposite process (the dephosphorylation), the AKAPs bind to phosphodiesterases (PDEs), which break down cAMP phosphatases and dephosphorylate downstream PKA targets and also other kinases (PKC and MAPK)(Michel and Scott, 2002; Gold *et al.*, 2006; Luconi *et al.*, 2010). Some AKAPs can bind both regulatory subunits (RI & RII) of PKA and are dual-specific AKAPs (D-AKAP1 and D-AKAP2). The fact of RS present AKAP proteins in its composition maw suggest phosphorylation/dephosphorylation mechanisms as rulers of the regulation of DA by RS/CPC (Porter and Sale, 2000). Additionally, *RSPH1* gene encodes a RS-head protein, mainly expressed in respiratory and testis cells, it is important for the proper building of CPC and RS, since abnormal axonemal configuration with CPC and RS defects were observed in PCD patients with mutations in the RSPH1 (Kott *et al.*, 2013; Knowles *et al.*, 2014; Onoufriadis *et al.*, 2014).

b. Motile cilia functions

In mammals, motile cilia have a limited distribution, being only detected in the respiratory epithelial cells, brain ependymal cells, nodal cells, epithelial cells from the fallopian tubes and in sperm flagellum. Thus, their functions are highly specific and regulated (Choksi *et al.*, 2014; Mirvis *et al.*, 2018; Vladar *et al.*, 2018). I will briefly describe some vital functions in which motile cilia play an essential role.

Mucociliary clearance

During breathing, lungs are continuously expose to multiple pathogens, particles and chemicals. Consequently, a powerful and competent army is needed to defend it. Mucus and mucociliary clearance are the main soldiers of innate mechanical defences in the lungs(Munkholm and Mortensen, 2014; Bustamante-Marin and Ostrowski, 2017; Smith *et al.*, 2019).

Epithelial airway cells form a physical barrier to bacteria and viruses, allergens, dust particles, and air pollutants. In addition, the airway epithelial cells provide an antimicrobial function via production of mucus, immunoglobulins, and defensins. The airway epithelium also plays regulatory and pro-inflammatory roles through the release of neuropeptide degrading enzymes, endothelin, nitric oxide, TGF-β, arachidonic acid metabolites, and cytokines. Cilia are present in most of the airway epithelial cells, enabling mucous clearance, but also have sensory functions and can respond to mechanical and chemical stimuli (Yaghi and Dolovich, 2016)

About 50-80% of epithelial airway cells are ciliated, and each ciliated cell presents about 200-300 cilia. They are present from the nasal cavity to bronchi (Bustamante-Marin and Ostrowski, 2017). These cells present mitochondria in both the apical and the basolateral domains. The apical region is more-rich in mitochondria to ensure the availability of ATP that is used to sustain ciliary motility. Mitochondria also function as an internal Ca^{2+}-buffering system in airways, forming barriers to prevent/regulate global cell Ca^{2+} signalling upon plasma membrane receptor activation (Ribeiro *et al.*, 2003).

The mucus produced by airway epithelial cells is a viscous fluid, mainly composed by water and a complex network of proteins, electrolytes, enzymes and lipids. Mucins, a group of large glycoproteins, predominate in this mucus (Rubin, 2010). Airway mucus have multiple important functions: sequester, immobilize, and remove harmful particles and invading microorganisms from the airway; transport essential secreted substances that will help in lung defence, such as, defensins, collectins, antiproteases, and immunoglobulins; and provide a water proofing layer that acts to reduce fluid loss through the airways (Smith *et al.*, 2019).

The ciliated epithelium is covered by a mucus layer (above cilia tips) of gel-like consistency, and by a periciliary layer (between cilia). In the periciliary region, cilia are immersed in low viscosity medium containing water and macromolecules that prevent the passage of particles from the mucous layer. The mucociliary clearance is the process by which upper airways sequester and remove inhaled toxic particles and infectious microorganisms. To accomplish this function, the epithelium presents three main components: cilia, which transports the mucous; a mucus layer that traps inhaled toxic

particles and infectious agents; and a periciliary layer that provides a favourable environment for ciliary beating and cell surface lubrication (Bustamante-Marin and Ostrowski, 2017).

Proper ciliary function is critical for effective mucociliary clearance and for this, ciliated airway cells must move coordinated. Cilia beat in metachronal waves (movements produced by the sequential action of cilia, as opposed to synchronized, producing the appearance of a travelling wave), at a beat frequency of about 4.2 ± 1.0 Hz (Yaghi and Dolovich, 2016). The beat frequency is age-related, as the ciliary beat frequency in children is significantly greater than in adults (Chilvers *et al.*, 2003). Further studies should be conducted to further understand those differences observed between adults and children and eventually to analyse if gender and/ethnics differences may also affect beat frequency. This would also improve the diagnosis of ciliary diseases.

The ciliary beat frequency is also regulated by external factors, such as: (1) intracellular Ca^{2+} concentration (Ribeiro *et al.*, 2003; Shah *et al.*, 2009; Schmid and Salathe, 2011); (2) temperature. Since hydrolysis of ATP by dyneins require an optimal temperature of about 37ºC for the enzymatic reaction to occur, low temperatures decrease ciliary beat frequency and mucus transport (Sears *et al.*, 2015); and (3) intracellular pH. Changes in pH_i directly affect DA activity: intracellular alkalization stimulates the ciliary beat frequency, whereas intracellular acidification decreases the ciliary beat frequency (Sutto *et al.*, 2004).

Although the ciliary beat frequency is directly related with mucociliary clearance, it is not the only player. Mucociliary clearance is further influenced by the number and length of the cilia, the coordination of ciliary beating, and the ciliary beat amplitude (Yaghi and Dolovich, 2016; Bustamante-Marin and Ostrowski, 2017). Hence, although the full mechanisms and regulators are not fully understood, it is well established that cilia play a major role in lung defence, particularly in mucociliary clearance.

Left-right asymmetry

Polarization starts early in the preimplantation embryo, where blastomeres do not all divide in a synchronized manner, and are not genetically and morphologically symmetric (Gilbert, 2000). The mammalian body is not symmetrical. The heart begins its

formation at the midline of the embryo, but then moves to the left side of the chest cavity and loops to the right; the spleen is found solely on the left side of the abdomen; the large intestine loops right to left as traverses the abdominal cavity; the right lung shows one more lobe than the left lung and, in human males, the scrotum is not symmetrically positioned (Gilbert, 2000; Wright, 2001; Shinohara and Hamada, 2017).

Gastrulation occurs when the implanted blastocyst (blastula) folds inward and enlarges to create a gastrula. The node, or primitive knot, is an enlarged group of cells located in the anterior portion of the primitive streak in the developing gastrula. The node is the site where gastrulation, the formation of the three germ layers, first begins. The node determines and patterns the anterior-posterior axis of the embryo (Gilbert, 2000). Further it is the node that determines the left-right axis of the embryo (Okada *et al.*, 2005; Hirokawa *et al.*, 2006). Node monocilia are motile and produce a laminar leftward flow of the extraembryonic fluid surrounding the node. The leftward **node flow** generated by motile monocilia at the embryonic node is a key element in the establishment of the body axis (Hamada *et al.*, 2002; Hirokawa *et al.*, 2006; Shinohara and Hamada, 2017). The direction of the node flow is thought to be determined by a combination of two features of node cilia: their posterior tilt and clockwise rotation, which generates a leftward effective stroke and rightward recovery stroke near the surface of the cell (Nonaka *et al.*, 2005; Okada *et al.*, 2005); and the rotational movement of the cilia that may generate slow rightward fluid flow near the cell surface and fast leftward flow in the middle of the node cavity (Cartwright Julyan *et al.*, 2007). However, the precise mechanism by which the fluid flow results in symmetry breaking and thus leads to the left-right asymmetry is largely unknown. It has been hypothesized that the nodal flow may transport small molecules or vesicular particles to the left side of the embryo, which may orient the organ position or function as left-side determinant (Hirokawa *et al.*, 2006), or exerts mechanical pressure (Shinohara and Hamada, 2017). Other components assist in determining the LR axis: Ca^{2+} was suggested as being involved in flow sensing; the left-right dynein (Lrd) protein was proposed to be involved in the generation of leftward nodal flow; and the cation channel "polycystin-2" seems to be involved in sensing the nodal flow and to initiate an asymmetric calcium signal, inducing *Nodal* gene expression (Deng *et al.*, 2014).

Failure to properly establish left-right (LR) asymmetry results in laterality defects, and development of LR asymmetry depends on a mechanism capable of generating asymmetry from a pre-existing bilateral symmetry. Molecular asymmetry begins before morphologic asymmetry, being dependent on the restriction of *Nodal* gene expression at the left side of the node, where it codes for Nodal, a LR signalling protein. The main LR determining genes are *Nodal*, *Lefty* and *Pitx*, which are governed by the expression of several upstream genes not directly involved in asymmetry specification. The Nodal cascade consists of Nodal, a TGFβ-type growth factor, Lefty, a Nodal repressor, and Pitx, a transcription factor. Once activated, Nodal signaling directly promote its own transcription, leading to nodal mRNA expression in the lateral plate mesoderm. Further Nodal activates lefty and pitx2 transcription. Nodal expression is inhibited shortly by Lefty and *Cerl2* genes, whereas the expression of *pitx2* gene is maintained until organ morphogenesis, which is initiated right after Nodal expression has disappeared. Expression of this cascade is both necessary and enough to induce the correct asymmetric placement of organs, with the nodal flow generated by cilia triggering the asymmetric Nodal cascade (Hamada *et al.*, 2002; Deng *et al.*, 2014; Shinohara and Hamada, 2017)

Some of upstream genes are well studied and include KIF (kinesins) and Lrd (dynein of the ODA). Randomization of left–right asymmetry was observed in mice lacking KIF3B motor protein due to loss of nodal cilia generating leftward flow of extra-embryonic fluid (Nonaka *et al.*, 1998). Further, Lrd localizes to a centrally located subset of node monocilia, polycystin-2 is found in all node monocilia, and asymmetric calcium signalling appears at the left margin of the node coincident with nodal flow. These observations suggest that LR asymmetry is established by a set of molecular events in conjunction with the ciliary mechanism: Lrd-containing monocilia generate nodal flow, and nonmotile polycystin-2 containing cilia sense nodal flow, initiating an asymmetric calcium signal at the left border of the node (Nakamura and Hamada, 2012; Deng *et al.*, 2014).

Left-right determination is an extraordinary and puzzling process, with still several open questions that need to be answered, namely how cilia sense fluid flow during LR determination, how is this flow genetically determined, which external factors (beside

Ca^{2+}) are involved, what is the composition of nodal motile cilia that distinguish them from other, etc. The answers for these questions will certainly be of great value in understanding human development and some diseases' pathophysiology and will open new areas of cell, developmental, and evolutionary biology.

Cerebrospinal fluid

The cerebrospinal fluid, the major extracellular fluid of the central nervous system, is produced from arterial blood by the ventricular choroid plexuses. The choroid plexus consist of tufts of capillaries with thin fenestrated endothelial cells covered by modified ependymal cells (Skipor and Thiery, 2008). Although the biological function of this flow is not fully understood, the cerebrospinal fluid is known to play an important role in the development and organization of the central nervous system through neuronal guidance, and to provide important nutrients (Siyahhan *et al.*, 2014).

Coordinated beating of ependymal cilia promotes the flow of cerebrospinal fluid, being required for neuroblast orientation and for the formation of a concentration gradient in the subventricular zone (Sawamoto *et al.*, 2006). Ependymal cilia generate separate flows and rotations that may also be critical to establish and influence intraventricular boundaries that are responsible to concentrate locally released compounds, thus preventing their entry into off-target regions. Transient local changes in the beating pattern, whose causes are still unknown, induce changes in ventricular subdivisions, which possibly can be caused by transient changes in cell-cell interactions and in planar cell polarity (Sawamoto *et al.*, 2006).

Gametes motility

Sperm motility is an essential mechanism to life as it is vital for natural reproduction and offspring generation. The sperm flagellum is considered a special type of motile cilia, as it possesses the same cytoskeleton core, the axoneme 9d+2s, which drives sperm motility. However, the sperm flagellum presents very distinct structural components, and a different type of motility and its regulation Sperm do not develop a basal body and do not possess a transition zone. One centriole, the proximal centriole, binds to the nuclear base and becomes active after fertilization, where it triggers aster generation, pronuclei apposition and the first embryo division. The other cell centriole, the distal centriole,

originates the axoneme, and thereafter disaggregates in order that after fertilization the oocyte only receives one sperm centriole (the oocyte is devoid of centrioles). The other structural differences to cilia are that the axoneme is encircled by 9 outer dense fibbers, which confer resistance, and these by ribbons of the fibrous sheath, which in addition to confer resistance also contain glycolytic enzymes. The first third of the axoneme is further encircled by a mitochondrial helix, supposed to furnish the energy needed for membrane transformations occurring during the maturation process in the epididymis. The energy demanded for the capacitation process in the uterine cavity, sperm migration and penetration through the investments of the ovulated oocyte is provided from glycolysis occurring in the two thirds of the flagellum.

The main mechanisms that control sperm motility were previously reviewed by our group (Pereira *et al.*, 2014; Pereira *et al.*, 2017). Motile cilia in the efferent ducts facilitate transport of sperm to the epididymis (Boon *et al.*, 2013). Likewise, in the female Fallopian tubes, motile cilia help in the transport of the embryo towards the uterine cavity (Lyons *et al.*, 2006; Raidt *et al.*, 2015). Ciliated cells found in tubar fimbria, infundibulum, ampulla and isthmus are identical to respiratory cilia in terms of composition, and in the ciliary beating pattern and frequency (Raidt *et al.*, 2015). In the Fallopian tube, the ciliary beat frequency was proposed to be regulated by progesterone (Bylander *et al.*, 2010). The coordinated beating of tubar cilia generates a directed fluid flow towards the uterine cavity, which intrinsically interacts with muscle contractions, to allow the propulsion of embryos (Lyons *et al.*, 2006). Loss of tubar cilia motility reduces fertility and may increase the risk of ectopic pregnancies (Lyons *et al.*, 2006).

1.2. Primary Ciliary Dyskinesia (PCD)

Ciliopathies are a group of genetic disorders whose aetiologies lie in defective structure and/or function of cilia. Inherited defects in the function/ structure of motile cilia causes a rare ciliopathy called Primary Ciliary Dyskinesia (PCD, OMIM: 244400) occurring in an estimated 1:10,000 – 1:40,000 new-borns (Behan *et al.*, 2016). A precise prevalence of PCD is difficult to determine, mainly due to its diagnostics difficulties, as a consequence of the heterogeneous clinical manifestations and subsequent laboratorial examination (Boon *et al.*, 2013).

PCD (also called immotile cilia syndrome), was first described by Afzelius (Afzelius *et al.*, 1975; Afzelius, 1976). PCD is a genetically heterogeneous, autosomal recessive disease that affects the normal cilia function. As the motility apparatus of cilia is made of several components that can be affected differently, PCD patients present multiple combinations of phenotypes (clinical manifestations), with a varied range of severities, which change with age (Goutaki *et al.*, 2016; Leigh *et al.*, 2016; Rubbo and Lucas, 2017). The major clinical characteristics of PCD are recurrent airway infections. Infertility and organ laterality defects are also a common feature (Goutaki *et al.*, 2016). Kartagener syndrome (KS) is characterized by the combination of *situs inversus* (reversal of the internal organs), chronic sinusitis, and bronchiectasis (Leigh *et al.*, 2009), and occurs in about 50% of PCD cases.

1.2.1. General aspects of clinical characteristics

PCD is consequence of anomalies in motile cilia that lead to a motility dysfunction. As described in previously cilia motility is required for multiple vital functions from airway defence mechanisms to reproduction, thus patients may present diverse clinical characteristics. Notwithstanding, upper and lower respiratory tract symptoms, due to ineffective mucus clearance, is considered a hallmark of PCD patients. Clinical manifestations occur since birth and the disease progresses to a chronic stage, with actual treatments being directed to ameliorate symptoms and improve the respiratory capacity (Behan *et al.*, 2016; Goutaki *et al.*, 2016; Jackson *et al.*, 2016). Even though by distinct

mechanisms, PCD and cystic fibrosis are both characterized by impaired mucociliary clearance, and thus by recurrent respiratory infections. Further, in both conditions those infections are typically caused by *Haemophilus influenzae, Staphylococcus aureus, Streptococcus pneumoniae* and *Pseudomonas aeruginosa* and both share some clinical features (Wijers *et al.*, 2017). Therefore, lacking a standard treatment, most treatment practices in PCD are primarily based on medical personal experiences and extrapolations from cystic fibrosis care (Barbato *et al.*, 2009). However, this is not straightforward. For instance, in cystic fibrosis mucous viscosity was shown to improve, as well as the overall lung function, with nebulization's with a mucolytic agent, such as recombinant human DNAse I (rhDNAse) or N-acetylcysteine (NAC). However, in PCD patients this treatment had the opposite effect (Rubbo and Lucas, 2017).

About half of PCD patients additionally present sub-fertility or complete infertility. Males with PCD show total/partial sperm immobility or azoospermia (absence of sperm in the ejaculate). Structural deficiencies may arise at different components of the sperm flagellum axoneme, which explain the different degrees of sperm immobility. The rete testis conducts testicular sperm to the efferent ducts and these to the epididymis. As testicular sperm are mainly immotile, cilia present in the rete testis and efferent ducts are responsible by fluid flow and sperm transport. When the function of these cilia is compromised, sperm is not transported to the epididymis, and the patient presents with a picture of obstructive azoospermia (Munro *et al.*, 1994; Boon *et al.*, 2013; Pereira *et al.*, 2015; Pereira *et al.*, 2017)..

Females with PCD may also develop infertility. Dysfunction of motile cilia in the Fallopian tube may favour the occurrence of ectopic pregnancy or cause abnormal gamete transport towards the uterine cavity (Lurie *et al.*, 1989; Lyons *et al.*, 2006; Raidt *et al.*, 2015).

The action of motile monocilia in the embryonic node is responsible for nodal flow. The leftward nodal flow, working together with cascades of signalling molecules and ions, establish the left-right axis of the internal organs (Wright, 2001; Okada *et al.*, 2005; Hirokawa *et al.*, 2006). Laterality defects are thus a common feature of PCD. The most frequent laterally defect is *situs inversus totalis*, which has a prevalence of 1:10,000 in the

population, and occurs in about 50% of the PCD cases (Leigh *et al.*, 2009; Leigh *et al.*, 2016; Rubbo and Lucas, 2017).

Another feature of PCD, although much rare than the previous ones, is the hydrocephalus. Hydrocephalus, which is an accumulation of cerebrospinal fluid in the brain, caused by the loss of ependymal flow (Greenstone *et al.*, 1984; Pérez-Fígares *et al.*, 2001; Sapiro *et al.*, 2002; Lee, 2013). This condition was observed quite often in mouse models of motile cilia defects such as Mdnah5-/- mice which lack an ODA heavy chain in their cilia (Ibañez-Tallon *et al.*, 2002; Fernandez-Gonzalez *et al.*, 2009), but is rare in humans possibly due to anatomical or neuroanatomy differences between the species.

1.2.2. Existing diagnostic approaches

Given the complexity and heterogeneity of PCD, there is no single standard diagnostic test for PCD and diagnosis usually vary from centre to centre as it requires several technically demanding, sophisticated investigations not equally available to all (Jackson *et al.*, 2016; Rubbo and Lucas, 2017). As a result, clinical diagnosis may be largely under-diagnosis in some places, which leads difficulties in the comparisons among different PCD studies and hard to make standardization of diagnostic and therapeutic options (Jackson *et al.*, 2016; Rubbo and Lucas, 2017). In an attempt to minimize this problem the European Respiratory Society (ERS) launch a guideline for the diagnosis of PCD, in order to diagnostic be divided in the following categories: a 'definite positive diagnosis', a 'highly likely diagnosis', an 'inconclusive diagnosis' or 'highly unlikely diagnosis' (Rubbo and Lucas, 2017).

Following a suspicion in clinical evaluation (Behan *et al.*, 2016; Jackson *et al.*, 2016; Rubbo and Lucas, 2017), patients should perform several tests to validate this suspicion: nasal nitric oxide (NO) measurement; evaluation of cilia motility by high-speed video microscopy analysis (HSVMA) or determination of the variation in ciliary beat axis and ciliary deviation; evaluation of the ultrastructure of the axoneme of cilia by transmission electron microscopy (TEM); and genetic analysis. According to the ERS guidelines, a definite diagnosis corresponds to the presence of an ultrastructural defect (DN, CPC) in the cilia axoneme or of bi-allelic mutations in PCD-related genes; and a highly likely

diagnosis is made by the combined presence of low nasal NO levels and a dysfunction in ciliary motility (Rubbo and Lucas, 2017).

a. Nitric oxide measurements

Nitric oxide (NO) is produced by the NO synthase (NOS) family and in respiratory system is found at higher concentrations in the mucosa of the paranasal sinuses, a group of four paired air-filled spaces that surround the nasal cavity (Corbelli and Hammer, 2007). The physiological function of this high levels of NO in upper airways is remains to be fully elucidated but it seems to play a role in local host defence by its antimicrobial action, in the regulation of ciliary motility, and in the regulation of bronchial tone and pulmonary vascular resistance (Corbelli and Hammer, 2007).

Regarding PCD, nitric oxide was found to be markedly reduced in most PCD patients (Wodehouse *et al.*, 2003; Walker *et al.*, 2012). Therefore, NO measurements were proposed as a valuable a screening tool. However, some limitations to their use have been proposed, namely in small children whose specificity of the test could be unacceptably as children are unable to perform a breath-holding procedure properly (Collins *et al.*, 2016). Further some PCD patients have normal levels of NO (Walker *et al.*, 2012).

b. Evaluation of cilia ultrastructure and motility

Cilia can be collected by means of nasal brushings that allows a direct visualization of the cells allowing to evaluate its ultrastructure and motility (Rutland *et al.*, 1982).

In most PCD patients' cilia and/or flagella reveal ultrastructural defects on its axoneme, the *bigwig* of motility. Multiple ultrastructural defects underlying PCD have been characterized, among them the most common are defects in DA (either outer or inner, or both) microtubule transpositions, radial spokes defects, and the absence of DRC (nexin links) (Noone *et al.*, 2004; Leigh *et al.*, 2016; Pereira *et al.*, 2017). In addition, it was also found in PCD patient's absence or dislocation of the central microtubules (Chemes *et al.*, 1998; Chemes, 2000; Chemes and Rawe, 2003).

Nevertheless, about 20% of the PCD patients show an abnormal ciliary function, but present a normal axoneme ultrastructure, (Schwabe *et al.*, 2008; Pifferi *et al.*, 2010; Knowles *et al.*, 2012; Horani *et al.*, 2013). The mechanisms by which this happen are still

not understood Further, if is a inflammation and/or infection, which can alter the normal 9+2 arrangement, TEM could lead to an incorrect diagnose (Knowles *et al.*, 2013a). Additionally, TEM analysis may be accompanied with the determination of the variation in ciliary beat axis and ciliary deviation, which gives a quantitative estimation of the severity degree in ciliary motion abnormalities(Rutland *et al.*, 1982). TEM analysis require expensive equipment and highly skilled professionals to perform it correctly, therefore is not accessible to all centres.

The ciliary/flagellar motility by HVMA, could be used as a complementary or alternative analysis to TEM. HVMA allows the determination of both CBF and ciliary beat pattern, whose is proper function is critical to several vital functions, as discussed previously. Notwithstanding that importance, cilia motility dysfunction analysed by HVMA may not fully represent the patient cells motility as it could be influenced by any damage caused during sampling, inflammation of the epithelia cells, and by the transport as cilia motility is influenced by temperature and pH (Sutto *et al.*, 2004; Raidt *et al.*, 2014; Sears *et al.*, 2015) , thus puzzling the diagnostic. If an abnormal motility is observed, for these reasons is necessary to reanalyse cilia/flagella a second time to confirm that abnormalities are due to a congenital defect. To that, cells could be obtained either by performing a repeated brushing or by culture of the epithelial cells, which have been considered as a helpful tool in PCD diagnostic (Jorissen *et al.*, 2000; Hirst *et al.*, 2010).

c. Genetics

A recent remarkable study used an elegant affinity proteomic approach, specifically systematic tandem affinity purifications coupled to mass spectrometry, to unveil the ciliary proteome constitution that authors named ciliary protein landscape (Boldt *et al.*, 2016). Boldt and co-workers, identified 1,319 proteins and 4,905 interactions that count to human ciliary landscape (Boldt *et al.*, 2016). Of those, the majority were not previously associated to ciliary proteome, as well as novel protein-protein interactions not previously described to act together, with much of those their role in ciliary biology is still to defined. Their work, as well as, as other elegant works (Yuan and Sun, 2013; Blackburn *et al.*, 2017) highlights the great complexity of the genetic study in ciliopathies.

PCD is not an exception. PCD is a rare and genetically heterogeneous disease following the autosomal-recessive disease inheritance mode. Multiple genes were consequently associated to PCD (Table 1).

Given that the typical diagnostic of PCD is the absence of DA, the investigations into the genetic basis of PCD have started with analysis of DA proteins using a classical genetic approach, namely candidate gene approach and linkage analysis. ODA defects are among the most common defect in PCD affecting about 37 % to 40 % of PCD patients (Hornef *et al.*, 2006; Loges *et al.*, 2008; Zariwala *et al.*, 2013a). In contrast, axonemal disorganization associated to IDA defects are less frequent accounting up to 14% of all PCD cases (Antony *et al.*, 2013; Zariwala *et al.*, 2013a). Defects in both ODA and IDA, were estimated to be the reason for about 10% of PCD cases (Knowles *et al.*, 2013a; Zariwala *et al.*, 2013a). Defects in other structures such as CPC and RS defects were anticipated to justify a reduced number of cases, collectively not more than 10% of all PCD (Kott *et al.*, 2013; Zariwala *et al.*, 2013a).

Although several genes can lead to a PCD phenotype some gene present a higher prevalence, namely (by crescent prevalence order): *DNAH5, DNAH11, CCDC39, DNAI1, CCDC40, CCDC103, SPAG1, ZMYND10, ARMC4, CCDC151, DNAI2, RSPH1, CCDC114, RSPH4A, DNAAF1, DNAAF2, and LRRC6. DNAH5* is considered the most frequent mutated gene in PCD (estimated prevalence from 15 to 30%), while *LRRC6* accounting only to 1% of the cases (Zariwala *et al.*, 2013a; Davis *et al.*, 2015; Leigh *et al.*, 2016) .

The first gene associated to PCD was *DNAI1*, which is an axonemal dynein IC gene, found in the ODA, and localized on 9p13-p21 (Pennarun *et al.*, 1999). Mutations in *DNAI1* have been identified in patients with PCD, whose ODA are defective or even absent (Pennarun *et al.*, 1999; Guichard *et al.*, 2001; Zariwala *et al.*, 2006). This gene has been extensively study but a lower occurrence was reported, with no more than 10% of PCD patients caring causative mutations (Failly *et al.*, 2008; Djakow *et al.*, 2012).

DNAH5, the major disease-causing gene in PCD, localized at 5p15.2, is a gene known to encodes for components from the HC of the ODA (Kispert *et al.*, 2003; Hornef *et al.*, 2006; Morillas *et al.*, 2007; Djakow *et al.*, 2012; Davis *et al.*, 2015). Pathogenic variants in this gene were among the first ones to be linked not only with cilia immotility and PCD

but also with left-right asymmetry (Olbrich *et al.*, 2002; Hornef *et al.*, 2006). Null dnah5 mice showed normal IDA and absent ODA (Ibañez-Tallon *et al.*, 2002) corroborating the findings in PCD patients and suggesting that *DNAH5* is important to the function of the ODA complex.

DNAH11 (dynein axonemal heavy chain 11) encodes a ciliary outer dynein arm (ODA) protein and according to Zariwala *et al.* (2013a) pathogenic variants in this gene accounts to up to 9% of PCD patients, thus is the second disease causing gene in PCD. A particularity of mutations in DNAH11 is that leads to a normal ciliary, but a reduced waveform amplitude and hyperkinetic beating pattern (Bartoloni *et al.*, 2002; Schwabe *et al.*, 2008; Zuccarello *et al.*, 2008; Pifferi *et al.*, 2010; Knowles *et al.*, 2012; Lucas *et al.*, 2012).

Two other well-known PCD genes are the *CCDC39* and *CCDC40* collectively accounts to up to 10% of PCD phenotypes (Blanchon *et al.*, 2012; Antony *et al.*, 2013). The human *CCDC39* gene is localized at 3q26.33 and mutations were observed to cause failure to correctly assemble of IDA complexes, disorganization of the microtubule doublet, anomalies in the DRC, as well as, in the CPC and the RS, thus causing axonemal disorganization that is responsible for the motility dysfunction observed in patients with *CCDC39* pathogenic variants (Merveille *et al.*, 2011). The *CCDC40* gene, located at 17q25.3, was described as critical not only for cilia motility but also to left-right axis specification, as knockdown of CCDC40 in morpholino-injected embryos produced laterality defect (Becker-Heck *et al.*, 2011). Mutations in *CCDC40* gene were found in association with a heterogeneous pattern of axonemal defects, including disorganization of the microtubule doublets, absent or shifted central pair, reduction in the mean number or absence of IDA, and abnormal RS and nexin links. Nevertheless, in most cases the ODA appeared normal (Becker-Heck *et al.*, 2011; Blanchon *et al.*, 2012; Antony *et al.*, 2013; Pereira *et al.*, 2015; Sui *et al.*, 2016; Yang *et al.*, 2018). Moreover, CCDC40 protein appears to physically interacts with CCDC39 protein, forming a complex that may serve as a part of the axoneme structural scaffold, possibly as a new DRC component or as docking sites, along the doublet microtubules (Becker-Heck *et al.*, 2011). CCDC39 andCCDC40 proteins were proposed to be involved in N-DRC assembly or microtubule attachment. and are believed to also play a role in IDA attachment (Becker-Heck *et al.*, 2011; Merveille

et al., 2011). Therefore, in contrast with *DNAH5*, which plays a major role in ODA defects, *CCDC39* and *CCDC40* represent the major cause of PCD with IDA defects and axonemal disorganisation (Blanchon *et al.*, 2012; Antony *et al.*, 2013).

PCD is genetically highly complex with multiple genes associated (Table 1; (Leigh *et al.*, 2009; Knowles *et al.*, 2013a; Pereira *et al.*, 2014)), and most of those genes are composed by multiple exons (as an example, *DNAH5* contains 79 exons). initial classical genetic approach, namely resorting to Sanger sequencing, is no longer feasible (Pereira *et al.*, 2015), needing high-throughput techniques, namely NGS.

Table 1.1-1. List of main genes known to be associated to PCD phenotype and the main axonemal ultrastructural defects showed by PCD patients carrying pathogenic variants in those genes.

Gene	Main molecular function	Main ultrastructural defects found	*Sample analysed by TEM*	*Situs inversus*?[1]	Fertility problems	References
ARMC4	Axonemal docking and targeting of ODA components	Marked reduction of ODA	RC	Yes (8/12)	NA	Hjeij *et al.* (2013)
		Absence of ODA		Yes (4/5)	NA	Onoufriadis *et al.* (2013b)
				NA	NA	Davis *et al.* (2015)
CFAP298	DA assembly	Absence of ODA/IDA	RC	Yes (112/295)	NA	Austin-Tse *et al.* (2013)
CCDC39	Assembly of DRC and IDA	Defects of DA and RS	RC	Yes (10/22)	OAS	Merveille *et al.* (2011)
		Absence of IDA		Yes (8/25)	AZS	Blanchon *et al.* (2012)
		Defects of IDA and CPC		Not referred	NA	Davis *et al.* (2015)
		Defects in IDA and MT doublets disorganisation		Not referred	AZS	Antony *et al.* (2013)
CCDC40	Assembly of DRC and IDA	Misplacement of the CPC and defective IDA and DRC	RC	Yes (11/16)	NA	Becker-Heck *et al.* (2011)
		Absence of ODA	RC and flagellum	Yes (4/8)	AZS	(Blanchon *et al.*, 2012)
		Defects of IDA and CPC	RC	NA	NA	Davis *et al.* (2015)
		Defects in IDA and MT doublets disorganisation	RC	NA	AZS	Antony *et al.* (2013)
CCDC103	Fundamental factor for DA assembly	Partial loss of ODA complexes	RC	Yes (3/10)	NA	Panizzi *et al.* (2012)
		Absence of both DA	RC and flagellum	Yes (2/6)	OAS	Pereira *et al.* (2015, 2017 in peer review)
		Absence of ODA	RC	Yes (2/3)	NA	Casey *et al.* (2015)
				NA	Not referred	Zariwala *et al.* (2013b)

Legend: TEM-Transmission electron microscopy; 1- the number of patients referred to be affected in the total number of the patients analysed.; CPC- central pair complex; DA- dynein arms; ODA- outer dynein arms; IDA-inner dynein arms; OAS-oligoasthenospermia; MT- microtubules; RS-radial spokes; NA not analysed or not referred; No-means that no patients do not show situs inversus or fertility problem.

Table 1.1-1. (continuation)

Gene	Main molecular function	Main ultrastructural defects found	*Sample analysed by TEM*	*Situs inversus?*	Fertility problems	References
CCDC114	Component of the ODA docking complex	Absence of ODAs	Respiratory cilia	Yes (6/16)	No	Onoufriadis *et al.* (2013a)
				No	NA	Knowles *et al.* (2013b)
				NA	NA	Davis *et al.* (2015)
CCDC151	Required for ODA assembly	Absence of ODA	Respiratory cilia	Yes (NA)	NA	Hjeij *et al.* (2014)
DNAAF1 (LRRC50)	Cytoplasmic preassembly of DA	Marked reduction of both ODA and IDA	Respiratory cilia	Yes (2/3(?))	NA	Loges *et al.* (2009)
		Absence of ODA and IDA		NA	NA	Davis *et al.* (2015)
DNAAF2 (KTU)	Cytoplasmic preassembly of DA	Absence or defects of both DA	Respiratory cilia and flagellum	Yes (2/3)	NA	Omran *et al.* (2008)
		Absence of DA	Respiratory cilia	NA	NA	Davis *et al.* (2015)
DNAAF3	Assembly of axonemal DA and dynein complexes	Absence of both DA	Respiratory cilia	Yes (4/10)	NA	Mitchison *et al.* (2012)

Legend: TEM-Transmission electron microscopy; 1- the number of patients referred to be affected in the total number of the patients analysed.; CPC- central pair complex; DA- dynein arms; ODA- outer dynein arms; IDA-inner dynein arms; OAS-oligoasthenospermia; MT- microtubules; RS-radial spokes; NA not analysed or not referred; No-means that no patients do not show situs inversus or fertility problem.

Table 1.1-1. (continuation)

Gene	Main molecular function	Main ultrastructural defects found	*Sample analysed by TEM*	*Situs inversus?*	Fertility problems	References
DNAH5	Important for function of the ODA complex	Absence of ODA	Respiratory cilia	Yes (7/17)	NA	Olbrich *et al.* (2002)
				NA	NA	Kispert *et al.* (2003); Hornef *et al.* (2006); Kano *et al.* (2016)
		Absence of both DA		Yes (6/9)	NA	Djakow *et al.* (2012)
				Yes (49/89)	NA	Failly *et al.* (2009)
DNAH9	Essential for ODA type 2 assembly	Distal absence of ODAs	Respiratory cilia	Yes (5/5)	NA	Loges *et al.* (2018)
DNAH11	Encodes a ciliary ODA protein	Normal axoneme ultrastructure	Respiratory cilia	Yes (11/20)	NA	Knowles *et al.* (2012)
				Yes (2/3)	NA	Pifferi *et al.* (2010)
				Yes (1/14)	NA	Schwabe *et al.* (2008)
				Yes (NA)	NA	Bartoloni *et al.* (2002); Lucas *et al.* (2012)
DNAI1	Important for function of the ODA complex.	Absence of ODA	Respiratory cilia	Yes (NA)	NA	Pennarun *et al.* (1999)
				NA	NA	Zariwala *et al.* (2006); Failly *et al.* (2008); Djakow *et al.* (2012)

Legend: TEM-Transmission electron microscopy; 1- the number of patients referred to be affected in the total number of the patients analysed.; CPC- central pair complex; DA- dynein arms; ODA- outer dynein arms; IDA-inner dynein arms; OAS-oligoasthenospermia; MT- microtubules; RS-radial spokes; NA not analysed or not referred; No-means that no patients do not show situs inversus or fertility problem.

Table 1.1-1. (continuation)

Gene	Main molecular function	Main ultrastructural defects found	*Sample analysed by TEM*	*Situs inversus?*	Fertility problems	References
DNAI2	Assembly of ODA complexes	ODA defects	Respiratory cilia	Yes (62/105)	Yes (1/49)	Loges *et al.* (2008)
DNAL1	Involved in the interaction of the axonemal LC 1 with DHC and tubulin.	Absence or markedly shortened ODA.	Respiratory cilia	Yes (48/86)	NA	Horváth *et al.* (2005); Mazor *et al.* (2011)
DRC1	Role in the assembly of DRC, IDA and RS.	Defects DRC structure	Respiratory cilia	Yes (1/20)	NA	Wirschell *et al.* (2013)
DYX1C1	Important for axonemal dynein assembly	Disruptions of ODA and IDA	Respiratory cilia	Yes (5/12)	Yes	Tarkar *et al.* (2013)
HEATR2	Assembly or transport of DA	Absence of DA	Respiratory cilia	Yes (4/9)	Yes (2/6)	Horani *et al.* (2012)
				NA	NA	Davis *et al.* (2015)
HYDIN	Encodes a CPC-associated protein	Defects in CPC	Respiratory cilia	NA	AZS	Olbrich *et al.* (2012)
NME8	Critical role in the ODA assembly.	Partial lack and reduction of ODA	Respiratory cilia	YES (NA)	NA	Duriez *et al.* (2007)
RSPH1	A radial-spoke-head protein	CPC and RS defects	Respiratory cilia	NA	Yes	Kott *et al.* (2013); Knowles *et al.* (2014); Onoufriadis *et al.* (2014)

Legend: TEM-Transmission electron microscopy; 1- the number of patients referred to be affected in the total number of the patients analysed.; CPC- central pair complex; DA- dynein arms; DLC- dynein light chain; ODA- outer dynein arms; IDA-inner dynein arms; HC- heavy chain; OAS-oligoasthenospermia; MT-microtubules; RS-radial spokes; NA not analysed or not referred; No-means that no patients do not show situs inversus or fertility problem.

Table 1.1-1. (continuation)

Gene	Main molecular function	Main ultrastructural defects found	*Sample analysed by TEM*	*Situs inversus?*	Fertility problems	References
RSPH4A	A radial-spoke-head protein	Abnormalities in RS	Respiratory cilia	NA	NA	Castleman *et al.* (2009); Daniels *et al.* (2013)
RSPH9	A radial-spoke-head protein	Abnormalities in CPC and RS	Respiratory cilia	NA	NA	Castleman *et al.* (2009); Davis *et al.* (2015)
SPAG1	Assembly and/or trafficking of the axonemal DA	Defects in ODA and IDA	Respiratory cilia	Yes (9/14)	NA	Knowles *et al.* (2013c)
ZMYND10	Required for both DA assembly	Absence of ODA and IDA	Respiratory cilia	Yes (4/10)	Yes (8/10)	Moore *et al.* (2013)
				Yes (8/15)	Yes (1/15)	Zariwala *et al.* (2013b)

Legend: TEM-Transmission electron microscopy; 1- the number of patients referred to be affected in the total number of the patients analysed.; CPC- central pair complex; DA- dynein arms; DLC- dynein light chain; ODA- outer dynein arms; IDA-inner dynein arms; HC- heavy chain; OAS-oligoasthenospermia; MT-microtubules; RS-radial spokes; NA not analysed or not referred; No-means that no patients do not show situs inversus or fertility problem.

1.3. Application of next-generation sequencing (NGS) in PCD

1.3.1 NGS in clinical genetics

From all the developments in medicine that have been attained since the beginning of the 21st century, genetics is probably among the most promising and fast-growing fields, but probably has not reached its full potential yet.

Genetics in health care basically started as a way to address the clinical uncertainty around such rare and complex diseases, as it provides the most accurate and definitive diagnoses. It has even revolutionized the diseases classifications being now progressively known by the genetic entity involved. In practical terms, based on establishing a genetic diagnosis, it is possible to predict the disease outcome and further guidance towards the best possible options of care for the patient or even genotype-based therapies if available. A genetic diagnosis also opens new possibilities to prevent disease recurrence for those afflicted families, starting by carrier screening but also allowing the best reproductive options, through gametes selection or preimplantation or prenatal genetic diagnosis.

From the DNA three-dimensional structure (the alpha helical structure) determined by X-ray crystallography (Watson and Crick, 1953), it took nearly two decades for the development of techniques to determine the DNA sequence at the single-base level. The noble prize winners Paul Berg (Jackson *et al.*, 1972), Frederick Sanger (Sanger and Coulson, 1975) and Walter Gilbert (Maxam and Gilbert, 1977), made possible the developments to perform Sanger sequencing. Sanger sequencing is a method based on the generation of DNA fragments labelled radioactive nucleotides and later with a fluorescent dye. These PCR products were then resolved in large acrylamide gels or later detected by automated DNA sequencers (capillary electrophoresis and associated with a laser and optics detection system) for size separation and to create electropherograms.

The first automated sequencer that was commercially available was Applied Biosystems Prism 373. This system was launched in 1987 and relies on the automated Sanger method. Later, ABI PRISM 3700 DNA Analyzer, allowed the completion of the human genome (Venter *et al.*, 2001), which opened a new road in clinical genetics

accelerating the detection of variants and consequently open horizons for the implementation of target screening for highly penetrant germline mutations, such cystic fibrosis (Zielenski *et al.*, 1991) and sickle cell anaemia (Saiki *et al.*, 1985). This method had dominated genetics for the last two decades and is still considered the "gold-standard" for genetic to confirm a suspected diagnosis. However, when complex disease involving dozens of genes need to be study, simultaneous analysis of all these genes by traditional sequencing is not possible

During the last decades, new and decisive steps were taken with the development of new methods to improve (and eventually replace) this Sanger method, collectively referred as next-generation sequencing (NGS). This technology made possible to analyse dozens of genes and produce a massive amount of data per instrument run, in a faster and cheaper way. A fast pace growing list of companies are developing machines to perform NGS: Illumina, Ion Torrent, BGI-Shenzhen, Qiagen, PacBio and Oxford Nanopore Technologies.

The typical NGS workflow can be subdivided into the primary, secondary and tertiary analysis (Fig. 1.3-1). The primary data analysis consists of the detection and analysis of raw data. Follows secondary analysis, where the reads are aligned against the reference human genome (or de novo assembled) and the calling is performed. The last step is the tertiary analysis, that includes the steps critical to address the biological context of the variants, also known as "making sense" or data interpretation. This third step includes the variant annotation, variant filtering, prioritization, data visualization and reporting.

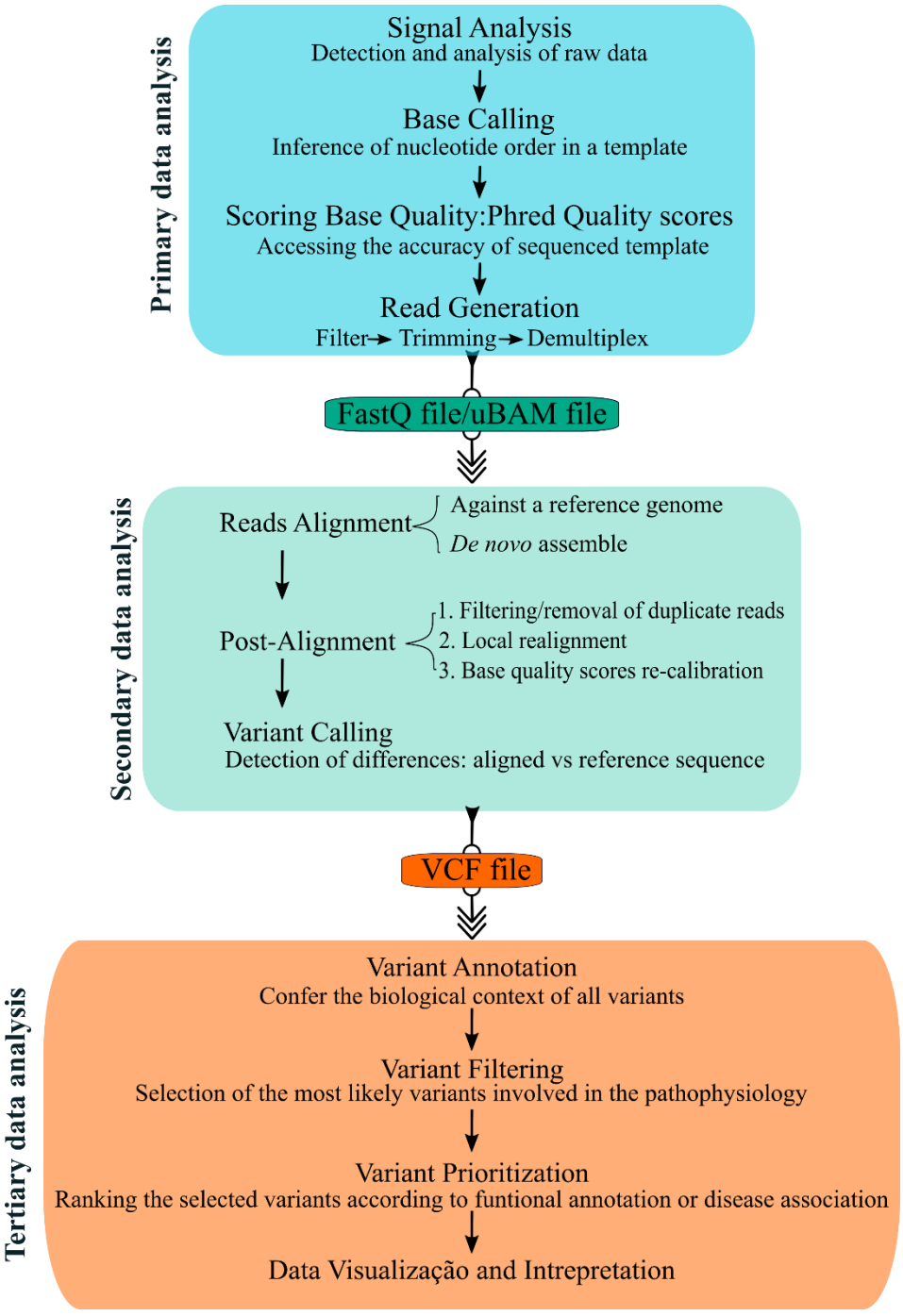


Figure 1.3-1. Schematic overview of the main steps of NGS. The NGS bioinformatics is subdivided in the primary (blue), secondary (green) and tertiary (orange) analysis. The darker green and darker orange are, respectively, files generated during the primary analysis and secondary analysis and that contains information important to procedure the secondary and tertiary analysis.

Prior to start an NGS experiment is crucial to be very clear regarding the biological question as the as sample preparation and the proper choice of workflow/platform are dependent on that question. The typical scenario in clinical genetics is the study of DNA variants that could cause a pathogenic phenotype, namely single nucleotide variants

(SNV), small insertion or deletion events (indels) or other types of large structural variants (SV) that may occur in genome. In this case, a DNA sample is used to investigate the whole genome (Whole Genome sequencing (WGS)) or to investigate the coding regions and its boundaries, through whole exome sequencing (WES). Within last years a great discussion has emerged concerning which - WES or WGS - is better for clinical diagnostics (Fig. 1.3-2).

WGS allows a complete examination of SNVs, Indels and SV both in coding and non-coding regions of the genome, whereas WES is targeted to coding regions and neighbouring intronic sequences (donor or acceptor splice-sites) and lacks the non-coding regions and regulatory regions. Coverage uniformity with WGS is superior to WES, likely due to differences in the hybridization efficiency of sequence capture probes in WES, which results in target regions that have little or no coverage. WGS do not have the limitation of reference bias, whereas WES do, since capture probes tend to preferentially enrich reference alleles at heterozygous sites producing false negative SNV calls (Belkadi *et al.*, 2016; Meienberg *et al.*, 2016).

No doubt WGS is more powerful and able to detect efficiently all types of genomic variants (Belkadi *et al.*, 2016), however, it comes with the drawback of higher sequencing costs, great informatics requirements both in terms of hardware (i.e. to store the data) and software (i.e. to handle the data), longer turnaround time, and great personal requirements since expert bioinformaticians are required to deal with harder and complex data analysis. Taking in consideration that WES is much cheaper, faster, requires less computational demanding and the data analysis is simpler comparing with WGS, as well as, the fact that 85% of the known disease-causing variants are located within the coding sequences, is comprehensible that this approach, is not common on clinical basis and is limited to gene discovery projects in large genome sequencing centres or service providers. Consequently, WES has emerged has a powerful approach for researchers elucidating genetic variants underlying human diseases (Antonarakis *et al.*, 2000; Bamshad *et al.*, 2011; Lee *et al.*, 2014; Jamuar and Tan, 2015).

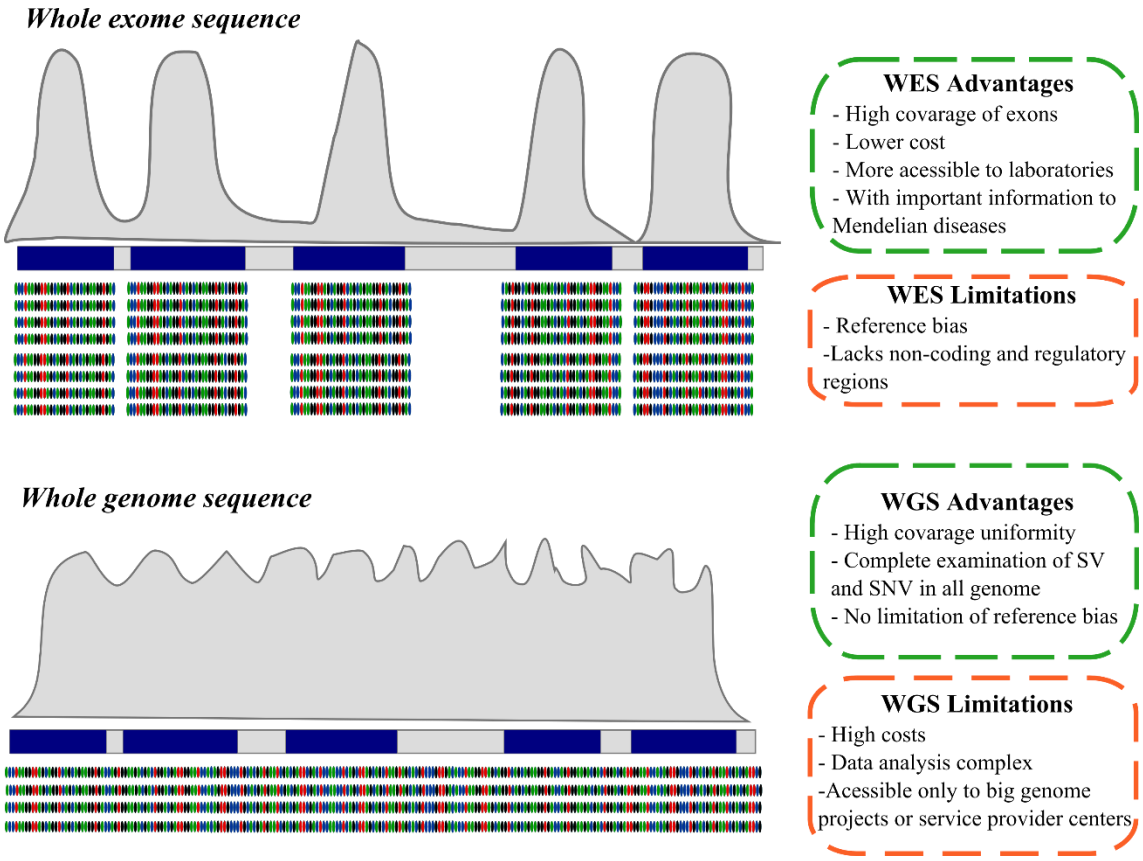


Figure 1.3-2. Overview of the pros&cons of WGS and WES in clinical practice.

Notwithstanding, WES generate hundreds of variants, which is still a great number of variants to deal. Accordingly, researchers have been working on approaches that narrow even more the NGS analysis. Bearing in mind that there is a great risk of losing information, in some specific cases, is possible to sequence a selected group of genes or genomic regions that were potentially implicated in a diseases/case study, known as gene panels. Consequently, gene panels offer a high sequence quality at even lower costs and with a reduction in the size of data files and analysis time comparing to WES (Mendeliome Group, 2015). These advantages are of upmost importance in clinical practice where a faster diagnostic with lower costs are needed to make this accessible to more laboratories and, so to more patients. Besides the risk of losing information, a major limitation of gene panels is the difficulty in defining the appropriate list of genes for a given phenotype owing to the remarkable heterogeneity of clinical features. Illumina, Ion Torrent and other companies have commercially available gene panels for a couple of diseases (e.g. cardiopathies, neurologic disorders, muscular dystrophies) and to

reproductive pre-natal diagnostic, but significant differences could be found between them (Xue *et al.*, 2015). Further, is also possible to customize gene panels and create one that fits a specific study or research aim. An important example of a gene panel is the Mendeliome panel, a sequencing-based multiplexing assay that was designed to facilitate the screening of patients with Mendelian disorders (Mendeliome Group, 2015).

For certain cases the analysis of WES or WGS are probably not enough as further understanding about the consequences of a pathogenic variant at other molecular level is needed. Nowadays, the recent advances in NGS technologies enable to combine multiple -omics and study disease in a system integrative way (Ohashi *et al.*, 2015; Tebani *et al.*, 2016). Particularly, the analyse the differential gene expression through transcriptome analysis (RNA analysis), called RNA-seq(uencing) (Wang *et al.*, 2009). NGS also allows more complex studies such as study of protein – DNA interactions, known as ChIP-Seq (Park, 2009), well as, the study of DNA patterns of methylation, called bisulfite-seq (Li and Tollefsbol, 2011).

1.3.2 NGS as a tool to unveil PCD genetic aetiology

PCD, as mentioned before, is a highly heterogenous diseases with a myriad of clinical phenotypic characteristics that are translated into a range of disease severities, which are closely related with the complex ciliary and flagellar structure and the sophisticated molecular mechanisms that enable its proper function. Cilia and flagella molecular and genetic composition is concordantly extremely complex. Recent studies applying recent NGS tools have demonstrated that cilia are composed by more than 1,000 proteins (Yuan and Sun, 2013; Boldt *et al.*, 2016; Blackburn *et al.*, 2017). Bearing that in mind, is reasonable to think that hundreds of genes may be involved and are able to theoretically lead to PCD if mutated.

Regarding the molecular genetic diagnosis of PCD to be considered conclusive, needs to identify in suspected PCD patients biallelic disease causing gene variants in PCD associated genes (Jackson *et al.*, 2016). Despite the vast number of potential candidate PCD genes, at the moment, PCD specific gene list (i.e. list of genes in which disease-causing variants were reported in PCD/KS patients) accounts with no more than 40 genes (Table 1; (Zariwala *et al.*, 2013a; Paff *et al.*, 2018)). However, those genes do not

justify all cases as about 40% of PCD patients present a negative or inconclusive genetic result after searching in this list (Kim *et al.*, 2014). Moreover, even in the most prevalent genes such as *DNAH5*, that justifies around 15% to 30% of PCD cases, more than 90 disease-causing gene variants were already described (Pereira *et al.*, 2019). An additional challenge to PCD genetic diagnosis is that rarely the same variant is observed in unrelated patients and most of PCD genes are larger genes with several exons (for instance *DNAH5* and *DNAH11* are composed of 79 exons and 82 exons, respectively). Altogether, finding the genetic aetiology among so many genes with several exons are unfeasible, time-consuming, laborious and expensive using the traditional Sanger sequencing approach (Berg *et al.*, 2011; Pereira *et al.*, 2015). Therefore, WES emerge as valuable tool in PCD diagnostic as it is faster and more cost-effective than traditional Sanger sequencing. Gene panels can also be good cost-effective option although given what was stated above about PCD complexity, the risk of negative or inconclusive genetic result is higher than in WES. Several commercial gene panels are already available in market typically with 30 to 35 genes, with some variations in the gene list. There is also the possibility, to customize this list of genes. In both WES and gene panels, whenever possible, "trios" analysis (i.e. the simultaneous analysis by NGS from the patient and the parents) should be performed, given the autosomal disease inheritance pattern is very helpful in analysis of NGS data to allow effective variant filtration. Consequently, a good option could be start analysis with gene panels and then, if a negative result, progress with WES. WES could be also followed by targeted CNV analysis, which increase the success rate to up to 76% (Marshall *et al.*, 2015).

Even though, the success rate is not 100%, and there is a propose from UK's 100,000 Genomes project that diagnostic of PCD will be initially by WES or a large panel analysis, which, if not give a positive result, will be followed by WGS (Wheway and Mitchison, 2019). Notwithstanding, WGS provides many opportunities to improve our understanding of ciliopathies and will undoubtedly give an answer to many unsolved cases, however it is still expensive, and the analysis is complex to be applied routinely in a clinical base.

A major challenge, however, to clinicians and researchers regarding NGS and PCD, is to correlate the findings with the relevant medical information, which may not be a

simpler task. In PCD is hard to find a same variant in different non-consanguineous patients, even with a very similar phenotype, and newer variants are frequently reported. Further, several new genes are believed to be still associated to PCD phenotype, which will require additional efforts to validate the pathogenicity of variants (which in a clinical setting may not be feasible). More importantly, both clinicians and patients must be clearly conscious that a positive result, although providing an answer that often terminates a long and expensive diagnostic journey, does not necessarily mean that a better treatment will be offered nor that it will be possible to find a cure (Biesecker and Green, 2014). Although some efforts are being done, namely a recent work showed a novel transfection protocol, using lentiviral shRNA from *DNAH5* gene, that provides a prolonged transgene expression in vivo with significantly reduction in toxicity and immunogenicity (Munye *et al.*, 2016), which will be of great value for further studies. Nevertheless, having a genetic cause identified could still be important to orient the treatment.

Chapter 2

Methodology

2. Methodology

2.1. Clinical data

2.1.1. Adults

Infertile male patients: five patients with total sperm immotility

Patient's clinical features and the sperm detailed ultrastructural analysis was previously reported in detailed (Sousa *et al.*, 2015). Briefly, one patient (thereafter referred as **male Patient 1**, but in Sousa *et al* (2015) was identified as patient 5) presented *situs inversus*, severe respiratory symptoms (nasal polyps, chronic sinusitis, rhinitis and bronchitis), all clinical features compatible with KS. This 45 years old male from northern Portugal, went for infertility treatments as the couple did not achieve a spontaneous pregnancy. Further, the karyotype was normal and there were no Y chromosome microdeletions, cystic fibrosis transmembrane receptor (CFTR) mutations or adverse CFTR polymorphisms, and hormonal values were in the normal range. Semen analysis showed a volume of 5.6 ml, a pH 7.8, and presence of very rare total immotile sperm. The patient undertook testicular sperm extraction (TESE) in November 2006, followed by intracytoplasmic sperm injection (ICSI) with fresh testicular in-situ motile sperm. No pregnancy ensued. In 2008 he performed another TESE cycle with fresh sperm but again no pregnancy ensued. A third attempt was scheduled with cryopreserved testicular sperm and a successful twin pregnancy was achieved with birth of two male twins, who are now 8 years old (born in 2010) and are perfectly healthy.

Male patient 2, 4 and **5** (in Sousa *et al* (2015) were named as patient 1, 3 and 4, respectively) presented dysplasia of the fibrous sheath (DFS), which is a clinical condition that arise during spermiogenesis as a failure of a cytoskeletal structure surrounding the sperm axoneme named fibrous sheath. These defects occur in association with abnormalities of other flagellum structures, namely the axoneme(Eddy *et al.*, 2003). **Male patient 3** showed also DFS but had a different presentation, which was associated with an intact annulus and disorganization of doublets in his sperm axoneme.

Female KS patient

We also included in our study a fertile female patient of 53 years old from northern Portugal. She is currently followed by the department of Otorhinolaryngology from S. Sebastião Hospital, Hospital Centre of entre Douro e Vouga. She reported chronic respiratory complaints, namely bronchiectasis and chronic rhinosinusitis, and *situs-inversus totalis*, all clinical features well-suited to suspect of KS. This patient was thereafter referred as **female patient 2.**

2.1.2. Children

Three unrelated children were included in this work. They were followed at the Maternal Child Centre of North/ Hospital and University Centre of Porto (CMIN/CHUP), due to upper respiratory complains and *situs inversus totalis*. Parents are healthy and non-consanguineous, and pregnancies and deliveries occurred without problems. As for the parents of **child 1** are 45 years (mother) and 52 years (father). The present age of the parents of **child 2** are 32 years (mother) and 36years (father). Regarding **child 3**, his mother is 37 years old and presents asthma controlled with daily budesonide /formoterol fumarate dihydrate, and the father is 40 years old and has allergic rhinitis. Child's 3 mother developed pre-eclampsia at 30 weeks of gestation, needing 2 cycles of antepartum betamethasone.

Neither children present neonatal transient tachypnoea, neonatal pneumonia (meconium aspiration), and signs of surfactant protein deficiency (absence of respiratory insufficiency at birth) or interstitial lung disease. Respiratory function tests were not performed due to age. The alpha-1-antitrypsin assay, the immunological study and the atopy test were normal. Cystic fibrosis was also clinically excluded.

Child 1

This female child, currently with 7 years of age (born in 2011), was born at 39 weeks of gestation by vaginal delivery. At birth, she presented an Apgar score of 8/10, a birth weight of 3190 g (z-score -0.09), length of 50 cm (z-score 0.46) and head circumference measure of 35 cm. She was referred to the consultation at 16 months of age due to acute bronchiolitis of difficult resolution, followed by recurrent respiratory infections, some requiring hospitalization due to exacerbation. In the first moments of life, cardiac sounds

were observed on the right side, with chest X-ray evidencing an inversion of the cardiac silhouette and gastric chamber on the right, raising the suspicion of a *situs inversus totalis*. Later, she developed poor weight evolution, recurrent acute otitis media (AOM), mucopurulent rhinorrhoea and productive daily cough, with no signs of asthma. During development, the child did not acquire chest deformities or digital clubbing, discarding chronic respiratory insufficiency development, and the pulmonary auscultation remained normal. Currently, she weighs 20,4 Kg (z-score -0.61), has 120 cm height (z-score -0.15) and a head of circumference of 50 cm (z-score 0). She is under daily respiratory kinesiotherapy and sporadic cycles of antibiotic therapy.

Child 2

This male child, with 7 years of age (born in 2011), was born at 39.5 weeks of gestation by vaginal delivery (with a suction cup). At birth, he presented and Apgar score of 10/10, a birth weight of 2958 g a (z-score -0.84), a length of 52 cm (z-score 1.12) and a head circumference measure of 34.5 cm (z-score 0) and a length of 52 cm. He was hospitalized with respiratory distress syndrome (up to day 13) and right opacification on chest X-ray compatible with right pneumonia, without any microbiological agent being identified. He presented clinical improvement after antibiotic therapy cycle. During hospitalization, cardiac sounds were found on the right during auscultation. Chest X-ray confirmed the inversion of cardiac laterality with the presence on the right of the gastric chamber, raising the suspicion of *situs inversus totalis*. During the neonatal period (up to day 28), virus screenings in secretions were negative. Up to about 2.5 months of age, control chest X-ray did not evidence opacities compatible with pneumonia, having developed nasal obstruction, productive cough, purulent nasal secretions, without wheezing, fever or AOM. From 1 year of age, he developed sporadic and self-limiting crises of a productive cough and wheezing, without fever, AOM or mucopurulent rhinorrhea. During development, the child did not attain chest deformities or digital clubbing, which is compatible with the absence of chronic respiratory insufficiency, and the pulmonary auscultation remained normal. Currently, he weighs 27,4 Kg (z-score 1.25), has121 cm (z-score -0,14) and a head of circumference of 52.5 cm (z-core 1.25). He is under daily respiratory kinesiotherapy.

Child 3

This male child, with 2 years and 9 months of age (born in 2016), was born at 38.1 weeks of gestation by vaginal delivery, without complications. At birth, he presented an Apgar score of 8/9/9, a birth weight of 2970 g (z-score 0), a length of 47 cm (z-score 1) and a head circumference measure of 38 cm (z-score 2).

By 1 month of age, he developed acute bronchiolitis, without any microbiological agent being identified, and no need of hospitalization. Since then, he presented persistent productive cough and wheezing, generally without fever, with slight improvement in summer. At Pediatric consultation cardiac sounds were found on the right during auscultation, he was sent by 6 months of age to be followed at the hospital. During the follow-up, cardiac sounds were confirmed on the right-side during auscultation and the electrocardiogram showed *situs inversus*, dextrocardia, dextroapex. Due to age, a chest CT scan was not performed. The abdominal ultrasound revealed *situs inversus*. The ecocardiogram was normal. Respiratory function tests were not performed due to age. The alpha-1-antitrypsin assay, the immunological study (immunoglobulins), the atopy test (Phadiatop infant) and the sweat test were normal. During evolution, the child did not develop chest deformities or digital clubbing, which is compatible with the absence of chronic respiratory insufficiency, and the pulmonary auscultation remained normal between the respiratory agudizations.

Actually, he presents bilateral otitis serosa, persistent purulent nasal secretions and episodic wheezing. At the last examination (2 years e 9 months), he weighs 14 Kg (z-score 1.5), has 93 cm (z-score 2.5) and a head Circumference of 51.5 cm (z-score 3). Currently, he is under daily inhaled corticosteroids and salbutamol in crisis.

2.2. Biological sample collection

2.2.1. White blood cells

White blood cells (WBC) samples, used for DNA, RNA and protein extraction, from both patients, their unaffected parents (whenever possible) and healthy controls, were collected from peripheral blood in EDTA tubes (VACUETTE, Porto, Portugal). Genomic DNA was extracted from peripheral blood leukocytes from patients, their unaffected

parents (whenever possible) and from healthy controls following the salting out method (Miller *et al.*, 1988), quantified by a NanoDrop spectrophotometer ND-1000 (Version 3.3; LifeTechnologies; California, United States of America (USA))and stored at 4ºC.

2.2.2. *Respiratory epithelial cells*

Respiratory epithelial cells were obtained by nasal brushing from male patient 1, female patient 2 and children, children's parents and healthy controls, using a cytology soft sterile brush (Endobrush, Biogyn SNC, Mirandola, Italy), in both nostrils from the inferior nasal turbinate(Rutland *et al.*, 1982). Briefly after a nasal wash with a saline solution to remove the mucus, the brushing was performed in both nostrils by means of circular movements from the inferior nasal turbinate. Excess mucus in the brush was removed and discarded. The obtained samples were placed in fixative for TEM, or in RPMI 1640 Medium (Gibco™, Thermo Fisher Scientific, Massachusetts, USA), for RNA and immunofluorescence analysis. The children were premedicated with oral paracetamol (15 mg/kg/dose), about 20 minutes before harvesting. All individuals were continuously monitored for vital signs, pain complaints and/or signs of bleeding.

2.2.3. *Reproductive samples*

Excedentary testicular tissue from patient-1 and from men with conserved spermatogenesis (controls) was obtained under infertility treatments. Ejaculated sperm from normozoospermic men was obtained from cases undergoing spermiogram evaluation. Excedentary oocytes were obtained under infertility treatments. Those reproductive cells were kindly provided by the centre for Reproductive Genetics Professor Alberto Barros (Porto, Portugal) and Hospital Center of Vila Nova de Gaia/Espinho (CHVNG/E), according to the Guidelines of the Local, National and European Ethical Committees.

Sertoli cells were obtained from primary cultures originated from testicle biopsies of men with conserved spermatogenesis (Bernardino *et al.*, 2018), kindly provided by . Raquel Bernardino (PhD) and Ana Maria Silva (MSc) from biology and genetics of reproduction group from UMIB/ICBAS.

2.3. Sample processing for transmission electron microscopy

For TEM analysis of respiratory epithelium and testicular sperm samples were fixed in 2.5% glutaraldehyde in 0.1 M cacodylate buffer (pH 7.4) for 3 h (room temperature) and then stored overnight at 4ºC in cacodylate buffer. If settling was incomplete at any of these stages, the tube was centrifuged gently. The samples were then post-fixed with 2% osmium tetroxide in buffer containing 0.8% potassium hexacyanoferrate for 2 h at 4ºC and dehydrated in a graded ethanol series. At the final dehydration step, the samples were treated with 1% tannic acid in 100% ethanol for 10 minutes and embedded in epoxy resin (Epon). Suitable areas of ciliated cells were in semithin sections (1µm) stained with methylene blue-Azur II. Ultrathin sections were cut on a LKB-ultramicrotome and retrieved on copper grids. They were double- contrasted with aqueous uranyl acetate and lead citrate, observed and photographed in a JEOL 100CXII transmission electron microscope, operated at 60kV.

2.4. Evaluation of cilia morphology and orientation

To assess the internal axoneme structure, electron micrographs of transverse sections of cilia were taken with a final magnification of 40,000–53,000x. At least 100 sections were observed. The following characteristics were evaluated: presence or absence of a 9+2 pattern, ODA and IDA, radial spokes, CPC (central microtubule pair, central bridge, central sheath). A diagnosis was based on the presence of a systematic defect (Afzelius and Srurgess, 1985). Signs of infection, if present, were quantified, which included: compound cilia (several axonemes inside a common cytoplasmic membrane), naked cilia (partial or total absence of the cytoplasmic membrane), blebs (cytoplasmic membrane projections) and disorganized microtubules (disorganization of the 9+2 pattern).

To assess the ciliary beat axis, electron micrographs were taken at a magnification of 13,000-16,000x, with a minimum of 100 transverse sections examined after printing (26,000-32,000x final magnification). In the printed images, a line was drawn which passed along the central microtubules. A reference line was then chosen based on the

main orientation of the lines drawn. The reference line is drawn perpendicular to the central microtubules and passes between them. The angle of each line is calculated to the reference line and the mean determined. Each angle is then subtracted to the mean. The mean of these differences is near or equal to zero. The standard deviation (SD) of these differences corresponds to the ciliary beat axis (De Iongh and Rutland, 1989).

2.5. Polymerase chain reaction and Sanger sequence

The gene regions of interest regions were amplified, using the primers previously designed (see Attachment 2)., by polymerase chain reaction (PCR), in a thermocycler of models 9700, 9800 or VERITI (LifeTechnologies). The PCR conditions were listed in attachment 1 and the PCR reaction mixture contained: 15µl of PCR Master Mix (Promega, Madison, USA); 12 µl of sterile bidistilled water, 1 µl of each primer at 10 pmol/µl (Thermo Fisher Scientific, Einsteinstrasse, Germany or NZYTech, Lisbon, Portugal) and about 50 ng of DNA. For sporadic regions, a High-GC PCR condition, was applied which the reaction mixture includes 12,5 µl of PCR Master Mix (Promega); 1,5 µl of sterile bidistilled water; 5 µl of Betaine 5M (Sigma-Aldrish, St. Louis, USA); 1,5 µl of DMSO (Bioline, London,United Kingdom (UK)), 1 µl of each primer at 10 pmol/µl

Following successful PCR amplification, reactions were enzymatically purified using Illustra ExoStar kit (GE Healthcare- Buckinghamshire, UK). Subsequently a new asymmetric PCR was prepared based on Sanger sequencing method, using BigDye Terminator v1.1 Cycle Sequencing Kit (LifeTechnologies). Next these reactions were purified using a size selection column method Performa DTR (for Dye Terminator Removal) (EdgeBio, Maryland, USA). Finally, the obtained products were resolved and analysed by high resolution electrophoresis in a 3130xl genetic analyzer (LifeTechnologies).

2.6. Exome sequencing by next-generation sequencing

The exome of the **male patient 1**, the **female patient 2** and the 3 children, all with *situs inversus totalis* and clinical features that raised the suspicion of KS, were sequenced using

the AmpliSeq strategy in the Ion Proton NGS platform, according to (Oliveira *et al.*, 2015) A total of 75-100 ng of high-quality DNA from the patients were amplified with the Ion AmpliSeq™ Exome Library Preparation kit (Life Technologies). The sample was barcoded with the IonExpress Barcode Adapter (Life Technologies) to enable pooling of two exomes per chip. To evaluate the quality of the library a High Sensitivity DNA Kit in Bioanalyzer (Agilent, California, USA) was used. The library was quantified using Ion Library Quantitation Kit (Life Technologies). Subsequently, library fragments were clonally amplified by emulsion PCR using the Ion PI Template OT2 200 kit v2 and the Ion OneTouch 2 System (Life Technologies), and the positive Ion Sphere Particles enriched in the Ion OneTouch exome sequencing machine (Life Technologies). Finally, these enriched positive spheres were loaded in Ion PI chip v2 and sequenced in the Ion Proton System (Life Technologies) at Genoinseq (Biocant, Cantanhede, Portugal).

Latter reads were mapped against the human reference genome hg19 using Torrent Mapping Alignment Program (Life Technologies). Variant calling was performed by running Torrent Variant Caller plugin, using manufactures' optimized parameters for exome sequencing (Life Technologies). All variants were listed in a Variant Call Format (VCF) file that was annotated and filtered using the Ion Reporter™ Software version 5.2 (http://ionreporter.lifetechnologies.com/) and VarAFT 2.10 (http://varaft.eu). Alamut Visual v2.10 software (Interactive Biosoftware, France) assisted variant interpretation. As a quality control we defined a threshold of 20x for coverage (i.e. that is the percentage of high-quality mapped reads, higher the best) and a phred score of 10 (Q10) (i.e. means an accuracy of 90,0%) and given the technical limitations of sequencing platform (Homer, 2010), homopolymers longer than 6 bp were removed from analysis.

To analyze variants obtained, an autosomal recessive disease model was used. We firstly filter-out the more deeply intronic placed variants, synonymous substitutions and variants with allele frequency above 1% in human population databases, assuming that are improbable to be deleterious (Bamshad et al., 2011). Further, we excluded genes with only one heterozygous variant and selected the variants based on the following gene ontology (Ashburner et al., 2000) keywords: to cilia and ciliopathies, namely cilia/flagellum; primary ciliary dyskinesia, cilium movement, cilium beat, cilium

morphogenesis and organization, cilium assemble, axoneme assemble, cilium component, axoneme component, determination of left/ right axis and symmetry, cell projection, dynein arm, locomotion and development. Further, *HYDIN* gene variants were also excluded, due to the high false discovery rate in mutation calling of this gene (Pongor et al., 2015). Subsequently, variants predicted as pathogenic, by at least two bioinformatic predictors included in the filter and annotation software analysed, were selected.

All the suspected variants were manually checked on the Binary Alignment Map (BAM) file through GenomeBrowse version 2.0.2 (Golden Helix, USA) and sanger sequencing was applied to validate the candidate variants.

2.7. Structural analysis

To further understand potential molecular consequence of the identified missense compound heterozygous variants in *DNAH7*, we decided to include a structural analysis using *in silico* tools. As the molecular protein structure of any dynein coding protein, namely DNAH7, were not yet solved in humans, to preview the potential effects of variants in DNAH7 protein structure, we applied a combination of cytoplasmic dynein structural models was used: human cytoplasmic dynein-1 (PDB ID:5NUG) (Zhang *et al.*, 2017) and dynein-2 (PDB ID:4RH7) (Schmidt *et al.*, 2015), *D. discoideum* cytoplasmic dynein 1 (PDB ID:3VKH) (Kon *et al.*, 2012) and *S. cervisiae* cytoplasmic dynein 1 (PDB ID:4AKI) ((Schmidt *et al.*, 2012). Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and ESPript (http://espript.ibcp.fr/ESPript/ESPript/) were used for multiple sequence alignment and to obtain secondary structure information from aligned sequences, respectively. The figures were rendered with protein data bank (PDB) data and the PyMol program.

2.8. Runs of homozygosity (ROH)

To understand the correlation between ROH parameters comparison between heterozygous and homozygous autosomal recessive, we analysed the data obtain from HomozygosityMapper (http://www.homozygositymapper.org/) (Seelow *et al.*, 2009).

Statistical analysis performed in IBM SPSS statistics v.24 and graphic in GraphPad Prism.

2.9. Copy number variation analysis

SNP-Array analysis was performed on genomic DNA of the patient using the Affymetrix1 CytoScan HD array according to the manufacturer's instructions. Arrays were scanned with the Affymetrix GeneChip1 Scanner 3000 7G and genotypes were analysed using Affymetrix Chromosome Analysis Suite Software version 3.3 (ChAS 3.3) and Annotation Net Affx Build 32.3. Interpretation was based on human reference sequence (GRCH37/hg19, Feb. 2009).

2.10. Gene expression analysis

Total ribonucleic acid (RNA) from cell suspensions containing nasal cells, sperm, germ cells, Sertoli cells and WBC, was extracted using the PerfectPure RNA Cell & Tissue Kit (5 PRIME GmbH, Hamburg, Germany), according to manufacturer instructions. For oocytes, due to the limited cell number we used the single cell RNA purification kit (Norgen Biotek, Thorold, Canada). The concentration and purity of RNA samples were determined on a Nanodrop spectrophotometer ND-1000 (Version 3.3; LifeTechnologies). Only samples with an A260/A280 ratio between 1.8 to 2.1 were selected as it is an indicative of highly purified RNA. RNA was then converted to complementary DNA (cDNA) using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems, California, USA), according to manufacturer instructions and including the optional step of "DNase treatment" that is intended to digest genomic deoxyribonucleic acid (gDNA) to avoid gDNA contamination.

For the gene expression analysis specific PCR primers were designed (see attachment 2) and amplified by PCR, as described above. Subsequently, Real-time quantitative PCR (qPCR) was performed to evaluate the mRNA expression of the interested genes and whenever was possible compare control with patient's cells. *B2M* and *GAPDH* genes were used as housekeeping gene, to normalize gene expression levels. qPCRs were carried out in a Bio-rad CFX96 (Bio-Rad, Hercules, USA) and efficiency of the amplification was determined for all primer sets using serial dilutions of cDNA. qPCR

amplifications used 1µL of synthesized cDNA reaction containing NZY qPCR Green (NZYTech) and 5 pmol of each primer. The cDNA dilutions were adjusted based on RNA concentration (used to cDNA synthesis) to ensure that samples had an equivalent cDNA concentration. Reaction conditions were the following: 10 min denaturation at 95ºC, followed by 40 cycles at 95ºC for 30 sec and 60ºC for 30 sec. For melting curves, the conditions were 95ºC for 60 sec, 55ºC for 30 sec followed by an increment of 0.5ºC until 95ºC. Samples were run in triplicate in each PCR assay. Fold variation of gene expression levels was calculated following the mathematical model proposed by Pfaffl (2001), using the formula 2-ΔΔCt (Pfaffl, 2001). The non-parametric tests run in the GraphPad Prism (version 6.01, GraphPad Software, California, USA) to access the differences on expression levels between cells.

2.11. Immunohistochemistry

Human testicular tissue was analysed from formalin-fixed (VWR) paraffin-embedded blocks (Merck). Immunohistochemistry (IHC) was performed in tissue sections of 3 µm thickness attached to adhesive slides (Starfrost, Braunschweig, Germany). Tissue sections were deparaffinised with xylene (VWR) and serially hydrated in a decreasing scale of ethanol followed by a wash in distilled water. Heat antigen retrieval was performed with citrate buffer (Merck), pH 6, in a microwave (600W) for 20 min. Tissue sections were intensely washed in water. Endogenous peroxidases were inhibited with 3% of hydrogen peroxide (Merck) and non-specific background staining was minimized with 5% of non-fat milk (Nestlé, Vevey, Switzerland) in PBS pH 7.4 (Panreac, Barcelona, Spain), with both treatments being performed for 30 min at RT.

Sections were then incubated with primary antibodies rabbit anti-CCDC103 (Abcam, Cambridge, England) and mouse anti- acetylated α-tubulin (Santa Cruz Biotechnology, California, USA) for 60 min at RT. A negative control was included. Ultraview universal DAB anti-rabbit and anti-mouse detection kit (Ventana Medical Systems, Arizona, USA) was used to reveal expression of CCDC103 and acetylated α-tubulin antigens, respectively. After dehydration, sections were mounted on Coverquick 2000 at RT (VWR). Results were observed in a light microscope BX41 (Olympus, Tokyo, Japan).

2.12. Immunocytochemistry

For immunocytochemistry (ICC), smears were made from cell suspensions containing nasal cells, sperm, germ cells and Sertoli cells on adhesive slides (Starfrost). Cell smears were air-dried and directly frozen at -80ºC until use. Afterwards, cells were fixed with 4% paraformaldehyde (PFA, Merck) in PBS for 20 min at RT, followed by washes in PBS. For permeabilization, cells were incubated in 0.2% Triton X-100 (Sigma-Aldrich) in PBS for 15 min at RT. Then washed in PBS and incubated with 5% non-fat milk in PBS for 60 min at RT, to inhibit non-specific binding. Afterwards, cells were incubated overnight at 4ºC, with antibodies rabbit anti-CCDC103 (Abcam); rabbit anti-DNAH5, anti-DNAH7, anti-CRHR1, anti-KRT34 and anti-USP11 all from Biorbyt (Cambridge, United Kingdom); and mouse anti- acetylated α-tubulin (Santa Cruz Biotechnology). For each experiment, a negative control was included. DyLight™ 488 anti-rabbit (Biolegend, California, USA) and Texas Red anti-mouse (Santa Cruz Biotechnology) were used as secondary antibodies and applied to cells for 60 minutes at RT. Then, cells were counterstained with Vectashield mounting medium containing 4', 6-diamidino-2-phenylindole (DAPI: Vector Laboratories, California, USA). Results were observed in an epifluorescence microscope (Eclipse E400; Nikon, Tokyo, Japan) and also analyzed with a FluoView™ FV1000 laser scanning confocal microscope (Olympus, Tokyo, Japan) to screen an eventual co-localization of the signal.

2.13. Immuno-Gold Labelling

Immune gold experiences were conducted in sperm from control individuals. Cells were fixed for 1h at 4ºC, in a mixture of freshly prepared 2% PFA and 0.25% glutaraldehyde (Sigma-Aldrich) in PBS. Samples were rinsed with PBS and then with PBS in 1mM of glycil-glycine (Merck) for 10 min, 4ºC. Samples were partially dehydrated with 50%, 75%, 90% ethanol (10 min each) followed by 4 two changes in 100% ethanol (15 min each), before overnight infiltration with LR white resin (Sigma-Aldrich), RT. Cells were polymerized in a closed capsule for 24 h, 60°C. Ultrathin sections were collected on precleaned nickel grids (AgarScientific, Stansted, England).

Immunogold procedures were carried out as previously described (Kann and Fouquet, 1989).Two rabbit polyclonal antibodies were used: anti-⍟-Tubulin antibody non-tyrosinated (Merck) and anti-CCDC103 (Abcam).

The grids were pre-treated in 20 mM Tris buffer [TBS: 10mM Tris (Sigma-Aldrich), 150 mM NaC1 (VWR), pH 7.8] containing 0.1% BSA (fraction V, Sigma-Aldrich) and 10 mM glycine (Merck) and incubated for 30 min at RT. Then were incubated with antibodies diluted in TBS-0.1% BSA for 2 h, at RT. The dilutions were the following: 1:50 for anti-⍟-Tubulin antibody and for CCDC103 several dilutions were tested from 1:5 to 1:40. After a rinse in a stream of TBS-BSA, grids were incubated with secondary anti-rabbit IgG 10 nm-gold labelled antibody diluted 1:20 with TBS-BSA for 2 h, at RT. Grids were subsequently rinsed in a stream of TBS-BSA and bi-distilled water before contrasted with 2% aqueous uranyl acetate and lead citrate. The grids were then observed in a JEOL 100CXII transmission electron microscope, operated at 60kV.

2.14. Western Blot

For WBC protein extraction, the peripheral blood was incubated with an erythrocyte lysis buffer [ELB: 155mM of NH4CL (Sigma-Aldrich), 10mM KHCO3 (Merck) and 1mM of EDTA (Sigma-Aldrich), pH 7,4] for 20 min on ice and centrifuged at 2000 rcf (or G-Force) for 10 min. The pellet was washed twice in ELB and then in cold PBS, centrifuged at 3000 rcf for 5 min at 4ºC and stored at -20ºC until protein extraction.

For the sperm, germ cells and Sertoli cells, cells were washed in cold PBS, centrifuged at 3000 rcf for 5 min at 4ºC and stored at -20ºC until protein extraction. For sperm, the pellet was stored with 20 ul of 1:100 solution of protease inhibitor cocktail (Sigma-Aldrich) in PBS. For protein extraction, the pellet was resuspended in lysis buffer [125 mmol/liter Tris-HCl (pH 6.8, Sigma-Aldrich), 2% SDS (Biorad), 4M Ureia (Sigma-Aldrich), 10% β-mercaptoetanol (Sigma-Aldrich), 1:100 of protease inhibitor cocktail], and incubated with soft agitation for 30min at RT. Protein concentration was estimated using Qubit® protein assay kit (Life Technologies). After protein measurement, the protein was precipitated in trichloroacetic acid (TCA, Merck) and acetone (Merck) and stored at -20ºC until further use. Then, the protein amount (50 to 100 µg), were adjusted with NuPAGE® LDS Sample Buffer (Life Technologies). Samples were boiled for 10 min

at 75ºC and more 2 min at 90ºC and electrophoresed on 4-12% NuPAGE Bis-Tris midi gels (Life Technologies). Gels were subsequently transferred overnight at 35V, 4ºC, to Amersham Protran 0.45 NC nitrocellulose membranes (GE Healthcare, Illinois, USA), then blocked with 5% non-fat milk with 0.05% Tween-20 (Sigma-Aldrich) in TBS. Afterwards probed with rabbit anti-CCDC103 antibody (Abcam) and a mouse anti-HSP 70 antibody (Santa Cruz Biotechnology), overnight at 4ºC. Following incubation for 2 h, RT, with an HRP conjugated goat anti-rabbit or anti-mouse secondary antibody (Santa Cruz Biotechnology), reactivity was detected by chemiluminescence using a LAS3000 digital imaging system (FujiFilm, Tokyo, Japan).

Chapter 3

Results and discussion

Contents

3. Results and discussion

3.1. Sperm immotility

3.1.1. Mutation analysis in patients with total sperm immotility

Adapted from: Pereira *et al*, 2015 J Assist Reprod Genet.
https://doi.org/10.1007/s10815-015-0474-6

The results that will be presented in this section were the starting point of this thesis and were already published in *Journal of Assisted Reproduction and Genetics* (Pereira *et al*, 2015)

Background

In this work we aimed for the genetic characterization of five patients with total sperm immotility using the classical approach, Sanger sequencing, to attempt the identification of a genetic variant that could justify the ultrastructural anomalies found in sperm cells of these five patients with total sperm immotility.

One patient (**male patient 1**) presented the clinical features of KS. He had more severe respiratory symptoms (nasal polyps, chronic sinusitis, rhinitis and bronchitis), *situs inversus totalis*, and total sperm immotility due to absence of DA and nexin bridges in his sperm axoneme.

The other four patients (**male patients 2 to 5**) presented dysplasia of the fibrous sheath (DFS). DFS is an anomaly found in spermatozoa of severe asthenozoospermic patients, in which cytoskeletal components such as the fibrous sheath and outer dense fibres are affected. The main characteristic of DFS is a marked hypertrophy and hyperplasia of the fibrous sheath. Patients showed also disruption of several axonemal structures and milder respiratory symptoms, namely chronic sinusitis and chronic bronchitis.

All patients presented several anomalies in sperm cell axoneme, namely anomalies in disorganization of microtubules doublets and CPC; absence of DA and nexin links; or reduced number of RS (Fig. 3.1-1). Having that in mind, we initiated our genetic research by studying through Sanger sequencing genes that code for proteins involved in these same structures and that are associated with total sperm immotility. Accordingly, we selected seven genes: *AKAP3 AKAP4, CCDC39, CCDC40, DNAH5, DNAI1* and *RSPH1*.

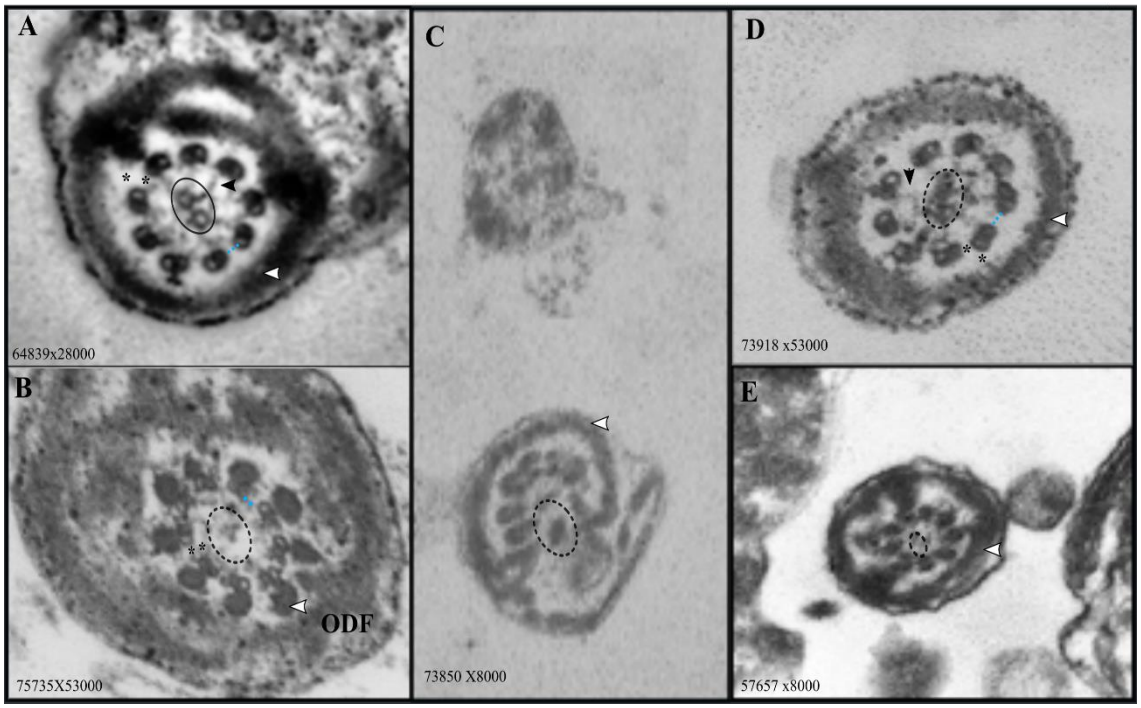


Figure 3.1-1. Ultrastructure of sperm axoneme from male patients 1 to 5, respectively A to E, showing the main ultrastructural anomalies found. **(A).** The axoneme from patient 1, shows a conserved central pair complex (circle) and radial spokes (black head-arrow); dynein arms (*) and nexin bridges (dotted blue line) are missing. White head-arrow represents the fibrous sheath. **(B)** The axoneme from patient 2, in which dynein arms (*) are present but the central pair complex and the radial spokes (dotted circle) are absent. White head-arrow represents the fibrous sheath. ODF-outer dense fibres. **(C)** The axoneme from patient 3, is disorganized presenting a reduced number of microtubules doublets; the central pair complex is present but misplaced (dotted circle) and fibrous sheath (white head-arrow) disorganized. **(D)** The axoneme from patient 4, shows absence of the central pair complex (dotted circle), a reduced number of radial spokes (black arrow), absence of dynein arms (*) and nexin bridges (dotted blue line). **(E).** The disorganized axoneme from patient 5, showing absence of the central pair complex and radial spoke (dotted circle).

Images provide by Professor Mário Sousa.

Briefly, the genes *AKAP3* and *AKAP4* are involved in the organization of the basic structure and assembly of the FS (Eddy *et al.*, 2003; Luconi *et al.*, 2010). Mutations are associated with DFS (Baccetti *et al.*, 2005a; Baccetti *et al.*, 2005b). *CCDC39* and *CCD40* genes are integral components of the DRC and the encoded proteins are essential for the assembly of dynein regulatory and IDA complexes. Mutations in these genes are associated to PCD and cause multiple disorganization of the axoneme (Becker-Heck *et al.*, 2011; Merveille *et al.*, 2011; Blanchon *et al.*, 2012; Antony *et al.*, 2013). *DNAH5* and *DNAI1* genes are essential to the function of the ODA complex and mutations are related to PCD with abnormalities or absence of ODA (Pennarun *et al.*, 1999; Olbrich *et al.*, 2002; Kispert *et al.*, 2003; Hornef *et al.*, 2006; Zariwala *et al.*, 2006; Failly *et al.*, 2008; Djakow *et al.*, 2012; Xu *et al.*, 2017). Finally, *RSPH1* gene codes a RS-head protein that is important for the proper building of CPC and RS, as mutations in the RSPH1 were observed to lead to an abnormal axonemal configuration with CPC and RS defects in PCD patients (Kott *et al.*, 2013; Knowles *et al.*, 2014; Onoufriadis *et al.*, 2014).

We designed specific primers for all exonic regions and intronic boundaries of the selected genes, with exception of the gene *DNAH5*. Due to the elevated number of exons, we concentrated our sequencing efforts in forty out of the total of seventy-nine exons composing this gene that are known to harbour the majority of the pathogenic variants (Djakow *et al.*, 2012).

Sanger Sequencing

After this initial screening, a total of 251 genetic variants were analysed. Most of the variants found were already described in online databases as polymorphisms (frequency in the global population being >1%), with only 13 being rare or new variants (Table 3.1-1).

The patient with *situs inversus* presented 43 DNA variants in the selected genes, with any variant with a predicted pathogenic impact.

In patient 2, it were observed 43 polymorphisms, 4 rare variants, 3 in the *DNAH5* gene (c.1537-100_1537-99delTT, c.3835-3delT and c.10282-81delT) and 1 in the *DNAI1* gene (c.81+61A>G), and one novel variant in the *CCDC39* gene (c.2540A >G). The new

heterozygous variant c.2540A>G, is located in the exon 18 of the *CCDC39* gene (NM_181426.1), and leads to a change, in the protein position 847, of glutamic acid by glycine. Both bioinformatic tools, Polyphen-2 and SIFT1, predicted this amino acid change as non-pathogenic. The rare heterozygous variant c.81+61A>G was located in intron 2 of the *DNAI1* gene and increases by 53.56% the score of the splice site acceptor (SSA, i.e. the splice site at the 3' end of an intron), meaning that a similar SSA may be generated, which may confuse the splicing machinery and disrupt the normal splice. The variants c.1537-100_1537-99delTT (heterozygous patient-2 and homozygous in patient 5) and c.10282-81delT (only detected in patient 2, and in heterozygosity) were both located at the *DNAH5* gene, and both were predicted as non-pathogenic. Regarding the rare variant c.3835-3delT that was found in the *DNAH5* gene, observed in all patients with DFS (patients 2 to 5), it is a heterozygous deletion of a T nucleotide three bases before exon 25. From analysis by five splice algorithms incorporated in Alamut Visual V2.4 software, it was foreseen that this variant, although not affecting the critical AG dinucleotide sequence at the 3' splice site, may lead to a slight reduction of the score of native SSA. This may likely disrupt the proper recognition by splicing components, and thus the normal protein function. DFS is thought to follow an autosomal recessive inheritance (Escalier and Albert, 2006), and therefore, even assuming the hypothesis of its pathogenicity, this heterozygous mutation may not be enough to cause the phenotype showed by the patients, namely the absence of DA. However, it can contribute to an effect on the flagellar movement. Notwithstanding the robustness of bioinformatic tools, expression analyses are always important to validate their predictions and to evaluate the effect of the variants on splicing. Unfortunately, these studies were not possible as patients did not provide further cell samples.

In patient 3 we found 50 DNA variants, 48 were polymorphisms and two were rare variants in the *DNAH5* gene, c.3835-3delT and c.7408-84_7408-83delAT. The rare variant c.7408-84_7408-83delAT (homozygous in this patient-3 and heterozygous in patient 5) was classified as benign, after the bioinformatic analysis (see Table 3.1 1) .

Patient 4 presented 51 DNA variants in selected genes. Of those, four are rare variants, one in the *CCDC39* gene (c.233G>A) and three in the *DNAH5* gene (c.1537-102T>A,

c.3835-3delT, and c.10872+84T>C), and three are new variants, one in the *CCDC40* gene (c.2620-92C>T), and two in the *DNAH5* gene (c.5882+133A>G and c.11570+124G>C). Apart from the variant c.3835-3delT, already discussed, all variants found in this patient were predicted by bioinformatic tools as polymorphisms.

Regarding the last patient, patient 5, we identified 51 variants, being four considered rare variants. Three of these four rare variants are in the *DNAH5* gene that were also observed in the other patients (c.1537-100_1537-99delTT, c.3835-3delT, c.7408-84_7408-83delAT); the other rare variant was found in the *CCDC40* gene (c.2682G>A), which was detected only in this patient and in heterozygosity. However, excluding variant c.3835-3delT, all variants in this patient represent polymorphisms without predicted pathogenic impact.

Table 3.1 1 Rare and novel DNA sequence variants found by Sanger sequencing and respective bioinformatic analysis

Gene	Patient	DNA change (Location)	Freq.*	Protein change	Bioinformatic analysis			
					PP- 2	SIFT	HSF	Mutation Taster
CCDC39	4 (He)	c.233G>A (exon 3)	0.56%	p.Arg78His	Benign (s=0.00)	T (s=0.28)	No effect on Splice	P (p=0.90)
	2 (He)	**c.2540A>G (exon 18)**	**New variant**	**p.Glu847Gly**	**Benign (s=0.17)**	**T (s=0.06)**	**New ESE wt: 72.9 mt: 82.9**	**DC (p=0.69)**
CCDC40	5 (He)	c.2682G>A (exon 16)	0.61%	p. (=)	-	-	No effect on Splice	P (p=0.90)
	4 (He)	c.2620-92C>T (Intron 15)	New variant	-	-	-	No effect on Splice	P (p=0.9)
DNAH5	4 (**Ho**)	c.1537-102T>A (Intron11)	ND	-	-	-	No effect on Splice	P (p=0.9)
	5 (**Ho**) 2 (He)	c.1537-100_1537-99delTT (Intron 11)	ND	-	-	-	No effect on Splice	P (p=0.9)
	2-5 (He)	**c.3835-3delT** (intron24)	**0.1%**	**-**	**-**	**-**	**New SSA mt:85.29 nt:85.85**	**DC(p=1)**
	4(He)	c.5882+133A>G (Intron 35)	New variant	-	-	-	No effect on Splice	P (p=0.9)
	3 (**Ho**) 5 (He)	c.7408-84_7408-83delAT (intron 44)	ND				No effect on Splice	P (p=0.9)
	2 (He)	c.10282-81delT (Intron60)	ND	-	-	-	No effect on Splice	P (p=0.9)
	4 (He)	c.10872+84T>C (Intron63)	ND	-	-	-	No effect on Splice	P (p=0.9)
	4 (He)	c.11570+124G>C (Intron67)	New variant	-	-	-	No effect on Splice	P (p=0.9)
DNAI1	**2 (He)**	**c.81+61A>G (Intron2)**	**0.3%**	**-**	**-**	**-**	**SSA wt: 54.03 mt: 82.97 nt: 83.19**	**DC (p=0.98)**

Human Splice Finder (HSF): http://www.umd.be/HSF/; PolyPhen-2 (PP-2): http://genetics.bwh.harvard.edu/pph2/; SIFT: http://sift.jcvi.org; and the MutationTaster: http://doro.charite.de/MutationTaster/ . Scores higher than 0.5 in Polyphen are treated as damaging and in SIFT scores less than 0.05 are considerate deleterious. In MutationTaster a probability (p) closer to 1 indicates a high 'security' of the prediction. *Freq.- Frequency within control population (based on Exome variant server and SNP database from NCBI); At blue bold the variants most promising in terms of predicted pathogenic impact.

Abbreviators: He: Heterozygous; Ho: Homozygous; c.-coding sequence, A-Adenine; G-Guanine, T-Thymine, C-Cytosine, ND: variant listed in databases but with a non-determined frequency, p.-protein sequence, Arg-Arginine, Hist-Histidine, Glu- Glutamic acid, Gly- Glycine, Thr-Threonine, Ser-Serine, s: score; p-probability, T-tolerated, D-Damaging, P-Polymorphism, Dc- Disease causing, SSA- Splice Site Acceptor, SSD- Splice Site Donor, wt- reference score (i.e. the score of non:mutated sequence), mt- mutant score (i.e. the score of mutated sequence), nt-native splice site (i.e. normally occurring splice site), ESE- Exonic Splice Enhancer.

In DFS patients (2 to 5), no candidate variants that may contribute to the patient's phenotype were identified. We had anticipated to found mutations in *AKAP3* and *AKAP4* given their roles in sperm motility (Luconi *et al.*, 2010) and previous studies that demonstrated that male mice lacking *AKAP4* were infertile due to reduced sperm motility, and that the FS was disrupted (Miki *et al.*, 2002). Further, deletions in *AKAP3/4* genes were associated to a diffuse signal on immunostaining for human AKAP4 protein in DFS patients (Baccetti *et al.*, 2005a; Baccetti *et al.*, 2005b). However, our work did not add further evidence towards the hypothesis that *AKAP3/4* are the genetic causes of the DFS phenotype. Our results are in agreement with other two reports, that did not detected gene mutations in these genes (Turner *et al.*, 2001; Moretti *et al.*, 2007). Altogether, these results demonstrate that the role of AKAP3/4 is still unclear and leads us to hypothesize that DFS, is also genetically complex as is PCD, showing that a classic gene by gene sequencing is not effective. DFS, may be a multigenic disease that related genes are not yet fully identified. Likely, those genes may be related with FS function and/or assembly; or could be caused by spermatogenic genes related to formation, transport or attachment of FS components. Further studies about molecular elements and mechanisms of intraflagellar transport of cell factors involved in human FS development and structure need to be done to fully understand the genetic causes of DFS.

Whole exome sequencing

Considering the phenotype of patient 1 (lack of DA and nexin bridges), would suggest a genetic basis well defined as compared with the other patients. However, using Sanger sequencing only polymorphic variants were detected. Given the complexity of dyneins, as briefly discussed in introduction of this thesis and as well as the complexity of the axoneme components, is comprehensible that this genetic analysis is not straightforward, and it is conceivable that several genes may be involved. As a result, a WES analysis was performed in this patient.

The WES run generated a total of 37,737,917 sequence reads, 99% of which were efficiently aligned against the hg19 human reference genome. The exome target regions were on average covered 108.9 times, with 93% having a coverage superior to 20%. These

values are above the defined threshold and thus acceptable for this study. In contrast to Sanger sequencing, from WES were identified 49,230 single nucleotide variants (SNV) and 3,479 insertions/deletions, which were listed in the Variant Call Format (VCF) file. To analyse such a great number of variants, we applied filter strategies (see methodology). Briefly, we selected rare or new variants located in exonic regions and those placed within intronic regions near exons (8bp from exon) from genes whose gene ontology was compatible with the patient phenotype (including PCD, axonemal component, flagellar/ ciliar motility). The synonymous variants were evaluated using bioinformatic tools and excluded for further analysis since no effects on splicing were predicted. Sixteen variants were select to be visually inspected on the BAM file GenomeBrowse version 2.0.2. (Golden Helix), with nine being considered sequencing artefacts and seven were selected for Sanger sequencing confirmation: c.104G>C in *CCDC103*; c.262_263delCC in *INSL6*; c.3167A>T in *DNAH6*; c.*7895C*>T in *DNAH10*; c.828C>G in *GAS8*; c.4445G>T in *SPAG17*, and c.8delC in *MYCBPAP* gene.

The variant in *MYCBPAP* gene was not confirmed by Sanger and 6 candidate variants were confirmed by Sanger sequencing. Of those, only the heterozygous variant in *SPAG17* was predicted to be a polymorphism by all bioinformatic tools. The remaining (Fig. 3.1-2; Table 3.1-2) were foreseen as potentially pathogenic.

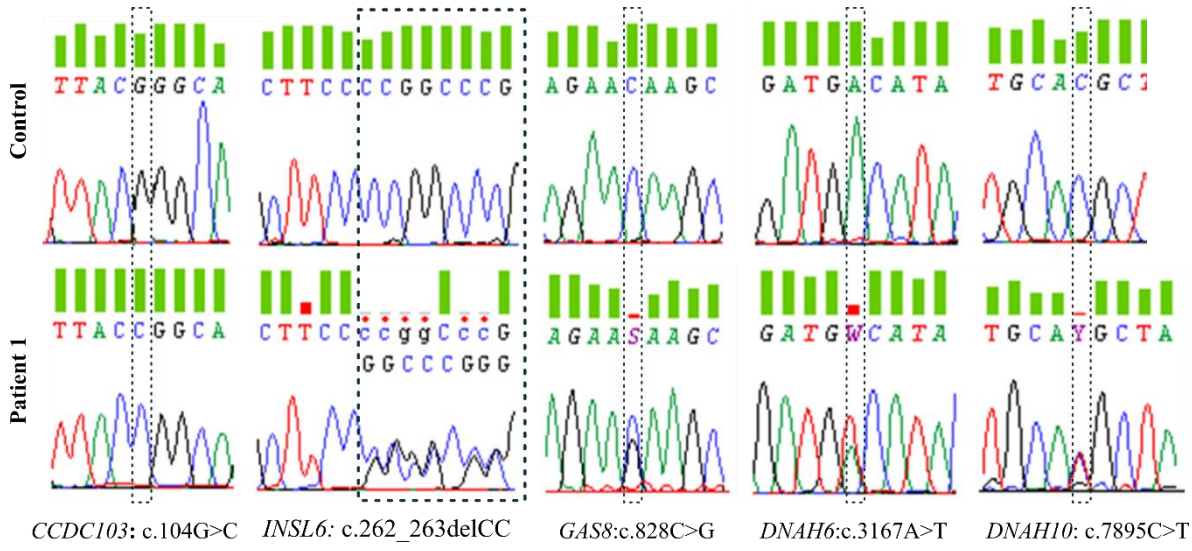


Figure 3.1.2. Sanger sequencing electropherograms of the variants selected from exome sequencing analysis from patient 1.

Table 3.1-2. List of the variants detected by exome sequencing and its respective bioinformatic analysis

Gene	Variant/ Zygoty	Protein Change	Freq.*	Bioinformatic analysis		
				SIFT	**PolyPhen2**	**Mutation Taster**
CCDC103	c.104G>C Homozygous	p.Arg35Pro	**New**	Damage (s=0)	Damage (s= 0.99)	*Disease Causing* (p=0.99)
INSL6	c.262_263delCC Heterozygous	p.Pro88Glyfs*27	**New**	na	na	*Disease Causing* (p=1)
DNAH10	c.7895C>T Heterozygous	p.Thr2632Met	0.13%	Damage (s=0.02)	Damage (s= 0.98)	*Disease Causing* (p=0.99)
DNAH6	c.3167A>T Heterozygous	p.Asp1056Val	**New**	Damage (s=0)	Damage (s= 0.98)	*Disease Causing* (p=0.99)
GAS8	c.828C>G Heterozygous	p.Asn276Lys	**New**	Tolerated (s=0,38)	Possible Damage (s=0.72)	*Disease Causing* (p=0.99)
SPAG17	c.4445G>T Heterozygous	p.Arg1482Leu	**New**	Tolerated (s=0,23)	Bening (s=0.005)	P (p=0.9)

SIFT: http://sift.jcvi.org ; PolyPhen-2: http://genetics.bwh.harvard.edu/pph2/ and the MutationTaster: http://doro.charite.de/MutationTaster/ .Scores higher than 0.5 in Polyphen are treated as damaging and in SIFT scores less than 0.05 are considered deleterious. In MutationTaster a probability (p) closer to 1 indicates a high 'security' of the prediction. Freq.*- Frequency within control population (based on Exome variant server (http://evs.gs.washington.edu) and SNP database from NCBI (http://www.ncbi.nlm.nih.gov/snp. Abbreviators: a, in silico prediction; D, Damaging; s, score; p, probability; na, nonapplicable .

GAS8 (actually named dynein regulatory complex subunit 4) was shown to be required for normal motility of cilia (Yeh *et al.*, 2002) being associated with tail maturation and attachment of DRC to microtubules (Bekker *et al.*, 2007; Colantonio *et al.*, 2008). The variant found in this gene occurs in a highly conserved central region (p.N276K, pfam number: 13851) and was predicted as possibly pathogenic. However being heterozygous, it likely is not sufficient to cause patient's phenotype

The missense variants c.7895C > T in *DNAH10* and c.3167A > T in *DNAH6* are both located in genes encoding a dynein heavy chain (HC) protein. Each HC comprises a N-terminal domain; and a C-terminal segment that includes a highly conserved central section with six AAA modules and a microtubule binding domain (see Introduction). According to the NCBI Conserved Domain Database, the variant in the *DNAH10* gene

codes for a residue located in the third AAA module, a highly conserved region, which contains a P-loop (P3) with an ATP binding site, in which mutations were showed to block ATP binding and hydrolysis, which could ultimately may affect motility (Silvanovich *et al.*, 2003). The variant in *DNAH6* gene corresponds to an amino-acid change in a residue located in the N-terminal region, which is important to specify the intracellular location of the dynein isoform by the binding of accessory proteins (Asai and Koonce, 2001). To our knowledge, neither of these two dynein HC genes have been associated with sperm motility in humans. An RNAi knockdown of *DNAH10* gene in *Trypanosoma brucei* lead to flagellum immotility and axoneme structural defects (Zukas *et al.*, 2012). Considering the inheritance mode of PCD/KS, by itself both variants are unlike to lead to absence of DA and consequently to lack of motility. Further studies in animal models are needed to analyse the combined effects of different genes that code for the same structure to understand the real effects of both variants and its possible interaction.

The *INSL6* gene, located at chromosome 9p24, belong to the relaxin family of peptide hormones whose members are involved in several reproductive functions (Lok *et al.*, 2000; Anand-Ivell *et al.*, 2013; Anand-Ivell and Ivell, 2014) . *INSL6* deficiency was shown, in mice, to cause an arrest of spermatogenesis at late stages of meiotic prophase, which lead to a reduction in sperm production and immotility, and thus proposed that the *INSL6* gene are required for normal spermatogenesis (Burnicka-Turek *et al.*, 2009). A heterozygous missense mutation in exon 2 of *INSL6* gene was found in a patient with spermatogenic failure, and the authors proposed that this mutation might be causal, due to the probable disruption of INSL6 pro-hormone processing, even in the heterozygous state (Chen *et al.*, 2011). We identified a novel heterozygous deletion of two C nucleotides in the coding region of *INSL6* (c.262_263del). Despite being heterozygous, this variant disrupts the reading frame, and thereby predictably affects the protein function. Thus, although not explaining the absence of both DA, ultimately, and according to previous reports about its gene, this deletion might compromise spermatogenesis and hence contribute to total sperm immotility and/or account for the reduced number of sperm cells in the patient semen.

The homozygous variant in the *CCDC103* gene, c.104G > C is predicted to be damaging and the mutated residue is highly conserved. This variant causes a change of an Arginine by a Proline in the protein α-helix region that contains the dynein attachment factor N-terminus domain. The wild-type Arginine is a polar amino acid that contains a side chain consisting of a 3-carbon straight chain ending in a guanidine group that is positively charged. This amino acid is frequent in protein binding sites, as the positive charge enables to interact with negatively charged groups, forming multiple hydrogen bonds (Betts and Russell, 2003). In contrast, the mutant amino acid Proline is smaller, has a neutral charge and is hydrophobic. The side chain of proline is a cyclic structure, named pyrrolidine ring, that confers nonpolar characteristics. This characteristic side chain is non-reactive and gives proline an exceptional conformational rigidity. Consequently, proline is not commonly involved in protein function (Betts and Russell, 2003), suggesting that this mutation may impact on the protein structure, namely preventing dynein binding (Venselaar *et al.*, 2010). These gene was proposed to be a DA attachment factor (Panizzi *et al.*, 2012; King and Patel-King, 2015). *CCDC103* mutations have been detected in PCD patients with absence of IDA and/or ODA, *situs-inversus* and paralysis of respiratory cell cilia beat (Panizzi *et al.*, 2012; Shoemark *et al.*, 2017). Those reports, however, did not focus on the patients' fertility, and thus, this is the first report of a DNA sequence variant in *CCDC103* with a predicted pathogenic impact associated with sperm motility in a PCD patient with total sperm immotility, absence of DA and *situs inversus totalis*.

Overall, both the new homozygous missense variant in *CCDC103* and the frame-shift variant in *INSL6* are the most likely to lead to, respectively, the absence of DA and total sperm immotility. *CCDC103* was selected to be further studied (shown in the following sections of this thesis).

In conclusion, this part of the Thesis highlights the difficulties in the identification of the genetic causes of complex diseases using conventional Sanger sequencing and demonstrates that WES is an efficient approach to increase our knowledge about the genetic causes of sperm immotility, infertility and PCD. Further, this study identified nine novel variants, with two being likely to be considered as genetic markers of sperm

immotility. Nonetheless, to understand the real impact of these variants, further studies, especially with animal models, are essential.

3.1.2. Characterization of CCDC103 expression profiles: further insights in primary ciliary dyskinesia and in human reproduction

Adapted from: Pereira et al, 2019, *J Assist Reprod Genet*.
https://doi.org/10.1007/s10815-019-01509-7

Background

In section 3.1, we described a group of male patients with total sperm immotility. Among these patients, it was included a PCD infertile male patient with *situs-inversus-totalis*, absence of sperm DA and in which, though WES, we identified a novel missense homozygous variant in the *CCDC103* gene (Pereira *et al.*, 2015). *CCDC103* was identified as a PCD gene in Pakistani PCD families with ODA defects (Panizzi *et al.*, 2012). *CCDC103* codes for an oligomeric coiled-coil domain protein that was found in the sperm axoneme and in cytoplasmic extracts of *C. reinhardtii* and *D. rerio* (Panizzi et al., 2012). Further, protein was shown to specifically bind polymerized microtubules, being critical to stabilize the microtubule polymeric structure, suggesting that may have a critical role in ODA assembly (King and Patel-King, 2015). Nevertheless, the knowledge about the role of this protein in ciliary biology is still very scarce, with *CCDC103* expression in human respiratory cilia and sperm having not yet been studied. With that in mind, we proposed to further investigate the role of protein CCDC103 in the ciliary structure and function. To achieve this, we examined the consequences of *CCDC103* pathologic variants in two PCD patients with *situs-inversus-totalis*, at the ultrastructural, RNA and protein level. Furthermore, we analysed the CCDC103 profile in different reproductive cells and somatic cells (nasal cells and white blood cells).

Patients with CCDC103 disease-causing variants display absence of dynein arms

As described previously, male patient 1 presents clinical features compatible with KS (PCD with *situs-inversus totalis)* and infertility due to the presence of rare total immotile sperm in the ejaculate. The ultrastructure of testicular sperm showed absence of DA and DRC (Sousa *et al.*, 2015). Here, we analysed nasal cilia and found the same ultrastructural

axoneme defects observed in sperm (Fig. 3.1-3A and B), which suggests that this gene affects both structures indistinctively. In humans, there are no studies that had simultaneously evaluated the effect of PCD gene mutations on the ultrastructure of the axoneme in both respiratory cilia and sperm flagellum. Although the axoneme structure of cilia and sperm are quite similar, some differences are known in the pathways regulating both. It is known that NOTCH signalling plays a crucial role in the differentiation of ciliated cells, with E2F transcription factors being required to activate centriole amplification genes, which are essential to the formation of multiple cilia. In contrast, WNT signalling plays a critical role in sperm flagellum formation, with E2F transcription factors being not so critical as sperm develop a single flagellum, with only one centriole (Choksi *et al.*, 2014). The present observations that this variant in *CCDC103* affected indistinctively both cilia and the sperm flagellum, suggest that this variant may act in a shared pathway of DA formation in both cell types. The presence of a pathogenic variant, the ultrastructural defects observed in axoneme structure, together with the high SD value (SD: 26.94) obtained from the determination of the ciliary beat axis and ciliary deviation data, established the diagnosis of PCD/KS in this patient.

We included also a female patient (female Patient-2) that presents chronic respiratory complaints, *situs-inversus totalis*, but no infertility. The axoneme ultrastructure of nasal cilia revealed absence of DA and DRC (Fig. 3.1-3C and D), and the determination of the ciliary beat axis and ciliary deviation showed a high SD (SD: 24.60), which established the diagnosis of PCD/KS in this patient.

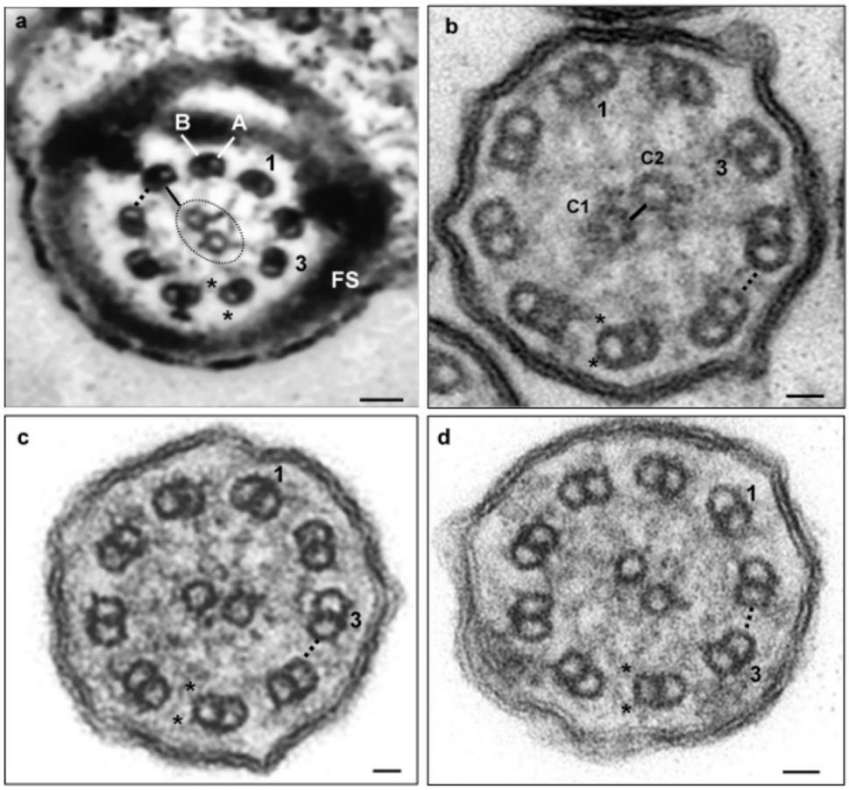


Figure 3.1-3 Ultrastructure of the axoneme. (a) sperm axoneme of patient-1; (b) axoneme of nasal cilia cells of patient-1; (c, d) axonemes of nasal cilia cells of patient-2. Note absence of dynein arms (*) and nexin links (dotted line). 1 and 3-numbering of the peripheral duplets, A-microtubule A of a peripheral duplet, B-microtubule B of a peripheral duplet, dotted circle-central sheath, C1 and C2-central microtubule pair, black bar-radial spokes, FS-fibrous sheath. Scale bars: 25 nm.

In both patients, WES analysis leads to the identification of novel homozygous, potentially pathogenic, gene variants in *CCDC103* (Fig. 3.1-4).

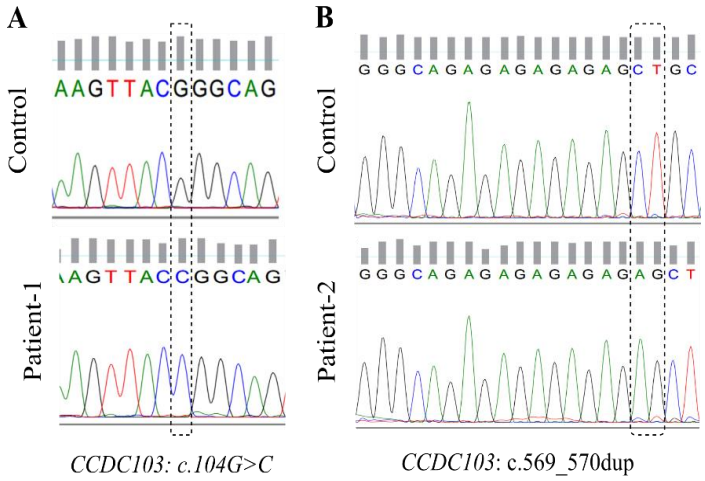


Figure 3.1-4. Sequencing electropherograms of variants found in the *CCDC103* gene (NM_213607.2).

(a) Variant c.104G>A of patient-1.

(b) Variant c.569_570dup of patient-2

Patient-1 showed a missense variant, already described, caused by a guanine to an adenine substitution at position 104 (c.104G>A), that corresponds to the replacement of

a positively charged arginine (Arg) by a nonpolar proline at protein position 35 and that is predicted as damaging with high confident scores (Pereira *et al.*, 2015). The female Patient-2 presents a duplication of two nucleotides (AG) between chromosomal position 17:42980013 and 17:42980014 (c.569_570dup), which results in a frameshift variant. This variant was submitted to ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), with accession number SUB5094211 (under processing). The mutant protein presents a p.Glu187Argfs*22 change at the RPAP3-Cter domain. Presently, only three genes, *RPAP3*, *SPAG1* and *CCDC103* genes have been found to code for an RPAP3 domain and, curiously, all those genes are directly (*CCDC103* and *SPAG1*) or indirectly (*RPAP3*) involved to cilia assemble and formation (Maurizy *et al.*, 2018). Although its function is not fully understood, the RPAP3 domain was suggested to participate in protein oligomerization and to be critical for maintaining ATPases in a conformation suitable for Client binding (Maurizy *et al.*, 2018). Client is the designation of a limited set of substrate proteins whose maturation is assisted by the heat-shock protein 90 (Hsp90), a molecular chaperone (Wayne *et al.*, 2011). Thus, this variant found in female patient-2 could affect the ATPases cycle and consequently DA assembly.

Patients with the CCDC103 variants have a reduction of expression at mRNA and protein level

As far as we know, no human studies regarding mRNA and protein expression of *CCDC103* in normal and pathologic conditions have yet been performed.

Patient and control cDNA, from WBC and nasal cells, was used to analyse *CCDC103* mRNA expression. In WBC of both patients, mRNA expression was significantly reduced ($p < 0.0001$) of about 0.2-fold (Fig. 3.1-5) to controls. In nasal cells of patient-1, no significant differences were observed, whereas, in patient-2, mRNA expression was significantly reduced ($p=0.0013$) of about 0.5-fold (Fig. 3.1-5) to controls. Although both variants are foreseen as pathogenic, the frameshift variant has a potentially higher impact to affect the phenotype, and therefore is comprehensible that this variant lead to a more perceptible expression reduction. Furthermore, the expected mRNA expression

reduction in nasal cells, caused by the missense variant, could have been masked by the low expression levels of *CCDC103* found in nasal cells (Fig. 3.1-10).

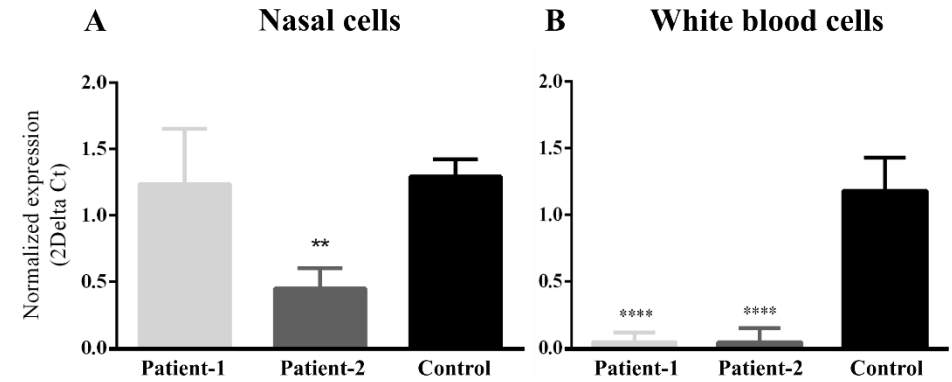


Figure 3.1-5. *CCDC103* mRNA expression levels in (a) nasal cells and (b) white blood cells from patients in comparison to controls. SYBR Green was the fluorescent dye used. Statistical significance was determined using the Mann-Whitney test, with alpha < 0.05. ** p < 0.01 and *** p < 0.0001. *B2M* and *GAPDH* were used as reference genes, with no statistical significance being observed between the normalizations performed with each locus.

Protein expression was evaluated by immunofluorescence in nasal cells of patients and controls. In controls, staining was localized in the cytoplasm and at the base of cilia (Fig. 3.1-6). In patients, cytoplasmic staining was strongly reduced, but again the reduction in patient-2 was more notorious, corroborating the hypothesis that the frameshift variant could have a more deleterious impact on the phenotype. In patient cells, we also did not observe staining at the base of cilia (Fig. 3.1-6), which suggests that *CCDC103* pathogenic variants affect protein expression but do not cause protein misallocation. Further studies on animal models are important to understand if those differences in patients' phenotypes are related to the type of variant or the variant location itself.

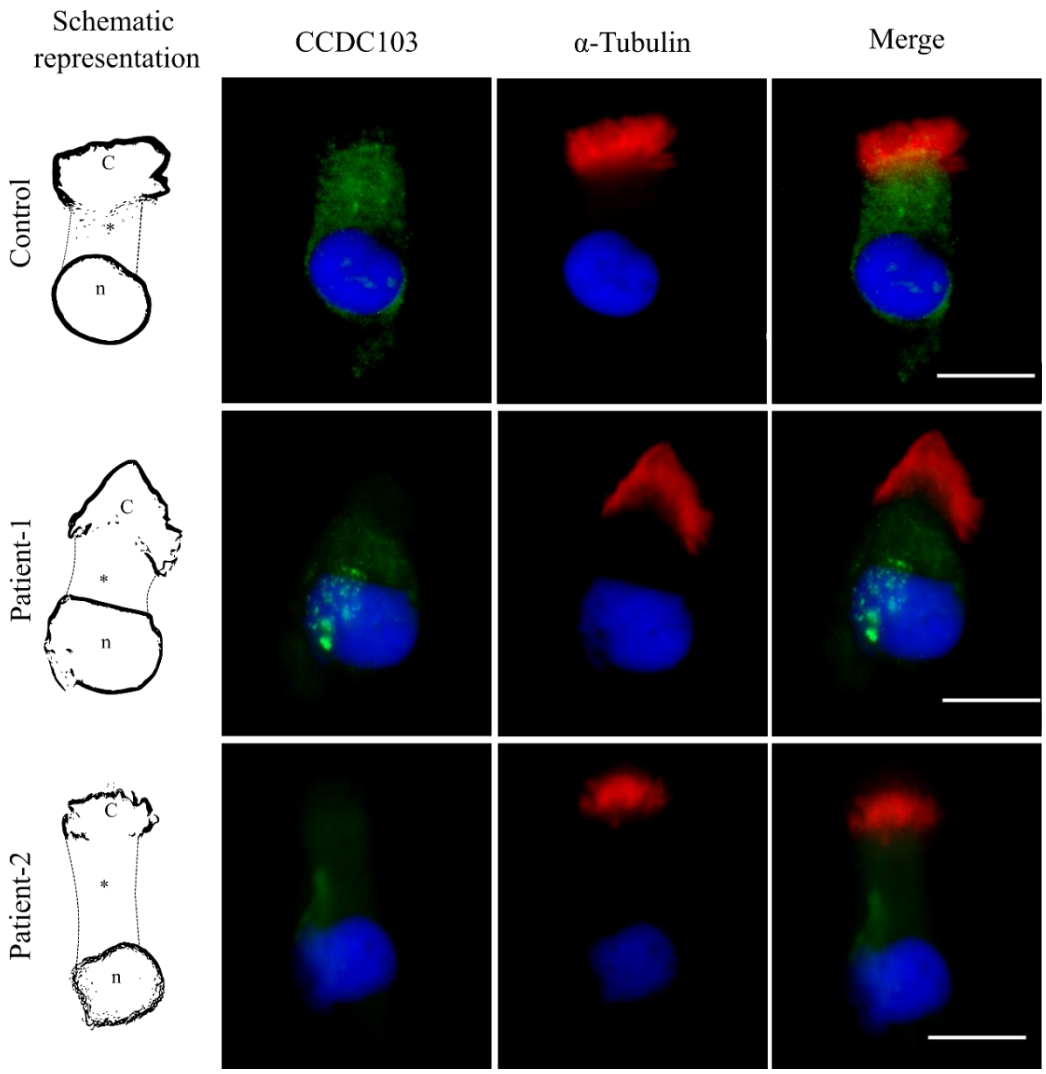


Figure 3.1-6. Immunocytochemical detection of CCDC103 (green) and of axoneme-specific acetylated α-tubulin (red), with merged images, in nasal cilia cells of controls, patient-1 and patient-2. Nuclei stained with DAPI (blue). C-cilia, *-cytoplasm, n-nucleus. Scale bars: 10 μm

Protein expression was also evaluated by ICC in testicular cells of patient-1 and controls, in ejaculated sperm of controls and in Sertoli cells. **In controls,** CCDC103 was localized in the cytoplasm and perinuclear region of Sertoli cells, in the cytoplasm of primary spermatocytes (ST1), secondary spermatocytes (ST2) and round spermatids (Sa), and in the midpiece of sperm (Fig. 3.1-7).

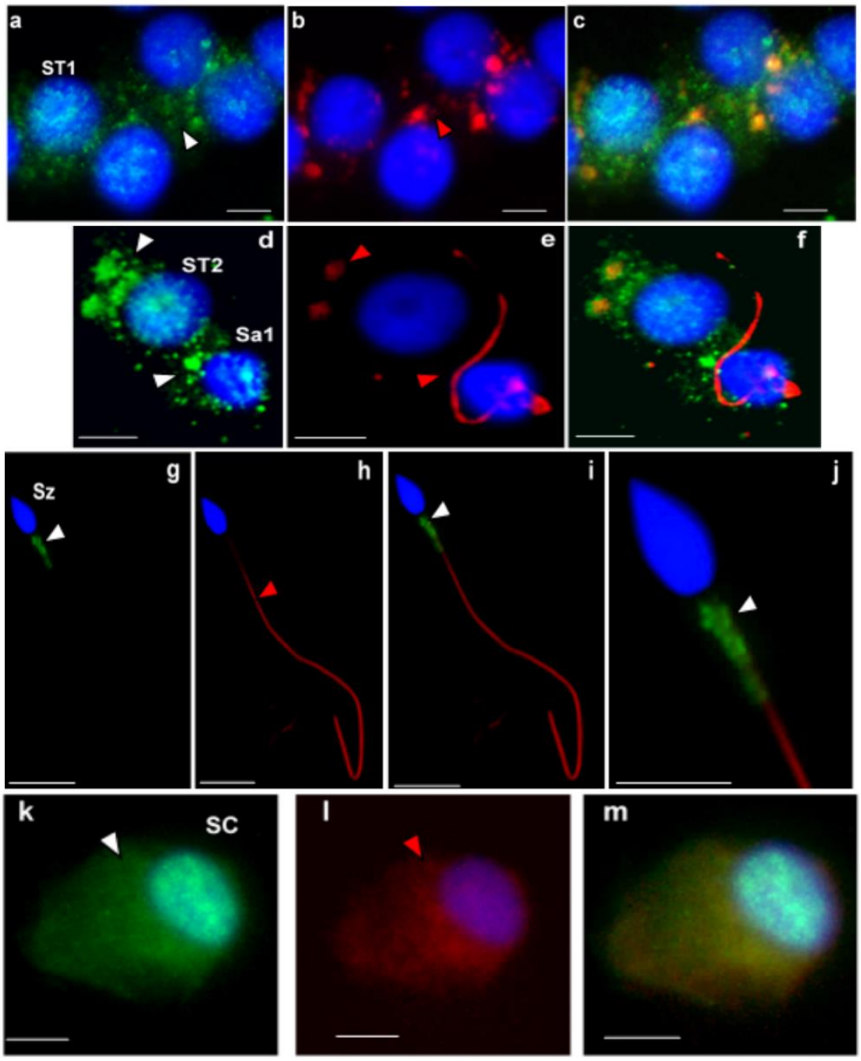


Figure 3.1-7. Immunocytochemical detection of CCDC103 (green) and of axoneme-specific acetylated α-tubulin (red), with merged images, in cells of the seminiferous tubules from controls. Staining was observed in the cytoplasm of primary spermatocytes (ST1: a-c), secondary spermatocytes (ST2: d-f) and round spermatids (Sa: d-f)), and in the midpiece of sperm (Sz: g-j). In Sertoli cells (SC: k-m), staining was observed in the cytoplasm and in the perinuclear region. Nuclei stained with DAPI (blue). White arrowheads-CCDC103 staining, red arrowheads-tubulin staining. Scale bars: 5 μm.

In patient-1, staining was observed in the cytoplasm and perinuclear region of ST1 and ST2, in the basal nuclear pole of Sa and in the midpiece of sperm (Fig. 3.1-8).

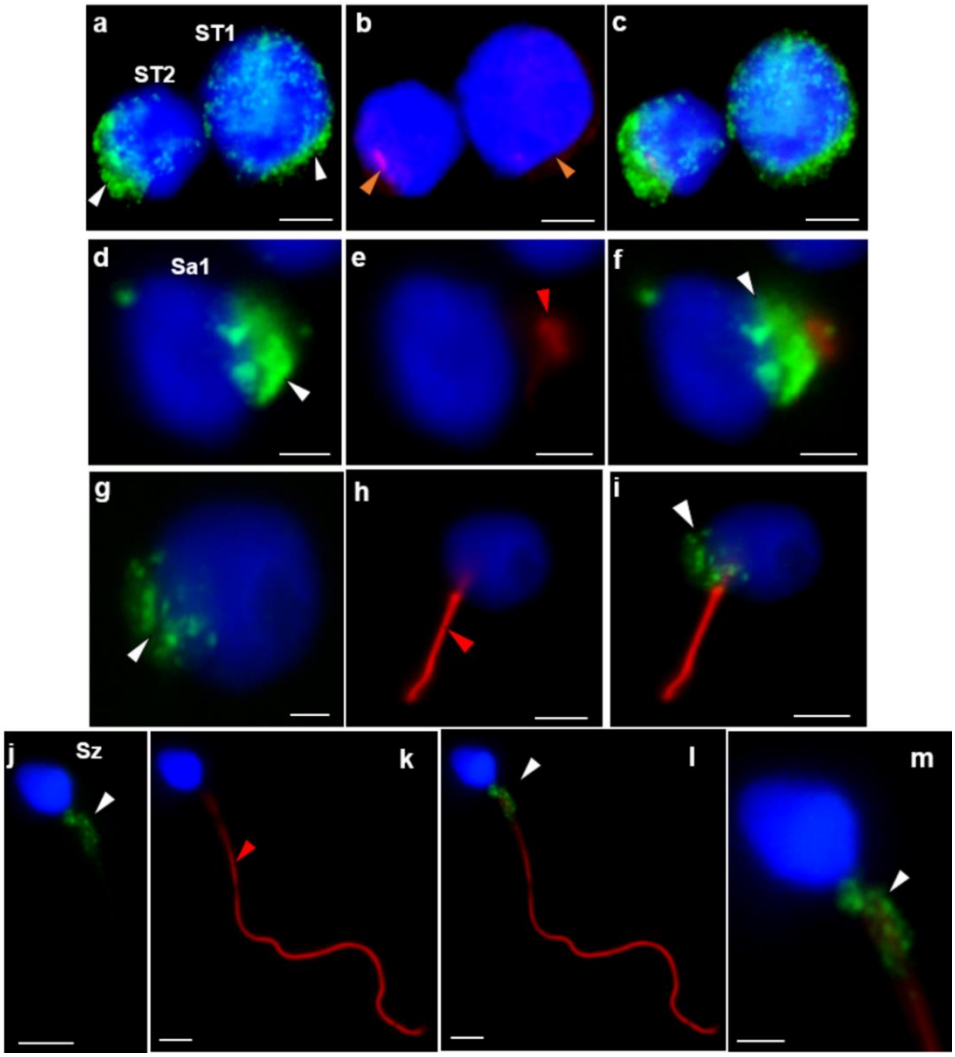


Figure 3.1-8. Immunocytochemical detection of CCDC103 (green) and of axoneme-specific acetylated α-tubulin (red), with merged images, in cells of the seminiferous tubules from **patient-1**. Staining was observed in the cytoplasm and perinuclear region of primary spermatocytes (ST1: **a-c**) and secondary spermatocytes (ST2: **a-c**), at the basal pole of the nucleus preceding axoneme extrusion in early round spermatids (Sa1: **d-f**) and late round spermatids (Sa2: **g-i**), and (granular appearance) in the midpiece of sperm (Sz: **j-m**). Nuclei stained with DAPI (blue). White arrowheads-CCDC103 staining, red arrowheads-tubulin staining. Scale bars: a-c: 4 μm; d-f: 2 μm; h, i: 4 μm; g: 2 μm; j-m: 2 μm.

Cytoplasmic staining of CCDC103 was previously observed in the cell body of *C. reinhardtii* (King and Patel-King, 2015). It was proposed that CCDC103 may assemble during axonemal growth, providing a high-affinity track along doublets to allow the association between the ODA/docking complex (King, 2016). Our observation of *CCDC103* expression in cells without motile cilia, altogether with its cytoplasmic staining, support previous findings that CCDC103 acts as a **cytoplasmatic dynein**

assemble factor. However, in contrast with what was described in the sperm of *C. reinhardtii* (King and Patel-King, 2015), we did not observe CCDC103 staining along the whole flagellum length of human sperm, but only in the sperm midpiece. *C. reinhardtii* flagella have been proved to be a suitable model species to study the molecular components of axonemes and the ciliogenesis process (Inaba, 2011). Further, *CCDC103* gene is well conserved. Nevertheless, the protein sequence of both species only shares 32% identity, which could explain those location differences. Moreover, both species may certainly have different posttranslational modifications that can likely regulate protein location.

Protein expression was further evaluated by Western-Blot analysis in the WBC from patients and controls. This task revealed very challenging due to CCDC103 specific biochemical characteristics. Further, all available commercial antibodies were designed using the same immunogen, a synthetic peptide directed towards the C-terminal region, thus we could not discard the hypothesis of incomplete access to native protein due to the presence of self-association or complex alpha helices in these regions. As there was a low degree of reproducibility, we were unable to quantify protein expression by Western-Blot. Nevertheless, it was possible to infer that both patients revealed reduced protein expression, especially patient-2 (Fig. 3.1-9).

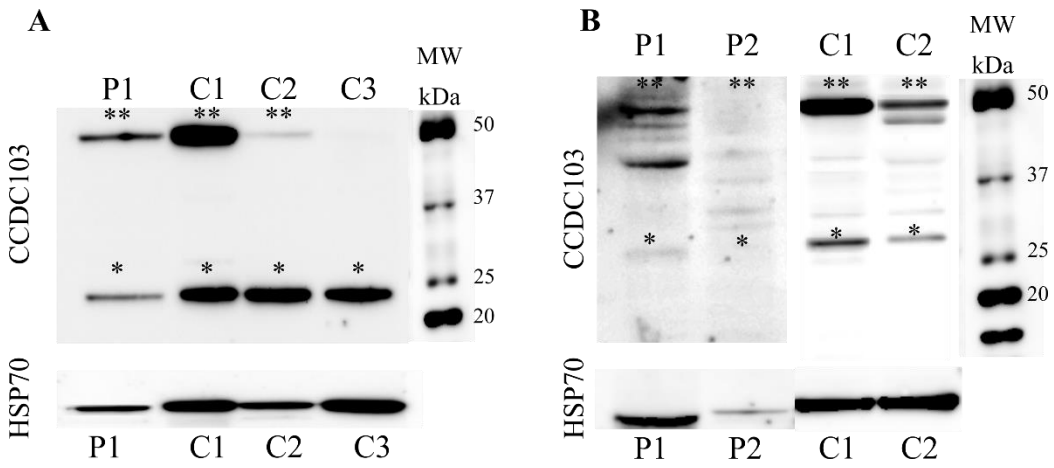
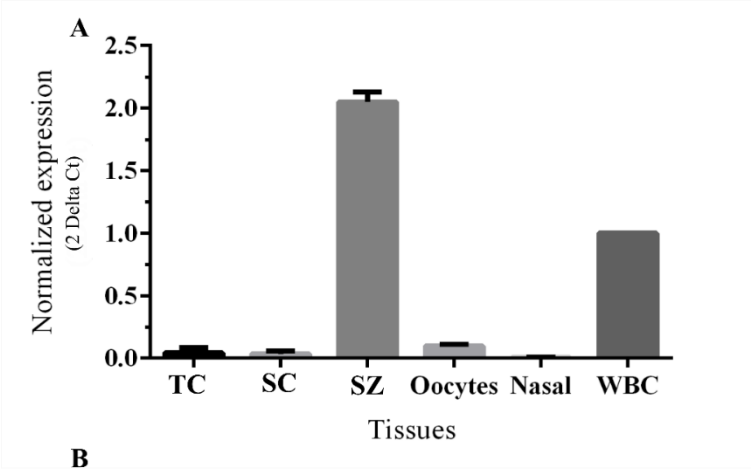


Figure 3.1-9. CCDC103 protein detection in white blood cells from controls, patient-1 and patient-2. *Possibly monomer form; **Possibly dimer and/or higher-order oligomers. HSP70 (molecular weight (MW) of about 70 kDa) was used as loading control.

CCDC103 is present in several cells at different expressions levels

To further increase the knowledge of *CCDC103* expression, control somatic and reproductive tissues were compared. Taken WBC as a reference, results showed that mRNA was expressed in all tissues but with different expression levels. Testicular cells (that are cells from the seminiferous tubules, which includes sperm cell progenitor cells, namely spermatogonia, primary and secondary spermatocytes and round spermatids) Sertoli cells, oocytes and nasal cells evidenced low expression levels (Fig. 3.1-10).



B

C \ S	WBC	TC	SC	SZ	Oocytes	Nasal
WBC		↓	↓	↑	↓	↓
TC	↑		⇔	↑	⇔	↓
SC	↑	⇔		↑	↑	↓
SZ	↓	↓	↓		↓	↓
Oocytes	↑	⇔	↓	↑		↓
Nasal	↑	↑	↑	↑	↑	

Figure 3.1-10. *CCDC103* RNA expression levels (a) from control testicular cells (TC), Sertoli cells (SC), sperm (SZ), oocytes, nasal cells (nasal) and white blood cells (WBC). SYBR Green was the fluorescent dye used. Statistical significance was determined using the Mann-Whitney test, with alpha < 0.05. WBC was used as a calibrator and B2M as the reference gene. Resume of all possible combinations between samples (b). The C (calibrator) column means that this tissue was used as calibrator. For example, using WBC as a calibrator, it was observed that samples (S) from TC, SC, oocytes and nasal cells present a reduced expression (↓), while SZ has a higher expression (↑) than WBC. ⇔ means that there were no statistically significant differences between the two groups

These results show that nasal cells present a much lower expression in comparison to WBC. A quantitative proteomic analysis of human airway cilia also observed that *CCDC103* expression was about 0.0015-fold lower than that found in *DNAH5* and *DNAH9*, two well-known cilia genes (Blackburn *et al.*, 2017). In contrast to cilia cells, we observed that sperm presented the highest expression levels, 10-fold more than oocytes

and about 2-fold more than WBC, which further confirms that *CCDC103* may have a special role in reproductive function and corroborates findings in reproductive cells of patient-1. We here also **firstly describe** the presence of *CCDC103* mRNA transcripts in human Sertoli cells and oocytes.

Western-Blot experiments in control somatic and germ cells showed that the protein CCDC103 forms monomers, dimers, and higher-order oligomers, whose size appeared tissue-specific (Fig. 3.1-11), corroborating previous findings that state oligomerization as a property of CCDC103 (King and Patel-King, 2015; Shoemark *et al.*, 2017). The theorical band size of about 27 kDa (monomer) was only observed in WBC, which also presented a dimer of about 50 kDa. In Sertoli cells, it was observed bands of about 37, 70 and 150 kDa. If we assume that in these cells the 37 kDa band as a Sertoli cell-specific monomer, thus the other correspond to dimer and higher-order oligomers. Testicular cells presented a similar profile to Sertoli cells, but with the absence of the 70 kDa dimer. In ejaculated sperm, the protein was hard to detect despite multiple extraction protocols tested and was only achieved when total protein was protected with proteinase inhibitors before and after pelleted cells. Sperm evidenced the nearly 70kDa dimer, comparably to Sertoli cells, and a higher-order oligomer of about 125 kDa. These results show the heterogeneity of this protein and likely reflect different functions or interactions. Studies in the sperm flagellum of *C. reinhardtii* highlighted that CCDC103 has great biophysical properties being hard to break its native conformation (King and Patel-King, 2015). Our experience supports these previous findings, as it was very difficult to detect CCDC103 protein in all cells, especially in sperm. We obtained a great variability among experiences, with the monomer bands not always being observed, which corroborates its strong dimerization proprieties. Additionally, these results could suggest that this protein may have a high turnover rate, making hard to have good reproducibility. Nevertheless, the great variability among experiments precluded us to comparatively quantify the protein in these experiments.

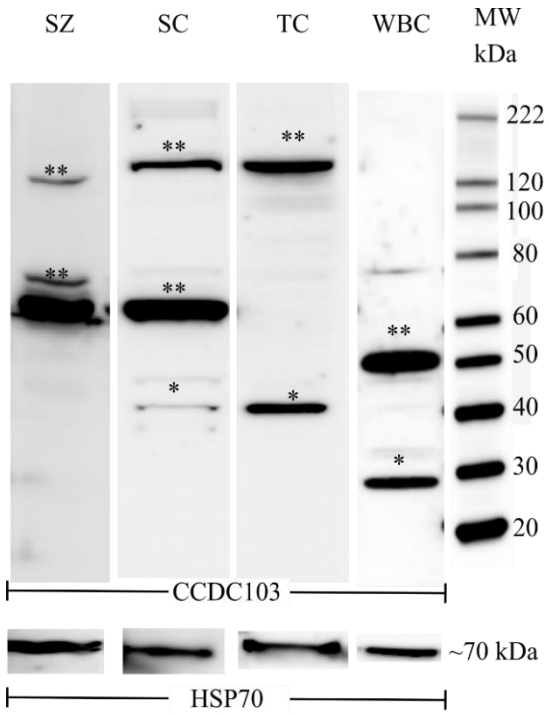


Figure 3.1-11. CCDC103 protein detection from control sperm (SZ), Sertoli cells (SC), testicular cells (TC) and white blood cells (WBC). *Possibly monomer form. **Possibly dimer and/or higher-order oligomers. Polyclonal CCDC103 antibody from ABCAM was used. HSP70 was used as loading control

To infer about the observed size differences among cells, we sequenced the full mRNA of *CCDC103*, including UTR, which are known to play crucial roles in the post-transcriptional regulation of gene expression (Matoulkova *et al.*, 2012). As no differences were detected between cells, data indicate that the differences could be explained by external factors that can regulate the dimerization process. For instance, post-translational modifications, such as tyrosine phosphorylation, were described as having a role in dimerization of the RACK1A protein (Sabila *et al.*, 2016). According to the NetPhos 3.1 Server (http://www.cbs.dtu.dk/cgi-bin/), several phosphorylation sites are predicted for the *CCDC103* sequence, and thus it is possible that tissue-specific protein regulation could explain differences in the dimerization process.

CCDC103 is differentially located in germ cells

To gain further insights about CCDC103 location in germ cells, IHC detection in paraffin sections of seminiferous tubules was performed. **Tubulin staining** was observed in the cytoplasm and perinuclear region of cells, being intense in spermatogonia (SG), moderate intense in ST1, weakly intense in ST2, Sa and Sertoli cells, and intense in the sperm flagellum (Fig. 3.1-12). **CCDC103 staining** was also observed in the cytoplasm and perinuclear region of cells but, in contrast to tubulin, staining was

weak in SG, intense in ST1 and ST2, moderate in Sa and Sertoli cells, and intense in the sperm midpiece (Fig. 3.1-13).

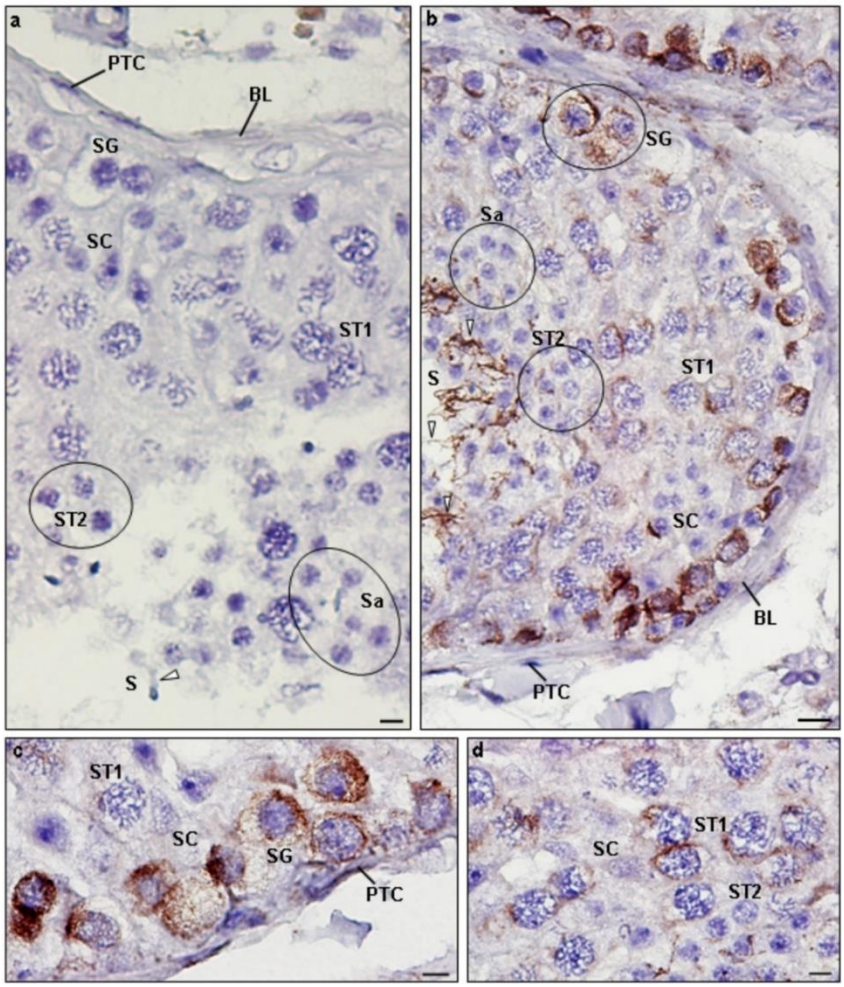


Figure 3.1-12. Immunohistochemical detection of tubulin in paraffin sections of control seminiferous tubules. (a) Negative control (absence of staining). (b-d) Staining was intense in the cytoplasm and perinuclear region of spermatogonia (SG), moderately intense in the cytoplasm and perinuclear region of primary spermatocytes (ST1), and weakly intense in secondary spermatocytes (ST2), round spermatids (Sa) and Sertoli cells (SC). The flagellum (white arrowheads) of late spermatids (Sd) presents strong staining. PTC-peritubular cells, BL-basal lamina. Scale bars: 10 μm

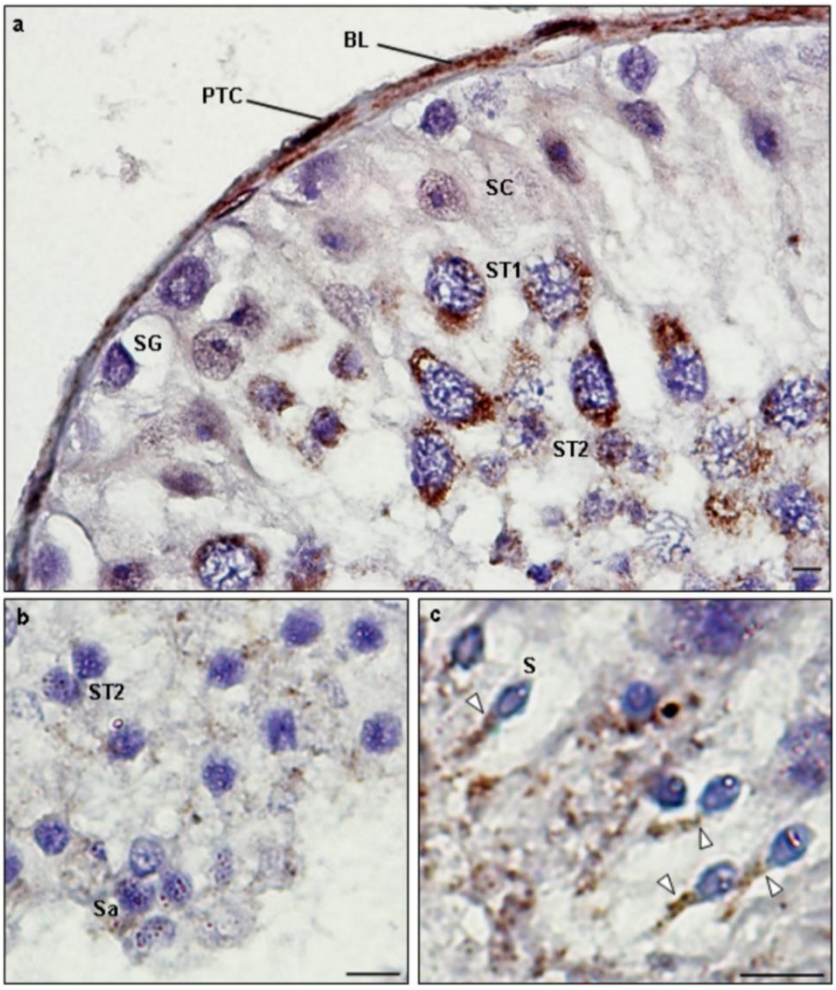


Figure 3.1-13. Immunohistochemical detection of CCDC103 in paraffin sections of control seminiferous tubules. Staining was weakly intense in spermatogonia (SG), intense in the cytoplasm and perinuclear region of primary spermatocytes (ST1) and secondary spermatocytes (ST2), and moderately intense in the cytoplasm and perinuclear region of Sertoli cells (SC) and round spermatids (Sa). The midpiece of the elongated spermatids (Sd) showed intense staining (white arrowheads). PTC-peritubular cells, BL-basal lamina. Scale bars: 10 μm

The results from immunohistochemical staining demonstrate that the presence of CCDC103 throughout spermatogenesis follows a cell-stage specific manner. The observed localization is similar to what was previously observed in ZMYND10. *ZMYND10* is another gene known to be involved in the assembly of DA, which was shown to be strongly expressed in the cytoplasm of mouse primary spermatocytes and spermatids (Mali *et al.*, 2018). This suggests that DA assembly could start in spermatocytes. CCDC103 location in sperm was also inspected by laser scanning confocal microscopy. Staining exhibited a helix shape throughout the midpiece region, involving it as a scarf.

To further understand the spatial and subcellular localization of CCDC103 on sperm, we attempted immunogold labelling. Tubulin labelling was observed in the microtubules of the proximal centriole and axoneme (Fig. 3.1-14). Despite multiple attempts with different staining methods and antibody dilutions, no staining was reproductively obtained for CCDC103. This could be due to the method needed for fixation and embedding in acrylic resins, which leads to poor preservation of cell membranes (Porter and Kallman, 1953; De Paul *et al.*, 2012). There still uncertainty concerning the CCDC103 sub localization: is it in the sperm midpiece connected either to MT, DA or in the cytoplasm. If it is cytoplasmic, it is then possible that the antigen was lost during processing. Another hypothesis is that the antibody used has no specificity for ultrastructural antigen detection. In fact, all commercially available CCDC103 antibodies were made using the same immunogen sequence, and none were tested for immunogold labelling. Additional studies may be needed, to verify if with another technique, such as cryo-transmission electron microscopy that better preserve structural information of cell membranes, the antigen is preserved and the CCDC103 labelling is more efficient. In contrast, the tubulin antigen was very well protected, with a perfect labelling being observed in the MT of the proximal centriole and of the axoneme.

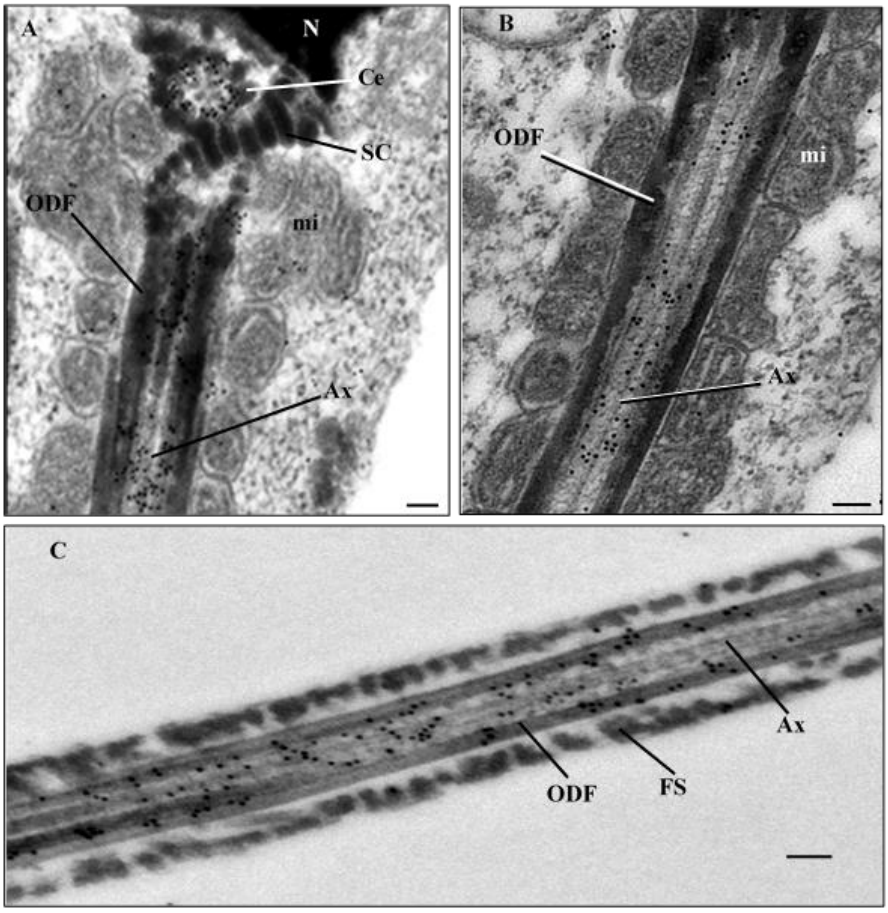


Figure 3.1-14. Immunogold staining of tubulin in control sperm. Gold particles were observed in the centriole (Ce) and axoneme (Ax). (A, B) Midpiece. (C) Principal piece. N-nucleus, SC-striated columns, ODF-outer dense fibers, mi-mitochondria, FS-fibrous sheath. Scale bars: 100 nm

In conclusion, here we firstly characterize the expression profiles of *CCDC103* in human cells, thus increases the knowledge regarding its expression and subcellular localization. We described that CCDC103 has a broad distribution, being present both in somatic tissues and in reproductive cells. As cytoplasmic dyneins are present in nearly all animal cells (Neely and Boekelheide, 1988), our present data corroborates that CCDC103 is a cytoplasmic dynein arm assembly factor and suggests that the *CCDC103* gene may be present in an initial pathway of the ciliogenesis process and/or that its function is not exclusive to axonemal DA assemble.

We also described for the first time the presence of CCDC103 in WBC and in testicular germ cells, with higher staining at the spermatocyte-stage and local staining in the midpiece of both testicular and ejaculated sperm. Further, we could show that the pathogenic variants found in the *CCDC103* gene lead to an absence of DA and to a

significant reduction of gene expression and protein expression that was also tissue-specific. Our data corroborate the involvement of CCDC103 in PCD and suggests also a role in infertility.

We believed that CCDC103 may have more unknown functions in cilia biology and male reproduction. Thus, these results added an additional piece to the complex puzzle of the axonemal dynein assembly process and ultimately may help to further understand the pathophysiology of PCD.

3.2. Primary ciliary dyskinesia: childhood focus

We previously focused our study on genetic bases of sperm immotility and studied the genotype-phenotype correlation, as well as, expression profiles of *CCDC103*, a PCD related gene that we suggested, given its expression profiles, to have a particular role in spermatogenesis.

In the follow-up of this thesis, we proposed to provide further insights on PCD genetic etiology. For that, we performed a comprehensive clinical, ultrastructural and genetic study of two unrelated Portuguese children with strong PCD suspicion. We included three children that are being followed at the Maternal Child Centre of North/ Hospital Centre of Porto (CMIN/CHP), due to upper respiratory complains and situs inversus totalis. All children have clinical features compatible with KS, and underwent extensive clinical assessments, ultrastructural analysis of motile cilia by transmission electron microscopy, whole exome sequencing and bioinformatics analysis to identify their genetic etiology. In this section we will firstly focus our attention in two of those children and then describe the genetic analysis of the third child, which genetic etiology was showed challenging to unveil.

3.2.1. Clinical and genetic analysis of children with Kartagener syndrome due to *DNAH5* and *CCDC40* mutations

Adapted from: Pereira et al, 2019, Cells https://doi.org/10.3390/cells8080900

Ciliary ultrastructure is disrupted in PCD children

Children-1 is a female child that was referred to the consultation at 16 months of age due to acute bronchiolitis of difficult resolution, followed by recurrent respiratory infections, some requiring hospitalization due to exacerbation. During her first moments of life, doubtful cardiac sounds raised the suspicion of a *situs inversus*. Later, she presented recurrent acute otitis media (AOM), mucopurulent rhinorrhea and productive daily cough (for more details on patient phenotype see Methodology).

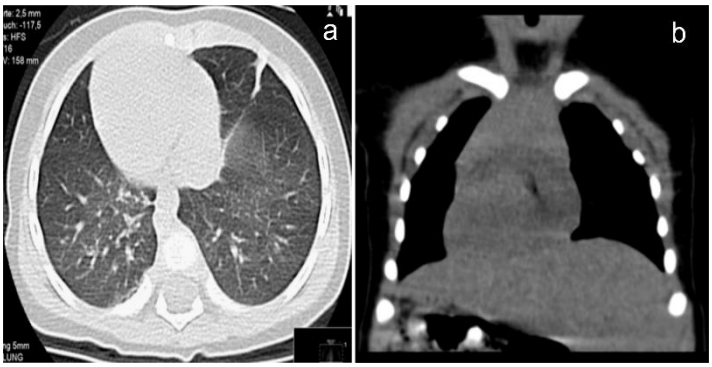


Figure 3.2-1. CT scans, without administration of intravenous contrast from child 1.
a Axial plane, pulmonary window. Presence of dextrocardia. There are no relevant changes in the lung parenchyma. b Coronal reformation. Presence of dextrocardia (right apex) and liver on the left flank.

The *situs inversus totalis* was confirmed after an echocardiogram and a thoracoabdominal CT scan, which showed a complete inversion of the heart and lungs, the presence of liver on the left and of stomach and spleen on the right (Fig. 3.2-1). No structural intracardiac anomalies were detected in the echocardiogram.

The ultrastructural analysis of the axoneme showed a heterogeneous pattern, with most of the cases lacking IDA and nexin links, presence of partial RS and a slight displacement of the CPC (Fig. 3.2-2a). Remainder axonemes evidenced an absence of the IDA or of both DA, CPC displacement, the presence of extra central pair microtubules and displacement of the peripheral doublets, always with absence of the RS and nexin

links. Evaluation of 192 axonemes revealed a variation in the ciliary beat axis deviation of 43.60 (36.4 - 50.7), which is compatible with PCD (De Iongh and Rutland, 1989).

Child-2 is male child that was hospitalized immediately after birth with a respiratory distress syndrome. On the chest X-ray, a right opacification was observed, with clinical improvement after antibiotic therapy. He was delivered by day 13. During hospitalization, as in child 1, cardiac sounds were found on the right and chest X-ray confirmed the inversion of cardiac laterality, emerging the suspicion of *situs inversus*. Up to about 2.5 months of age, control chest X-ray did not further show opacities compatible with pneumonia, but he developed nasal obstruction, productive cough and purulent nasal secretions, without wheezing, fever or AOM. From 1 year of age, he developed sporadic and self-limiting crises of productive cough and wheezing, without fever, AOM or mucopurulent rhinorrhoea. The *situs inversus totalis* was confirmed as in child-1. No structural intracardiac anomalies were detected in the echocardiogram. The ultrastructural analysis of the axoneme showed a homogenous pattern, with all axonemes lacking DA and nexin bridges (Fig. 3.2-2b). The RS and CPC were present. Evaluation of 228 axonemes revealed a variation in CBAD of 18.4, which is compatible with a bronchiectasis type of clinical severity. Even though not compatible with a typical PCD, it is as well, although milder, a PCD feature (De Iongh and Rutland, 1989).

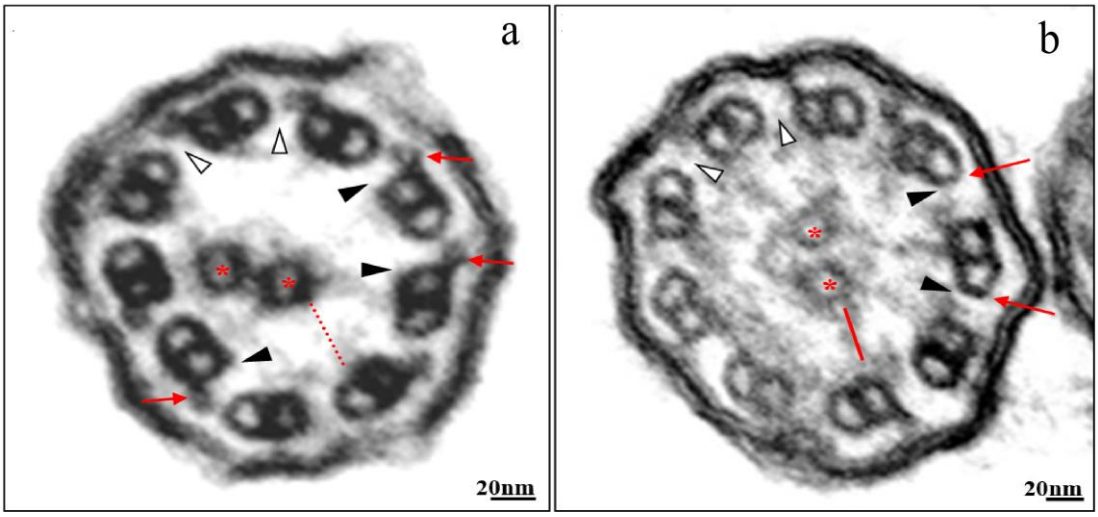


Figure 3.2-2 Ultrastructure of the axoneme. **a** Ultrastructure of the axoneme of child 1. Note absence of the inner dynein arm (black head arrow) with presence of the outer dynein arm (red arrows), absence of nexin links (white head arrows: absence of a link between doublets that here appears as a blank space), absence of the radial spokes (dotted red line) and a slight displacement of the central microtubules pair (*). **b** Ultrastructure of the axoneme of child 2. Note absence of both dynein arms, inner (black arrowheads) and outer (red arrows), and absence of nexin links (white arrowheads), with the presence of the radial spokes (red bar) and correctly located central microtubules pair (*).

WES analysis revealed gene variants in CCDC40, DNAH5 and DNAH7 genes

From WES analysis, in **child 1,** we found two compound heterozygous variants in the *CCDC40* gene: one substitution in a splice site, c.1989+1G>A, and one frameshift variant, c.2824_2825insCTGT (Fig.3.2-3a; Table 3.2-1). The c.1899+1G>A has an extremely low frequency (0.0008%) and was predicted to affect the wild-type donor splice site by all splice site prediction algorithms included in the Alamut software (Fig.3.2-3b). The c.2824_2825insCTGT variant has also a very low frequency (0.0042%), it was predicted to affect the open reading frame (p.Arg942fs), and classified as pathogenic according to the ClinVar database.

Previous studies suggested that *CCD40* gene plays an important role in cilia function and left-right axis specification(Becker-Heck *et al.*, 2011) and that CCDC40 forms a complex with CCDC39, which may function as docking sites for RS, nexin bridges and IDA. This is corroborated by the fact that several patients with CCDC40 defects, including ours, display a variable disorganization of axoneme ultrastructure

(Attachment 3) (Becker-Heck *et al.*, 2011; Blanchon *et al.*, 2012; Antony *et al.*, 2013; Sui *et al.*, 2016). Both variants found in this patient are disease-causing. The variant c.1989+1G>A is predicted to affect the donor splice site. The donor splice site (or 5' splice site, a key regulator of splicing) includes an almost invariant GU sequence. Variants that change this highly conserved GU residue, as in patient 1, may impair the binding of the U1 small nuclear ribonucleoprotein. This could potentially cause exon skipping, which would drastically affect the protein (Pagani and Baralle, 2004). Alternatively, a cryptic splice site (a sequence identical that of the donor splice site but that not regularly used) could be used, also leading to an abnormal RNA(Matlin *et al.*, 2005; Lewandowska, 2013). Unfortunately, during the time of this study, we were unable to obtain a newer nasal sample from both children, and it was thus impracticable to further characterize this variant. Regarding the variant c.2824_2825insCTGT this likely lead to a complete shift of the reading frame and putatively to a non-functional protein.

Table 3.2-1 List of pathogenic variants found in the patients included in this study.

Patient	Gene (Transcript)	Genomic location (hg19)	Variant (Exon/intron)	Protein change (Variant type)	Freq.	Variant origin
Child 1	*CCDC40* (NM_017950)	Chr17: 78055858	c.1989+1G>A (12i)	NA (Splicing)	0.0008%	P
		Chr17: 78063675	c.2824_2825insCTGT (17)	p.Arg942fs (Frameshift)	0.0042%	M
Child 2	*DNAH5* (NM_001369)	Chr5: 13864572	c.4530delG (28)	p.Asn1511fs (Frameshift)	New	M
		Chr5: 13830767	c.6000C>A (36)	p.Tyr2000* (Nonsense)	New	P
	DNAH7 (NM_018897)	Chr2: 196722306	c.8209G>A (44)	p.Gly2737Ser (Missense)	0.0008%	P
		Chr2: 196602773	c.11947C>T (65)	p.Arg3983Trp (Missense)	0.8666%	M

M- Maternal origin; P-Paternal origin

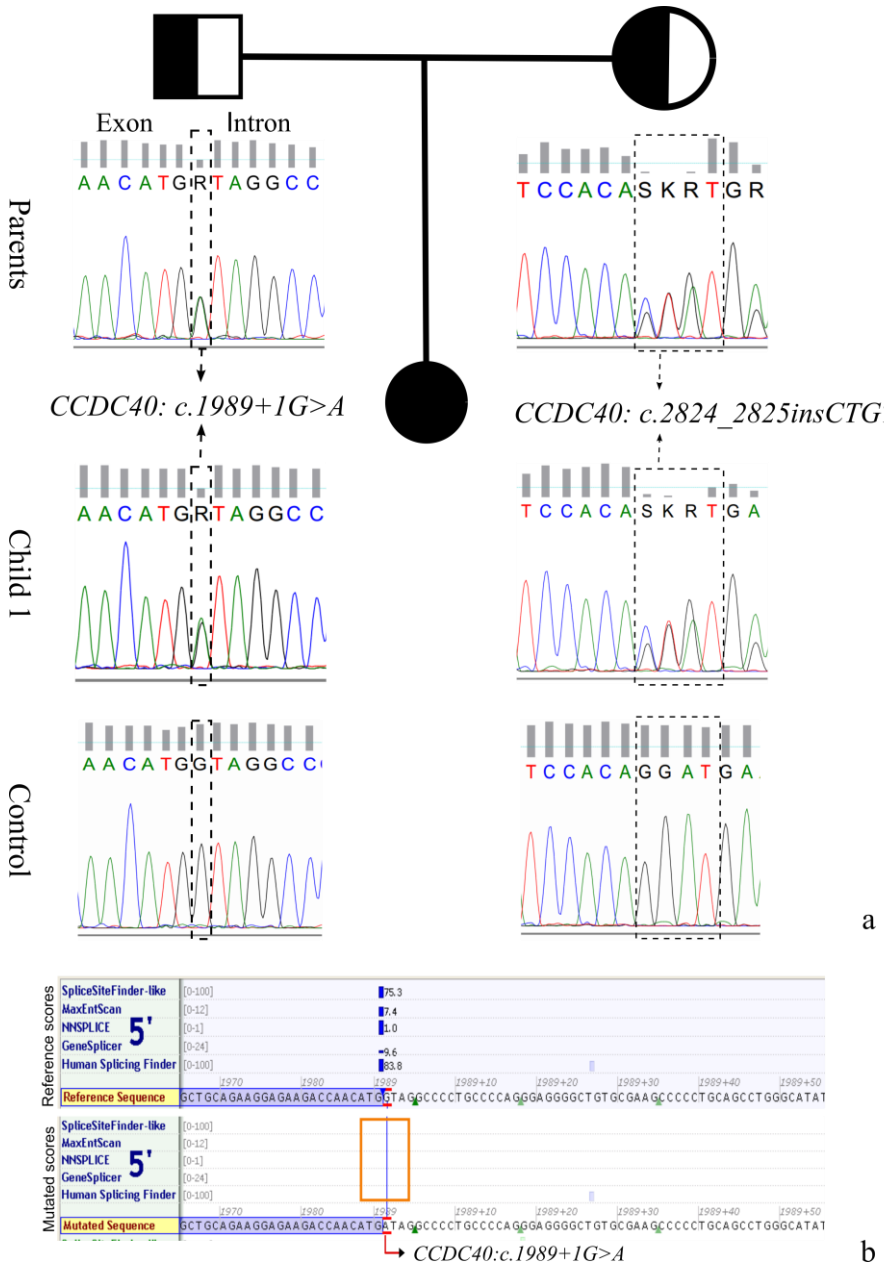


Figure 3.2-3. Genetic analysis of the child 1. a Pedigree and sequencing electropherograms from child 1 and both parents, showing the variants c.1989+1G>A and c.2824_2825insCTGT found in *CCDC40* gene; b Bioinformatic analysis of the splicing variant c.1989+1G>A of *CCDC40* gene. Top panel: reference sequence; bottom panel: mutated sequence, with the respective predictive scores for each algorithm. The reference sequence and mutated sequences are shown at the bottom of each panel with nucleotide positions indicated above. Blue vertical bars at the top panel indicate 5′ (donor) site scores. The orange box highlighting the loss of donor splicing site. Figure obtained from Alamut Visual software (v2.10).

In **child 2**, WES revealed two compound heterozygous variants in the *DNAH5* gene and other two in the *DNAH7* gene (Fig.3.2-4; Table 3.2-1). In *DNAH5*, we found **a frameshift** variant caused by a single base deletion at the coding position 4530 (c.4530delG); and **a nonsense** caused by a change from a cytosine to an adenine at coding

position 6000 (c.6000C>A). Both SNV were firstly detected in this study and reported on the ClinVar database (submission numbers: SCV000579338.1 and SCV000537861.1, respectively).

Regarding the *DNAH7* gene, two missense compound heterozygous variants were found. The first is a guanine to an adenine substitution at the position 8209 (c.8209G>A), which lead to a change from a hydrophobic glycine to a polar serine on amino-acid 2737 (p.Gly2737Ser). The other was found within the cDNA position 11947, consisting on a cytosine to a thymine substitution (c.11947C>T), leading to the replacement of the positively charged arginine to the hydrophobic tryptophan (p.Arg3983Trp).

In this child, we describe two novel variants in the *DNAH5* gene, which codes for a dynein HC that is localized on the ODA. It is a major disease-causing gene in PCD patients with ODA defects (Attachment 3). Both variants can potentially lead to a premature non-functional protein and both affect relevant domains of *DNAH5*. Particularly, the nonsense variant is in a hydrolytic ATP binding site of the dynein motor region D1 (or AA1), which is highly conserved (Mocz and Gibbons, 2001). It is the main ATPase site, and its movement is critical for dynein function (Kon *et al.*, 2012). Disruptions of this domain may lead to an enzymatically inactive DHC (King, 2000). Therefore, those variants are thought to justify the absence of ODA and the *situs inversus* in child 2. Few reports described an association between *DNAH5* pathogenic variants and absence of both DA (Attachment 3). However, null dnah5 mice showed normal IDA and absent ODA(Ibañez-Tallon *et al.*, 2002). As child 2 showed absence of both DA, it is intriguing to speculate that alterations in another gene product could act together with the *DNAH5*. Curiously, we have identified two additional compound heterozygous variants in the *DNAH7* gene. This *locus* is a member of the inner arm family of the DHC and was found to be missing in a patient with PCD who had absence of IDA and displayed abnormal cilia motility, but with normal ODA(Zhang *et al.*, 2002). In this child, both variants found in the *DNAH7* gene affect highly conserved amino acids and are predicted to be disease-causing according to several bioinformatic tools (Table 3.2-2).

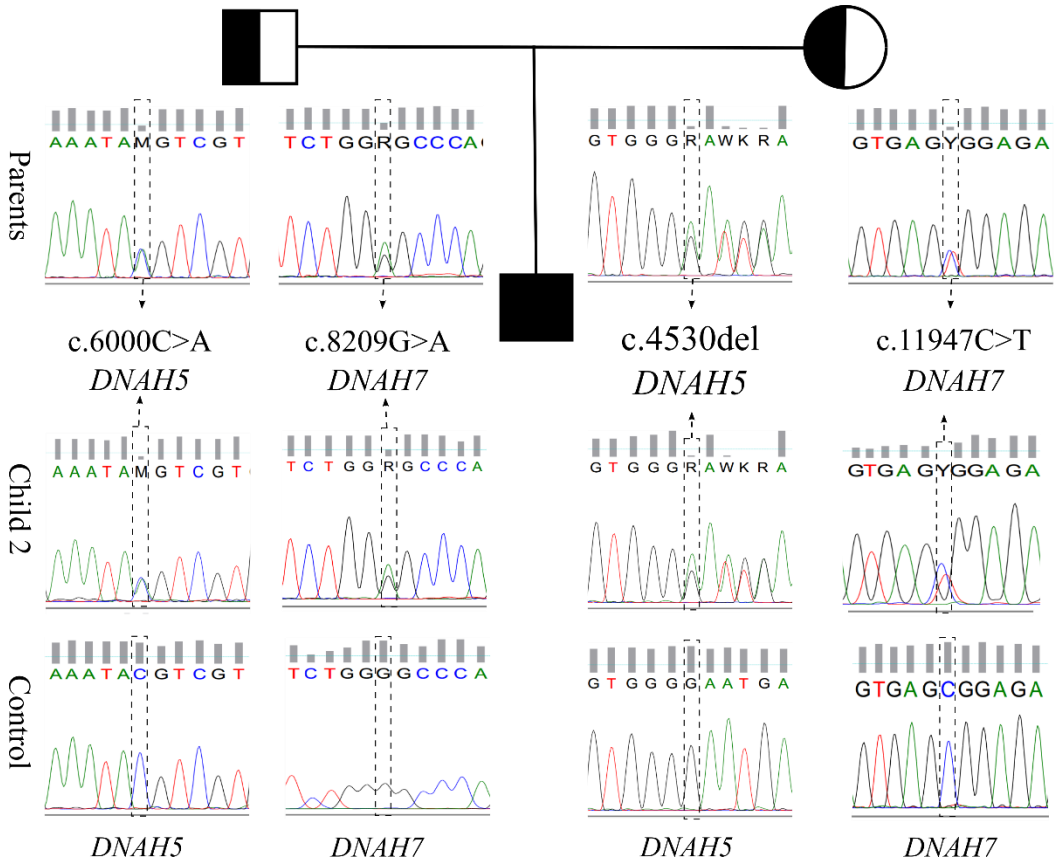


Figure 3.2-4 Sequencing electropherograms from patient 2 and both parents showing the variants found in DNAH5 gene (c.4530delG and c.6000C>A) and the variants found in DNAH7 gene (c.8209G>A and c.11947C>T).

Table 3.2-2. Bioinformatic analysis of DNAH7 variants found in the studied child 2 of this study.

Variant	SIFT	Poly-Phen-2	Align GVGD	Mutpred2 Score	MUpro	ENVISION		
						Prediction	AA Polarity (WT→ MT)	Delta PI
C.8209G>A p.Gly2737Ser	0.02	0.60	C55	0.66	ΔΔG =-0.37 ↓stability	0.93	Hydrophobic → Polar	2.39
C.11947C>T p.Arg3983Trp	0	1	C65	0.92	ΔΔG = -1.01 ↓stability	0.74	Positive Charge→ Hydrophobic	-0.45

SIFT (http://sift.jcvi.org/) scores from 0 to 0.05 are considered deleterious and scores >0.05 to 1.0 considered tolerated. PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) scores range from 0 to 1, with scores < 0.15 are considered benign, whereas > 0.15 considered damaging (scores >0.85 being more confidently predicted to be damaging). AlignGVGD (http://agvgd.hci.utah.edu/) combines the biophysical characteristics of amino acids and protein multiple sequence alignments and the score is divided into seven classes (C65, C55, C45, C35, C25, C15, C0) where the higher-class numbers are more predisposed to affect the protein function. The MutPred2 (http://mutpred.mutdb.org/) is a machine learning-based method that predicts the variant pathogenicity with a prediction score that is the average of the scores from all networks and ranges between 0 and 1; higher scores (>0.6) reflect a higher probability of pathogenicity.MUpro uses support vector machines to predict protein stability changes for single-site mutations (http://mupro.proteomics.ics.uci.edu/). Envision (https://envision.gs.washington.edu/shiny/envision_new/) combines variant effect measurements from nine large-scale experimental mutagenesis datasets, with damaging variants having scores of less than one. ↓ means decrease.

To gain a better understanding of the molecular alterations introduced by the variants in the *DNAH7* gene, we localized structurally the mutant residues. Because the atomic structure of any axonemal dyneins has not been solved, the location of the mutations reported here was predicted using multiple sequence and structural alignments from a combination of cytoplasmic dynein structural models whose structure were already solved, specifically the human cytoplasmic dynein 2 (Schmidt *et al.*, 2015), human cytoplasmic dynein-1 (Zhang *et al.*, 2017), *D. discoideum* cytoplasmic dynein 1 (Kon *et al.*, 2012) and the *S. cervisiae* cytoplasmic dynein 1(Schmidt *et al.*, 2012). These sequences reveal conformational variations, which presumably are not related to sequence differences, but instead are caused by the different nucleotide states that are assumed to be caused due to power stroke (an essential mechanism to dynein movement)(Schmidt *et al.*, 2015). Therefore, conformational changes are visible depending whether structures are obtained in pre-power stroke or in post-power stroke state (Burgess *et al.*, 2003). Sequence alignment of human DNAH7 with proteins deposited in the PDB reveals that the p.Gly2737Ser mutation is located in an ATP-sensitive microtubule-binding site (MTBD) and that the p.Arg3983Trp mutation is located in the C-terminal (Fig. 3.2-5a).

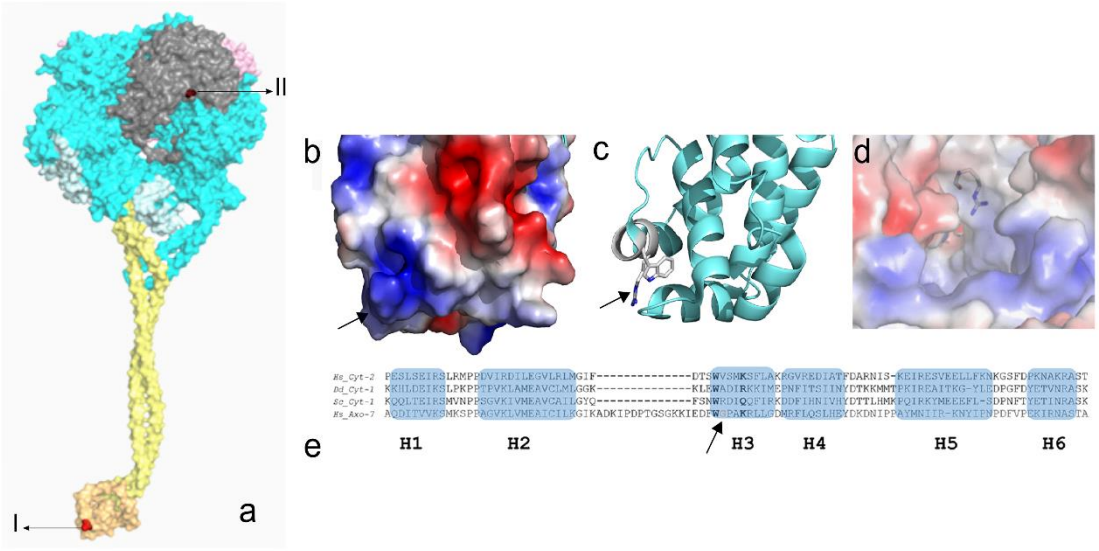


Figure 3.2-5. Structural analysis of DNAH7 variants. a The overall structure of the dynein motor domain. It consists of an N-terminal linker (pink), six AAA domains in a ring arrangement (blue), the stalk (yellow) containing the MTDB (orange) and the C-terminal (black). The position of mutations p.Gly2737Ser **(I)** and p.Arg3983trp **(II)** are highlighted in red (Crystal model PDB ID 4RH7: Crystal structure of human cytoplasmic dynein 2 motor domain). **b, c** Crystal structure of the MTBD of *D. discoideum* cytoplasmic dynein 1 (PDB ID: 3VKH). pGly2737Ser mutation is in H3. H3 is highly polar and assumes an important role in the interaction with the microtubule (electrostatic surface shown in b and H3 key residues in **c. d** close-view of the C-terminal of the crystal structure of human cytoplasmic dynein 2 motor domain (PDB ID 4RH7). The R side chain is disordered in 4RH7 model and exposed to the solvent. A putative position is shown in the figure to highlight the superficial position of the residue. **e** multiple sequence alignment, the aligned sequences are: Hs_Cyt-2-human cytoplasmic dynein 2 (PDB ID: 4RH7), Dd_Cyt-1-*D. discoideum* cytoplasmic dynein 1 (PDB ID: 3VKG), Sc_Cyt-1-*S. cervisiae* cytoplasmic dynein 1 (PDB ID: 4AKI) and Hs_Axo-7- human axonemal dynein 7. The secondary structure of Hs_Cyt-2 is indicated (alpha-helices H1 to H6).

DNAH7 codes for a DHC, which includes a AAA+ superfamily of mechanochemical enzymes, with six AAA domains that assemble into a ring structure ((Fig. 3.2-**5a, in blue**). One ring surface is covered by the N-linker (involved in dynein's power stroke) and, the opposite surface, is covered by the C-terminal (Fig. 3.2-**5a, in black**). DHC contains a stalk with an ATP-sensitive MTBD at its tip. MTBD is a small domain with 6 alpha-helical (H1-H6) at the end of the stalk responsible for binding the MT track. Alpha-helices H3 and H6 are major polar (H3 electrostatic surface shown in Fig. 3.2-5b) and establish important contact points to the microtubule(Redwine *et al.*, 2012). Alignment with cytoplasmic dynein proteins suggests that **mutation p.Gly2737Ser** occurs at MTBD H3 (Fig. 3.2-**5e**). The H3 sequence W2736XXXY2740 with Y highly polar (= K or R or Q), is conserved. W2736 is involved in an intermolecular interaction with the loop between

helices H1 and H2 and Y2740 establishes a salt bridge with the microtubule (Fig. 3.2-**5c**) (Redwine *et al.*, 2012). Nearly all of the MTBD residues involved in interactions with the MT are highly conserved, and mutating them results in defects in MT binding(Koonce and Tikhonenko, 2000; Gibbons *et al.*, 2005). Further, glycine is an unique amino-acid as it contains a hydrogen as its side chain and can play a distinct functional role, such as using its backbone to bind to phosphates, and consequently, changes in its conserved glycine might be damaging (Betts and Russell, 2003). It is thus plausible that the **p.Gly2737Ser mutation**, at MTBD H3, has an influence in the affinity towards MT. Regarding mutation **p.Arg3983Trp**, it is at the C-terminal, and despite no direct structural or functional role has been attributed to the residue, it is solvent exposed (Fig. 3.2-**5d**). Arginine is a highly polar and positively-charged amino acid that generally prefers to be on the surface of the protein(Betts and Russell, 2003), whereas tryptophan is a nonpolar residue that prefers to be buried in protein hydrophobic cores. Therefore, this may affect protein stability. Further, according to the web server MUpro, both variants gave a negative ΔΔG, which corroborates the prediction of decreasing the protein stability(Cheng *et al.*, 2005).

Gene variants in DNAH5 and DNAH7 genes lead to a decrease in patients mRNA expression and protein mislocation

To further explore the molecular consequences of the genomic variants identified in *DNAH5* and *DNAH7* genes from child 2, we analysed mRNA expression in nasal epithelial cells from this patient and his parents compared to controls. The *DNAH5* mRNA relative expression of child 2, who is a compound heterozygous for a frameshift and a nonsense variant, and his carrier parents, was about 0.2-fold lower comparing with healthy volunteers (Fig. 3.2-6), suggesting that those variants are harmful enough to *per si*, affect gene expression. The *DNAH7* mRNA relative expression of child 2, who is a compound heterozygous for two missense variants, was 0.5-fold lower comparing with healthy volunteers (Fig. 3.2-6). Carrier parents did not show statistically significant differences in *DNAH7* mRNA relative expression in comparison to controls.

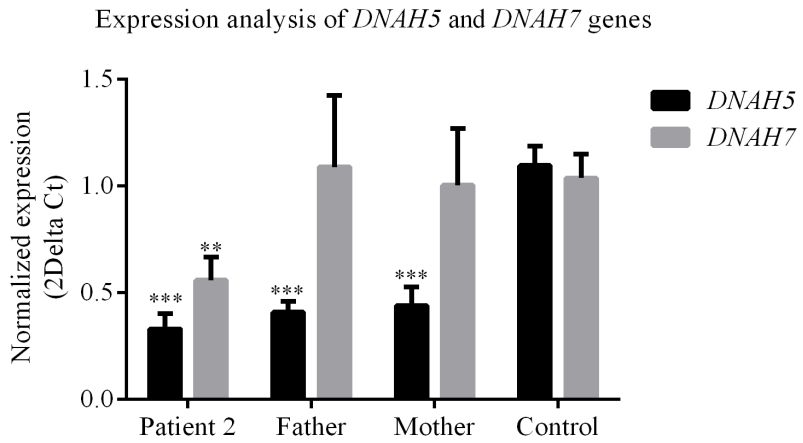


Figure 3.2-6. RT-PCR analysis of *DNAH5* and *DNAH7* mRNA expression levels in epithelial respiratory cells from Patient 2 and his parents, compared to healthy donors (a fluorescent dye used: SYBR Green). ECM7 was used as the reference gene. For *DNAH5*, patient and parents presented significantly lower levels compared to controls. For *DNHA7*, only the patient showed lower levels of controls. Statistical significance determined using the Mann-Whitney test, with alpha < 0,05. ** = P value < 0.01; *** = P value < 0.001.

Additionally, to assess the functional significance of the detected frameshift and nonsense variants in the *DNAH5* gene, as well from both missense variants in the *DNAH7* gene, we performed IF analysis in respiratory cells of child-2 and his carrier parents. In cells from healthy volunteers, DNAH5 intense staining was observed only in cilia axonemes. In child-2, DNAH5 staining was undetectable in cilia axonemes and a faint staining appeared in the cytoplasm. In the mother, a faint staining was observed in cilia axonemes and in the cytoplasm. In the father, staining was absent from cilia axonemes, with presence of an intense staining in the cytoplasm (Fig. 3.2-7).

In both patient and parents, it was observed in DNAH5, a statistically significant reduction of mRNA expression and a mislocation of the DNAH5 protein, with cytoplasmic accumulation instead of distribution along the ciliary axoneme as observed in control individuals. Previous reports also presented this same DNAH5 staining pattern in PCD patients with *DNAH5* mutations (Fliegauf *et al.*, 2005; Hornef *et al.*, 2006). Moreover, those authors observed a different staining pattern of DNAH5 depending on the type of mutation, demonstrating that there is a high regulation of protein trafficking, which are still not fully understood. Previous reports stated that the correct subcellular localization of proteins is as important as the amount of functional protein produced for

their function on cells and that an aberrant localization of proteins are implicated in the pathogenesis of several human diseases (Hung and Link, 2011). Further studies on protein location are vital in PCD patients, as it may help to bridge genotype and phenotype severity. Altogether, it gives an additional piece of evidence that both variants are causative and could justify the lack of ODA in our patient.

The fact that both parents presented a reduced *DNAH5* mRNA expression and mislocation of the DNAH5 protein, but do not exhibit the PCD phenotype, could be explained by a phenomenon described as variable expressivity, in which individuals carrying the same variant may display a myriad of phenotypes (Castel *et al.*, 2018). This agrees with the recessive pattern of the disease and demonstrates that the parents are asymptomatic carriers of single-mutant alleles, causing disease in the presence of homozygous or compound heterozygous variants. This can also support the existence of highly regulated modes of translation and/or protein-degradation regulation that differ in compound heterozygous leading to a severe phenotype. Furthermore, this data highlights that the correlation between mRNA expression levels, protein function and the severity of the phenotype is still a subject poorly understood, which needs further enlightenment.

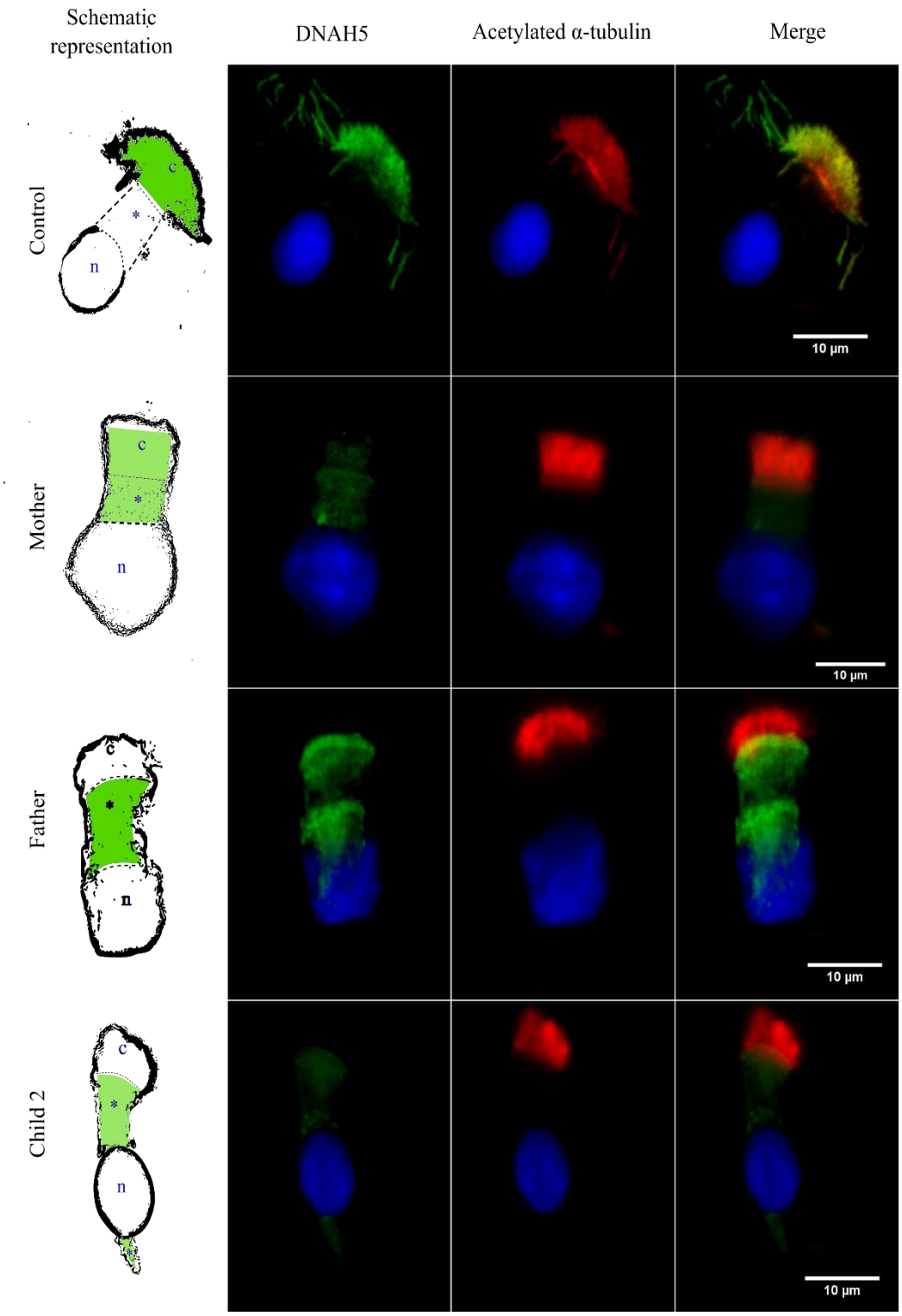


Figure 3.2-7. Expression and localization of specific DNAH5 protein by immunofluorescence in respiratory epithelial cells from healthy volunteers (Control), both parents of child 2 (Mother and Father) and child 2. Staining with anti-DNAH5 antibodies (green) and with antibodies against axoneme-specific acetylated α-tubulin (red). Nuclei stained with DAPI (blue). Schematic representation at left to better localize the staining (DNAH5 in intense or faint green): images were obtained with Inkscape 0.92.3 (2405546, 2018-03-11); c = cilia; * = cytoplasm; n = nuclei.

Regarding **DNAH7,** intense staining was observed in cilia axonemes of controls. In child-2, staining was absent form cilia axonemes, with a faint staining appearing in the cytoplasm. In the mother, an intense staining was observed in cilia axonemes. In the father, staining was absent form cilia axonemes, with an intense staining observed in the cytoplasm (Fig.3.2-8). It was observed a reduced mRNA expression of the *DNAH7* gene in patient nasal cells, but no significant changes were observed in his parents. Interestingly, it was observed a mislocalization of the DNAH7 protein to the cytoplasm in child-2 and his father, but not in his mother, which presented a typical location along the ciliary axoneme. The variant present in his mother is located at **C-terminal (c.11947C<T; Fig. 3.2-5)**, and thus data may suggest that the C-terminal may not be critical for protein location, in contrast to **the father's variant (p.Gly2737Ser),** already described as having an influence in the affinity towards microtubules, that seems to be important for DNAH7 location.

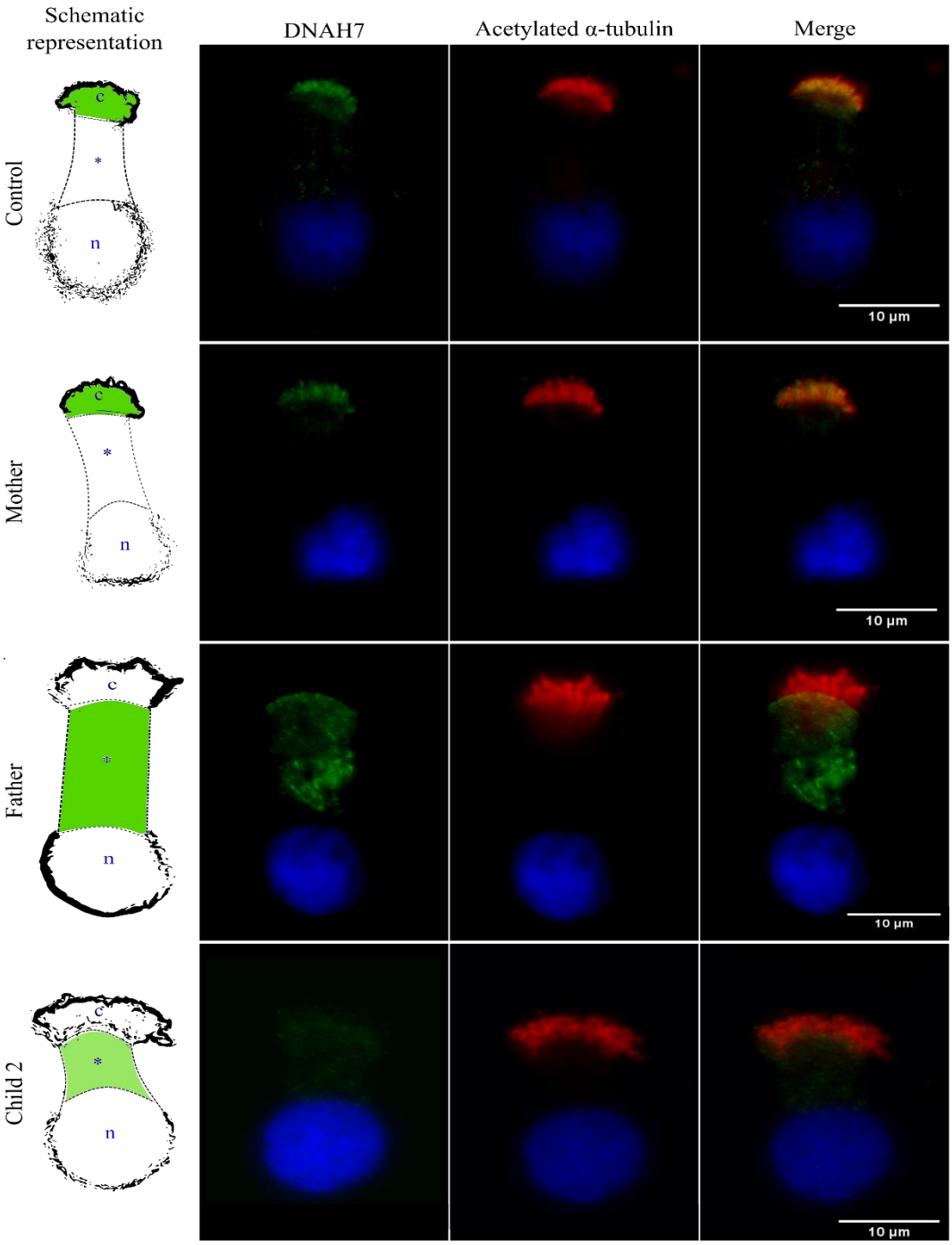


Figure 3.2-8. Expression and localization of specific DNAH7 protein by immunofluorescence in respiratory epithelial cells from healthy volunteers (Control), both parents of the child 2 (Mother and Father) and Child 2. Staining with anti-DNAH7 antibodies (green) and with antibodies against axoneme-specific acetylated α-tubulin (red). Nuclei stained with DAPI (blue). Schematic representation at left to better localize the staining (DNAH7 in intense or faint green): images were obtained with Inkscape 0.92.3 (2405546, 2018-03-11); c = cilia; * = cytoplasm; n = nuclei.

To the best of our knowledge, no pathogenic variant in *DNAH7* gene has been published so far and no study has shown the interaction of two genes in the PCD

phenotype. Considering the information given by previous knockout models and PCD complexity and heterogeneity, we hypothesize that it may exist an effect of *DNAH7* variants over the phenotype. Either genes products when defective may have thus caused the absence of both DA in Patient-2. The variants in *DNAH5* are potentially more damaging and are thus expected to be the major responsible for the child-2 phenotype. Nevertheless, *DNAH7* may have an accumulative effect, leading to the complete loss of both DA. Our functional data corroborates these assumptions as we observed a reduction of gene expression and protein mislocalization in both genes.

Nonetheless, we are aware that the data is still limited, as it represents only a small pedigree and thus requires further functional evidences. However, if such model is proven to be realistic, it would certainly increase the accuracy of the genotype-phenotype correlations in PCD. WES is undoubtedly a powerful and determinant technique for the discovery of new genetic causes of disease. The possibility of multiple variants in different genes contributing to the same phenotype opens a new door in genomic research and brings an additional important advantage of applying WES or multi-gene NGS panels over the (single) gene-by-gene approach by Sanger sequencing. Additional studies are pivotal to understand the effects of variants in multiple genes on the PCD phenotype. Further, it is also critical to increase our knowledge regarding the biological mechanisms that control the regulation of mRNA levels and protein abundances/degradation, as well as the co-translational and post-translational mechanisms that regulate protein location. Nowadays, the main challenge placed in complex genetic diseases, as PCD, is not only to know which genes and variants are associated with the phenotype, but also to obtain refined genotype-phenotype correlations, for each variant and the different genetic factors, explaining why two individuals holding the same genotype or similar variants may develop dissimilar phenotypes. Generating that knowledge would be critical to unveil the complexity of PCD, as well as their implications in reproductive function and fertility.

In conclusion, in child-1 we identified two compound heterozygous variants in the *CCDC40* gene, one affecting the splicing region, not indicated in ClinVar, and one frameshift, previously indicated in ClinVar, both here firstly reported. Previous studies

found that *CCDC40* pathologic variants are often of high impact (Attachment 3) and patients presented worst clinical features(Davis *et al.*, 2015). Our data corroborate these associations. Our studied child 1 presented a heterogeneous pattern of ultrastructural cilia anomalies, higher ciliary beat axis deviation, high impact variants, and worst clinical features. Taken together, these variants are highly suggestive to explain the ultrastructural anomalies, the *situs inversus*, and the severe clinical features observed in patient 1. Further, this indicates that although already described as a common PCD gene, new variants are still to be reported, difficulty standard genetic diagnosis.

In relation to Patient 2, we identified four compound heterozygous variants; two novel in the *DNAH5* gene, one frameshift and another nonsense; and two in the *DNAH7* gene, both missense and not indicated in ClinVar, all here firstly reported. Functional studies revealed decreased mRNA expression and protein mislocalization with cytoplasmic accumulation. These observations draw the attention for the possibility of the interaction of these two genes in the patient phenotype.

3.2.2. Unveiling genetic ethology of primary ciliary dyskinesia: when standard genetic approach is not enough

Adapted from: Pereira et al, 2019, Advances in Medical Sciences,
https://doi.org/10.1016/j.advms.2019.10.003

In this section, I will present the last child with KS that during this thesis was subjected to clinical assessments, ultrastructural analysis of motile cilia by transmission electron microscopy (TEM) and extensive genetic evaluation. In contrast to previous children, the identification of the genetic aetiology of this child was challenging to unveil. WES and CNV analysis were performed and later gene expression analysis and immunofluorescence to further study genetically this child.

Ciliary ultrastructure is disrupted

Our studied male child 3 has now 2 years and 9 months of age and presented clinical features compatible to KS.

Ultrastructural analysis of the axoneme showed the absence of DA and nexin links, with the RS and CPC normal (**Fig. 3.2-9**). Evaluation of 152 axonemes revealed a variation in the ciliary beat axis and ciliary deviation of 26.71, which is compatible with a bronchiectasis severity in cilia axis movement, also a PCD feature (De Iongh and Rutland, 1989). All measurements of ciliary beat axis and ciliary deviation were conducted always by the same observer.

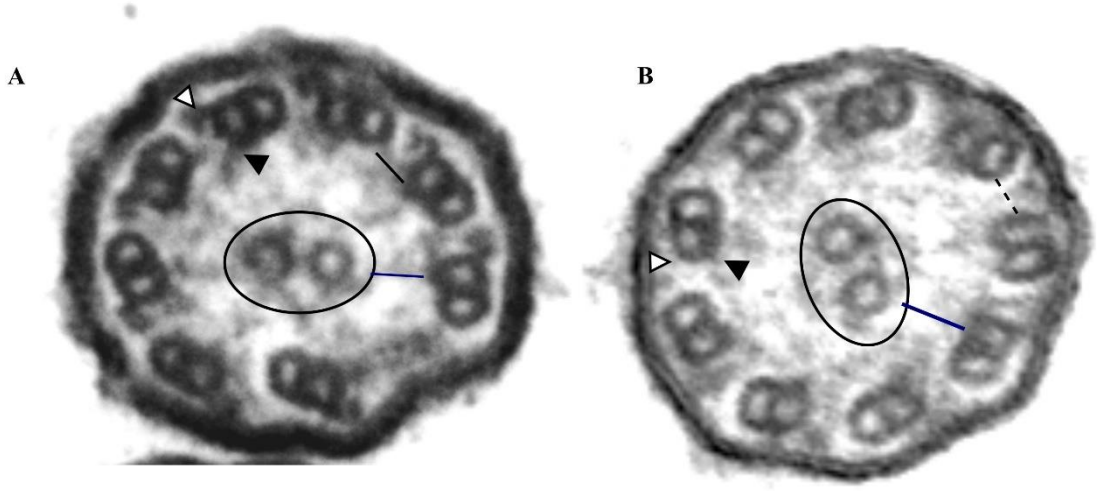


Figure 3.2-9. Ultrastructure of cilia axoneme from a control individual (**A**) and from child 3 (**B**). The axoneme is composed of nine peripheral pairs of microtubules and two single central microtubules. From the inner peripheral microtubule emerge two dynein arms: an outer (white arrowhead) and an inner (black arrowhead). The DRC that links the microtubule doublets are represented by black line. The radial projections, that bind the doublets to the two central microtubules (circle) are indicated by blue line. Note, that in this child (**B**) both dynein arms are absent, as well as, the absence of nexin bridges (dotted black line), with normal presence of the radial spokes (blue line) and central microtubules pair (black circle).

Exome sequencing metrics, analysis and filtering

To understand the genetic aetiology of PCD we perform a WES analysis of the patient. The exome was analysed according to the autosomal recessive disease model. In the first filter strategy applied (**filter 1**), we filter-out the more deeply intronic placed variants, synonymous substitutions and variants with allele frequency above 1% in human population databases, assuming that are polymorphisms and thus unlikely to be disease-causing variants (Bamshad *et al.*, 2011). Further, we selected the variants based on its predicted pathogenicity with algorithms included in Ion Reporter™ Software and VarAFT. From the obtained variants, we further selected the ones that gene ontology (GO) (Ashburner *et al.*, 2000) keywords were related to cilia and ciliopathies, namely cilia/flagellum; Primary ciliary dyskinesia, cilium movement, cilium beat, cilium morphogenesis and organization, cilium assemble, axoneme assemble, cilium component, axoneme component, determination of left/ right axis and symmetry, cell projection, dynein arm, locomotion and development. In this filter we also excluded the *HYDIN* gene variants, due to the high false discovery rate in mutation calling of this gene (Pongor *et al.*, 2015). The variants in the genes listed on Suppl.

table S2, were excluded as well, because are genes that were detected by our in-house database as false positive. Subsequently, variants predicted as pathogenic, by at least two bioinformatic predictors, were selected. This analysis resulted in a negative genetic test as no candidate variant were identified.

To procedure the search for a disease-causing variant in this PCD child, we broaden the search also to other genes that according to protein-protein association networks STRING database (Szklarczyk *et al.*, 2019), were associated to the previous gene list obtained from GO search (**Filter 2**). All the remaining filter condition from Filter 1, were maintained. Again, no candidate variant was identified. Thus we, enlarge once more the gene list, by include the GO keywords: cell motility, embryo development, embryonic organ development, embryonic organ morphogenesis, establishment or maintenance of cell polarity, flagellum motility, intraflagellar transport, lung development, reproduction, sperm development and spermatogenesis (**Filter 3**).

Given the negative results obtained, we widen the search to also include more intronic and synonymous variants that despite commonly assumed as improbable to be disease causing (Bamshad *et al.*, 2011), and therefore removed in **Filter 1**, those variants may indeed affect the splicing and consequently affect gene and protein function (Cartegni *et al.*, 2002). Moreover, we included also single heterozygous variants to infer about a possible *de novo* mutation, and we do not exclude genes according to GO. In that sense, in this **Filter 4** we remove variants with allele frequency above 1% in human population databases, homopolymers longer than 6 bp, and filter the variants according to the predicted pathogenicity with algorithms included in Ion Reporter™ Software and VarAFT. Algorithms included Alamut were used to analyse all the variants that could potentially affect the splicing, namely the more intronic and synonymous variants. Only variants that were predicted as pathogenic according to, at least, two bioinformatic predictors were included to further analysis. All suspected variants were manually checked on the BAM file through GenomeBrowse and then validated through Sanger sequence in patient, his parents and sister. Regarding the single heterozygous variants, we excluded all as they are also present in one of the progenitors, thus excluding the hypothesis of *de novo* mutation.

This last filter approach end-up with two candidate disease causing variants in two gene not previously reported as PCD genes: *cytochrome b-245 (CYBB)* and *Ubiquitin Specific Protease 11 (USP11)*. In the *CYBB* gene we found rare (0,0015%) a hemizygous missense variant,

NM_000397.3: c.1103C>T; p.(Ala368Val) (Fig. 2), that is considered as disease causing by multiple bioinformatic predictors, namely Poly-phen (http://genetics.bwh.harvard.edu/pph2/) predicts as possibly damaging with a score of 0.934 , as well as, Condel, that is a method to assess the outcome of non-synonymous SNVs using a CONsensus DELeteriousness score, with a score of 0.55. The unaffected mother and sister are carriers, whereas the father does not present the variant (Fig. 3.2-10a). Regarding the variant in *USP11* gene is a rare (0,0011%) a hemizygous synonymous variant, NM_004651.3: c.1599G>A; p.(=) (Fig. 2), that is predicted to affect the splicing by several tools included in Alamut software, likely by generation of a new acceptor splice site which may compete with the canonical site (Fig. 3.2-10b). In the same way that the variant in *CYBB* gene, the unaffected mother and sister are carriers, whereas the father do not present the variant (Fig. 3.2-10a). Both variants were firstly reported in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/).

The *CYBB* gene codes for a protein called cytochrome b-245, beta chain (also known as p91-phox), that is included in a group of proteins that forms an enzyme complex called NADPH oxidase, which plays an essential role in the immune system (Bustamante *et al.*, 2011). The complex cytochrome b-245, has two chains: the alpha chain (produced from the CYBA gene) and the beta from CYBB). Both alpha and beta chains are required for either to function and the NADPH oxidase complex requires both chains in order to be functional. The *USP11*, originally named ubiquitin C-terminal hydrolase on the X chromosome (UHX1), but recently renamed USPs (Ideguchi *et al.*, 2002). Propose as having a role in hereditary X-linked retinal disease (Ideguchi *et al.*, 2002). USP11 bound specifically to RanBPM and inhibited its ubiquitination and degradation (Ideguchi *et al.*, 2002). By its turn, RanBPM has been implicated in the regulation of multiple pathways and cellular processes and has been hypothesized to be a scaffolding protein. Therefore, is likely that USP11 may also be linked to other diseases.

No genetic variants in any of the genes were previously reported in PCD patients.

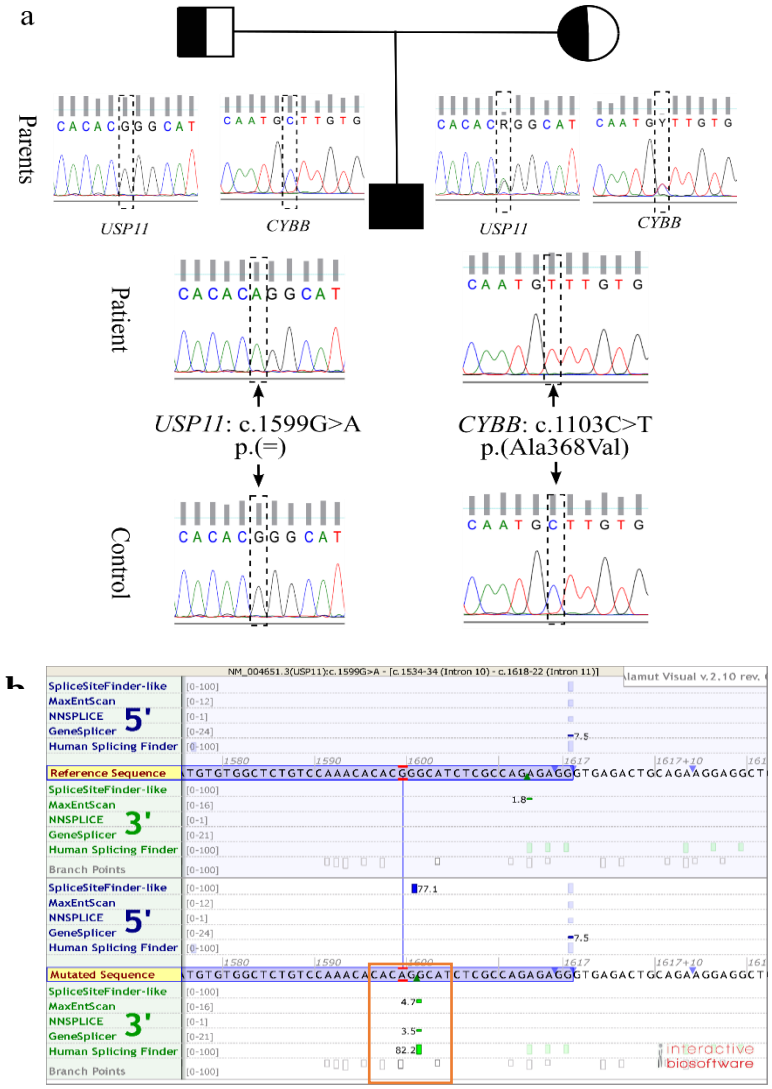


Figure 3.2-10.a) Pedigree and sequencing electropherograms of the variants in *CYBB* and *USP11* genes identified in this study, **b)** Bioinformatic analysis of the c.1599G>A; top panel = reference sequence; bottom panel = mutated sequence, with the respective predictive scores for each algorithm; the reference sequence and mutated sequences are shown at the bottom of each panel with nucleotide positions indicated above; green vertical bars indicate 3' (acceptor) sites scores; the orange box highlights the altered scores of the acceptor splicing site. Figure obtained from Alamut Visual software (v2.10).

CYBB and USP11 mRNA relative expression

To explore the molecular consequences of the genomic variants identified in *CYBB* and *USP11* genes, we analysed mRNA expression in nasal epithelial cells from this patient and his parents and sister compared to controls. Regarding *CYBB* no differences were identified in the mRNA relative expression between patient, parents, sister and controls (Fig 3.2-11a). Leading us to classify this variant as a variant of uncertain significance (VUS).

Regarding the synonymous variant in *USP11* gene, the patient and his carrier mother and sister, show a reduction in *USP11* mRNA expression of about 0.6-fold comparing with healthy volunteers (Fig. 3.2-11b). The mother and sister presents a similar pattern. This

suggests that those variants are harmful enough *to per si*, affect gene expression, as in controls and the father, who do not present the variant.

To comprehend the genetic mechanism that underlines the mRNA reduction observed in carriers from *USP11* variant, the *USP11* cDNA was sequenced aiming to verify if could be due to exon skipping. No changes in cDNA were observed. The variant c.1599G>A in *USP11* gene may lead to the activation of cryptic acceptor splice sites, thus a possible consequence is the generation of several alternative transcripts, with the wild-type transcript still being present (Anna and Monika, 2018). If this assumption is correct, it may justify the fact that we do not observe changes in the cDNA sequence analysed.

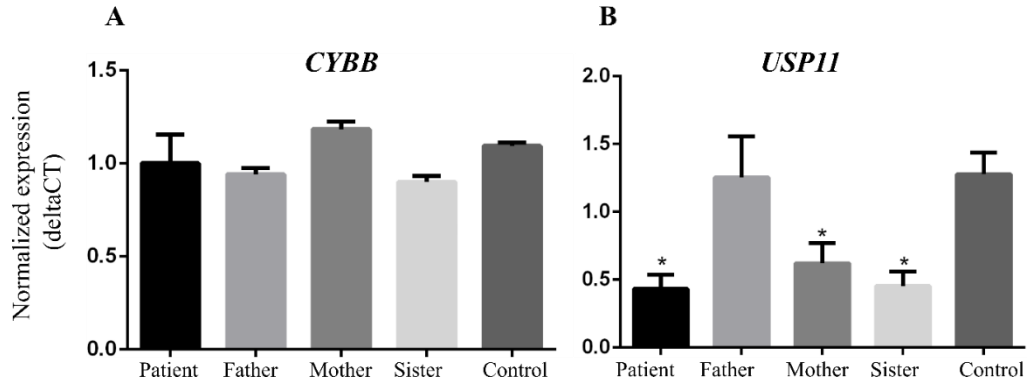


Figure 3.2-11. RT-PCR analysis of CYBB (A) and USP11 (B) genes mRNA expression levels in epithelial respiratory cells from patient and his parents and sister, compared to healthy donors (fluorescent dye used: SYBR Green). B2M and GAPDH was used as the reference gene. Statistical significance determined using the Mann-Whitney test. * P value < 0,05.

USP11 immunofluorescence analysis

The results obtained by RT-PCR prompted us to further explore the variant c.1599G>A found in *USP11* gene. To that we immunolocalize the USP11, through IF analysis, in respiratory cells from the present pedigree to further assess the functional significance of the detected variants. From the best of our knowledge, is the first report of IF analysis in respiratory cells of this protein. In cells from healthy volunteers and in father (that do not have the variant), USP11 staining was observed in the cytoplasm, in the base of the cilia, as well as in the cilia. In the mother and in the sister, staining was observed also in cytoplasm and cilia and although with a slightly reduction in intensity, namely in the sister. In Patient, USP11 staining was detectable also in the cytoplasm but staining is looked less intense, as in his sister, whereas the staining in the cilia or base is almost undetectable. Of note, is the fact that this patient, as well as his sister, who also present milder respiratory complications

without situs inversus, the ciliated cells are rare to observe and fragile which difficulted the IF analysis (Fig. 3.2-12).

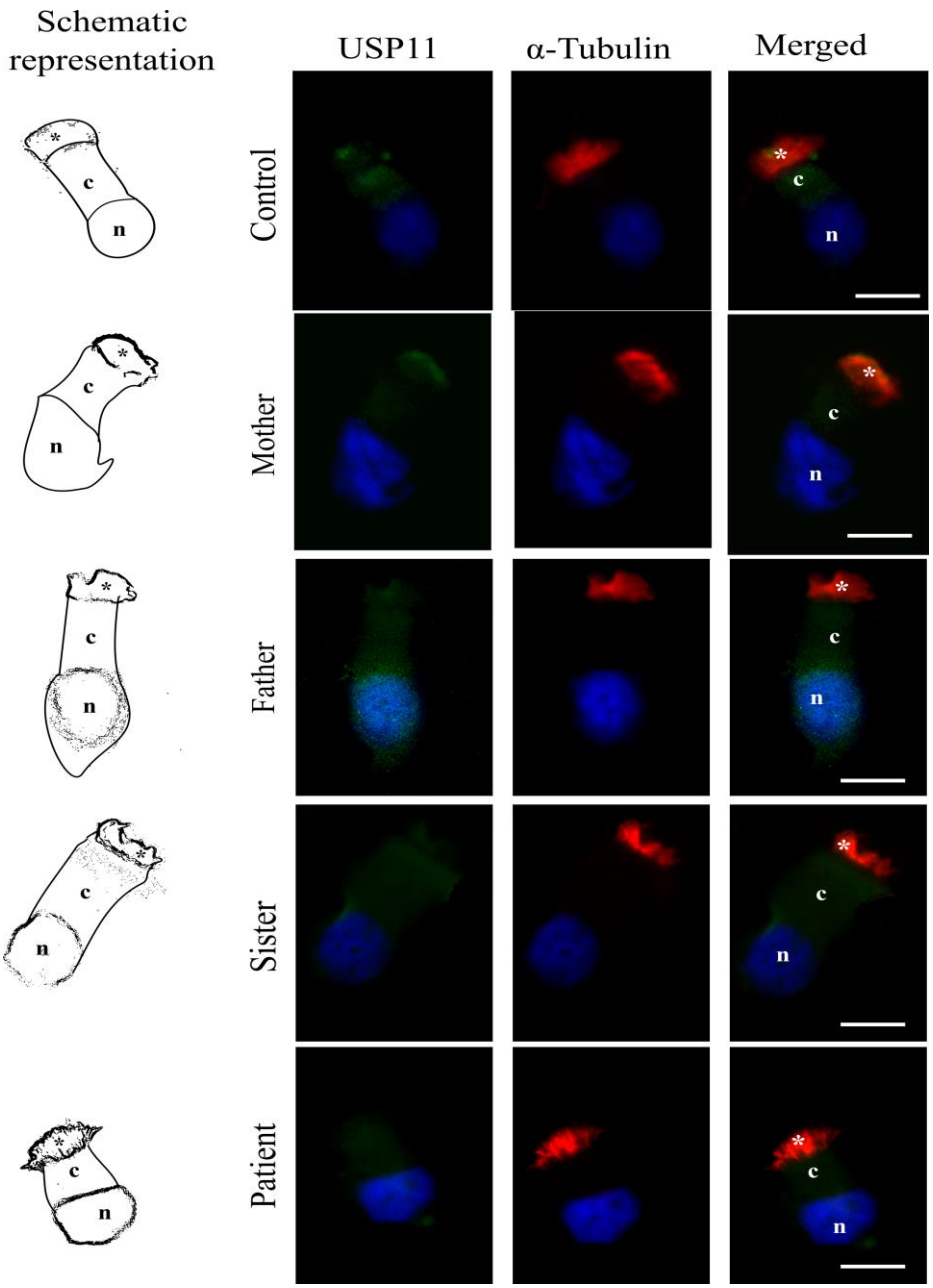


Figure 3.2-11. Expression and localization of specific USP11 protein by immunofluorescence in respiratory epithelial cells from healthy volunteers (Control), both parents (Mother and Father) of patient (Child 3), the sister and the Patient. Staining with anti-USP11 antibodies (green) and with antibodies against axoneme-specific acetylated α-tubulin (red). Nuclei stained with DAPI (blue). * = cilia; c = cytoplasm; n = nuclei.

Copy number variation analysis

WES is undoubtedly of up most importance in genetic analysis of PCD. Nevertheless, it is far from being a perfect solution, as lacks 99% of genomic information that may be involved in disease (Andersson *et al.*, 2014; Zhang and Lupski, 2015), and about 30% of PCD cases remain with inclusive genetic result (Kim *et al.*, 2014; Jackson *et al.*, 2016; Leigh *et al.*, 2016). Knowing that and given that the first approaches to unveil the genetic etiology in this patient gave inconclusive results, to find a genetic cause that may help to understand the ultrastructural anomalies found in patient's axoneme and his clinical features, we further performed a chromosomal microarray that is capable to detect regions of genomic imbalances, also known as copy number variation (CNV), at a higher resolution than classical techniques. A CNV arises when the number of copies of a segment of a chromosome, that could range from few tens base pairs (bps) to megabases (Mbs), differs from the typical two copies for autosomes and X chromosomes in females, due to gains or losses. CNVs are a major source of genomic diversity in human populations (Redon *et al.*, 2006). CNVs were as well observed in autosomal-recessive Mendelian disorders (Stankiewicz and Lupski, 2010) and already found in PCD patients (Loges *et al.*, 2009; Hjeij *et al.*, 2013; Knowles *et al.*, 2013c). Those may not be detectable by WES or Sanger sequencing, demonstrating the complexity of the genetic assessment of patients with PCD.

After chromosomal microarray, twenty-six CNVs were detected. Eight were discard as were verified in multiple in-house chromosomal microarray, which is highly suggestive that those are polymorphisms. The remaining were analysed in Decipher database (https://decipher.sanger.ac.uk/) to infer if those CNVs were already described. The exact CNV segment were not described for any case, but was observed that a larger segment that includes the CNV region were described as pathogenic for the eighteen cases. To further restrict the list, we search for CNV in genes related to PCD/KS. No CNV were found in any of the typical PCD genes (see Table 1 from introduction for further information). Thus, we searched for CNV in genes that could have an indirect relation to PCD. From this, five CNV segments were select to further study: three losses in the genes *ABCA3, CATSPER2* and *CRHR1* genes; and two gains in genes *KTR34* and *PDE4DIP* genes (Fig.3.2-13).

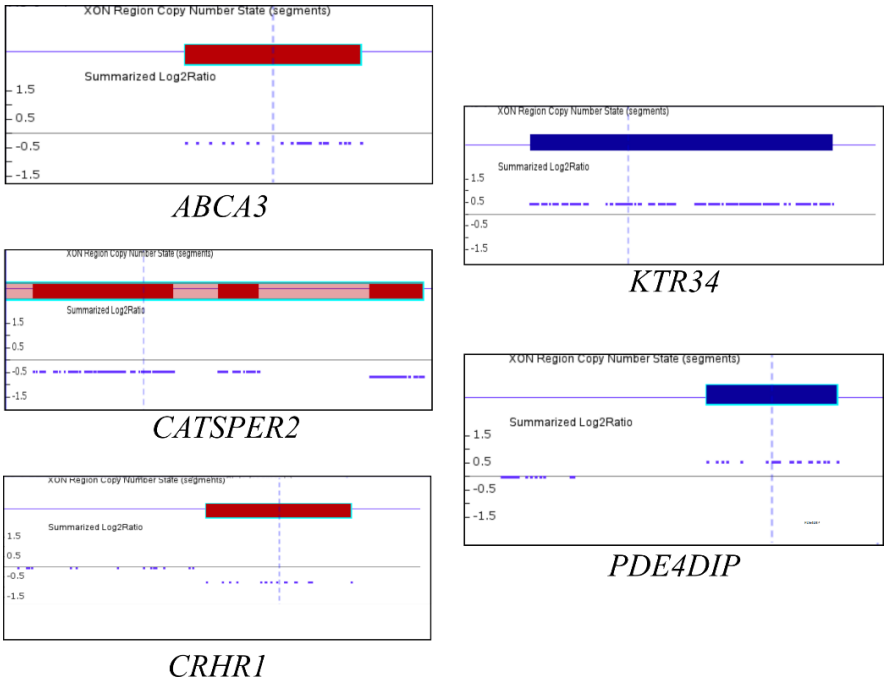


Figure 3.2-13. Results from analysis with Affymetrix Gene Chip1 Scanner 3000 7G viewed with Chromosome Analysis Suite. Values smaller than -0.3 are categorized as a "loss". CNV loss are represented in the left, in red, and were observed in *ABCA3* [16p13.3(2358100_2358648)], *CATSPER2* [15q15.3(43899889_43942147)] and *CRHR1* [17q21.31(43912285_43912973)] genes. Values larger than 0.3 are categorized as a "gain". CNV gains are represented in right, in blue, and were observed in *KTR34* [17q21.2(39533408_39539128)] and *PDE4DIP* [1q21.1(145017426_145018422)] genes.

ATP-binding cassette sub-family A member 3 (ABCA3) gene codes for ABCA3 protein that is a member of the ATP-binding cassette family, which are transmembrane proteins involved in membrane trafficking. The ABCA subgroup are predominantly involved in lipid transport across membranes. The gene for *ABCA3* gene is expressed in type II cells from pulmonary alveoli and the protein localizes to lamellar bodies (Mulugeta *et al.*, 2002). Type II cells are known to secrete pulmonary surfactant to lower the surface tension of water, helping with gases exchange. This lead to propose ABCA3 as having an important role in surfactant metabolism and ABCA3 transporter mutations have been identified as genetic causes newborn respiratory distress (Clark and Clark, 2005). A typical PCD feature is newborn respiratory distress (Ferkol and Leigh, 2006), therefore, *ABCA3* could indeed be related by a still not identified mechanism to PCD.

Cation channel sperm-associated protein 2 (CATSPER2) is a voltage-gated channel, one of the four proteins that compose the cation channels of sperm, which are responsible to form a Ca^{2+} permeant ion channel specific essential for the correct function of sperm cells (Avidan *et al.*, 2003; Qi *et al.*, 2007; Hildebrand *et al.*, 2010; Lishko and Kirichok, 2010). Mutations in

CATSPER2 were associated to deafness-infertility syndrome (DIS), characterized by early-onset deafness in both males and females and exclusive male infertility and hear impairment (Hildebrand *et al.*, 2010).

Corticotropin-releasing hormone receptor 1 (CRHR1) encodes a G-protein coupled receptor that binds neuropeptides of the corticotropin releasing hormone family that are major regulators of the hypothalamic-pituitary-adrenal pathway. *CRHR1* is the primary receptor mediating the release of adrenocorticotrophic hormone, which regulates endogenous cortisol levels and genetic variation in CRHR1 that is associated with improved pulmonary function response to inhaled corticosteroids (Tantisira *et al.*, 2004).

Keratin, type I cuticular Ha4 (KTR34) is a member of the keratin gene family. No direct association to cilia/flagellum or PCD was detected but this protein was detected proteomic analysis of multiple 9+0 cilia from choroid plexus epithelial cells from swine, in which components of motile cilium (9+2) were found, showing that both structures share genes (Narita *et al.*, 2012).

Phosphodiesterase 4D interacting protein (PDE4DIP) serves to anchor phosphodiesterase 4D to the Golgi/centrosome region of the cell. Defects in this gene may be a cause of myeloproliferative disorder (MBD) associated with eosinophilia. PDE4DIP (or Myomegalin) was also detected in protein studies of cilia (Narita *et al.*, 2012) and involved in regulation of microtubules at the centrosome (Roubin *et al.*, 2013).

Expression analysis

To further infer the effect of those CNVs in patient, the expression of these genes was analyzed through RT-PCR (Fig 3.2-14).

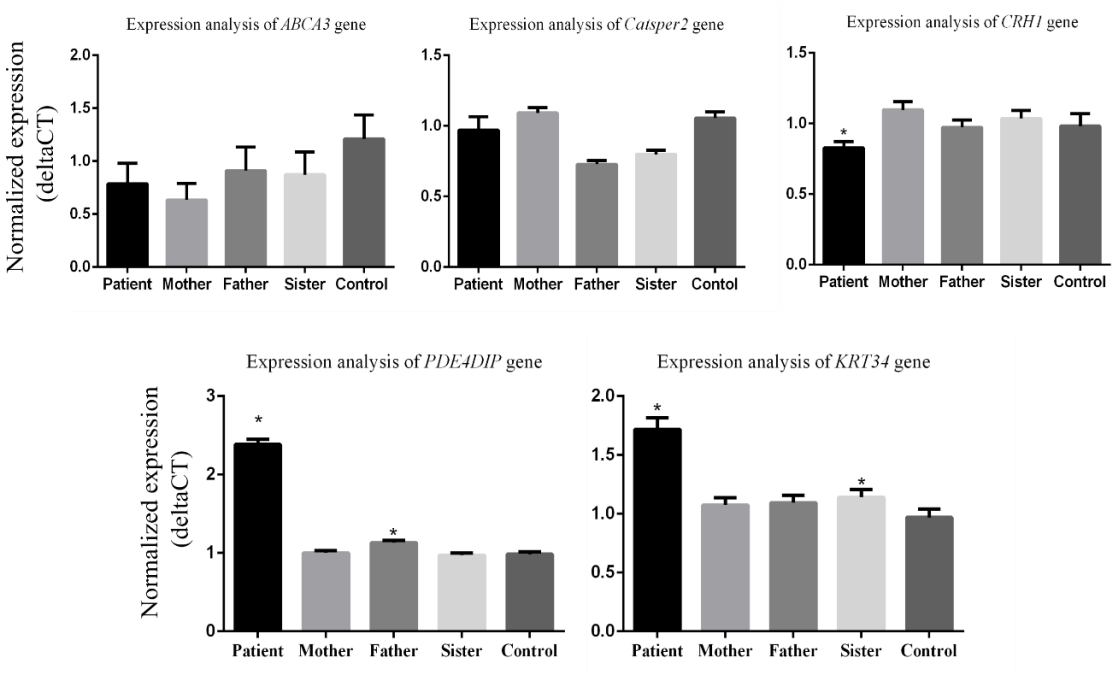


Figure 3.2-14. RT-PCR analysis of *ABCA3, CATSPER2, CRHR1, KTR34* and *PDE4DIP* genes expression levels in DNA extracted from peripheral blood cells from patient, his parents and sister, compared to healthy donors (fluorescent dye used: SYBR Green). *B2M* and *ACTA1* was used as the reference gene. Statistical significance determined using the Mann-Whitney test. * P value < 0,05.

From this analysis, we excluded the genes *ABCA3* and *CATSPER2* as candidate genes as no reduction in gene expression were verified. Suggesting that the CNV observed in these genes may be false positive. We also exclude the gene PDE4DIP as significative increase expression in comparison to controls were also observed and no clinical features are presented by the father, thus lead us to think that this could be a polymorphism without clinical significance. Regarding the *CRHR1* gene was observed a fold decrease of 0.16 only in this PCD child. Moreover, we observed a 0,7 fold increase in *KRT34* gene in this patient and a 0,3 fold increase, comparing to controls, in his sister, who present some respiratory complications. No association to PCD/KS is known for any of these genes, however given the complexity of cilia/flagellum proteome is plausible that many molecular components are still unknown. Further KTR34 were previously detected in ciliary proteome of mammalian cells (Narita *et al.*, 2012).

Immunofluorescence analysis

To increase the knowledge about these genes IF analysis were performed in respiratory cells of the presented PCD patient and his family (Fig. 3.2-15 and 3.2-16).

Relatively to CRHR1 (Fig. 3.2-15) in cells from healthy volunteers, mother and father CRHR1 staining was observed mostly in cytoplasm although part of the cells also displays staining in the cilia, as is shown in parents' cells. The cells from Patient and his sister, were hard to find, nevertheless staining was apparent to be present in both cytoplasm and cilia.

Regarding KTR34 (Fig. 3.2-16) in cells from healthy volunteers, mother and father staining was observed in cytoplasm and base of the cilia. The cells from patient and his sister, were again hard to find, nevertheless staining apparently disappear from cytoplasm being detectable only in the cilia.

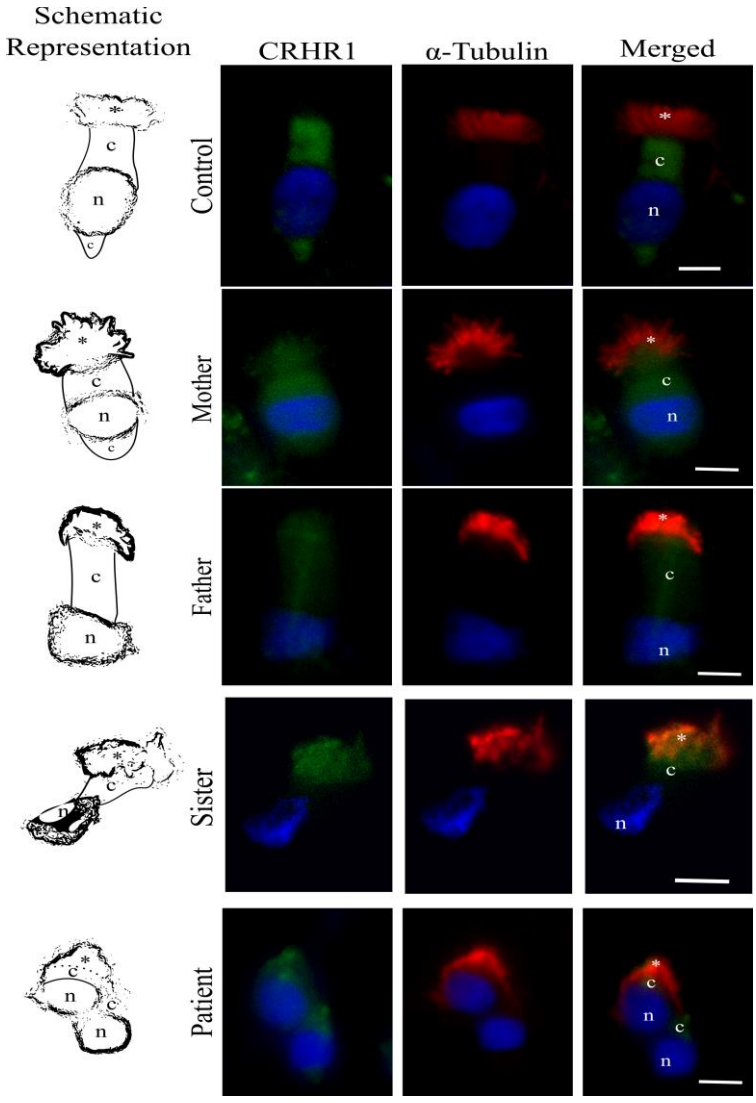


Figure 3.2-15. Expression and localization of specific CRHR1 protein by immunofluorescence in respiratory epithelial cells from healthy volunteers (Control), both parents (Mother and Father) of patient (Child 3), the sister and the Patient. Staining with anti-CRHR1 antibodies (green) and with antibodies against axoneme-specific acetylated α-tubulin (red). Nuclei stained with DAPI (blue). * = cilia; c = cytoplasm; n = nuclei.

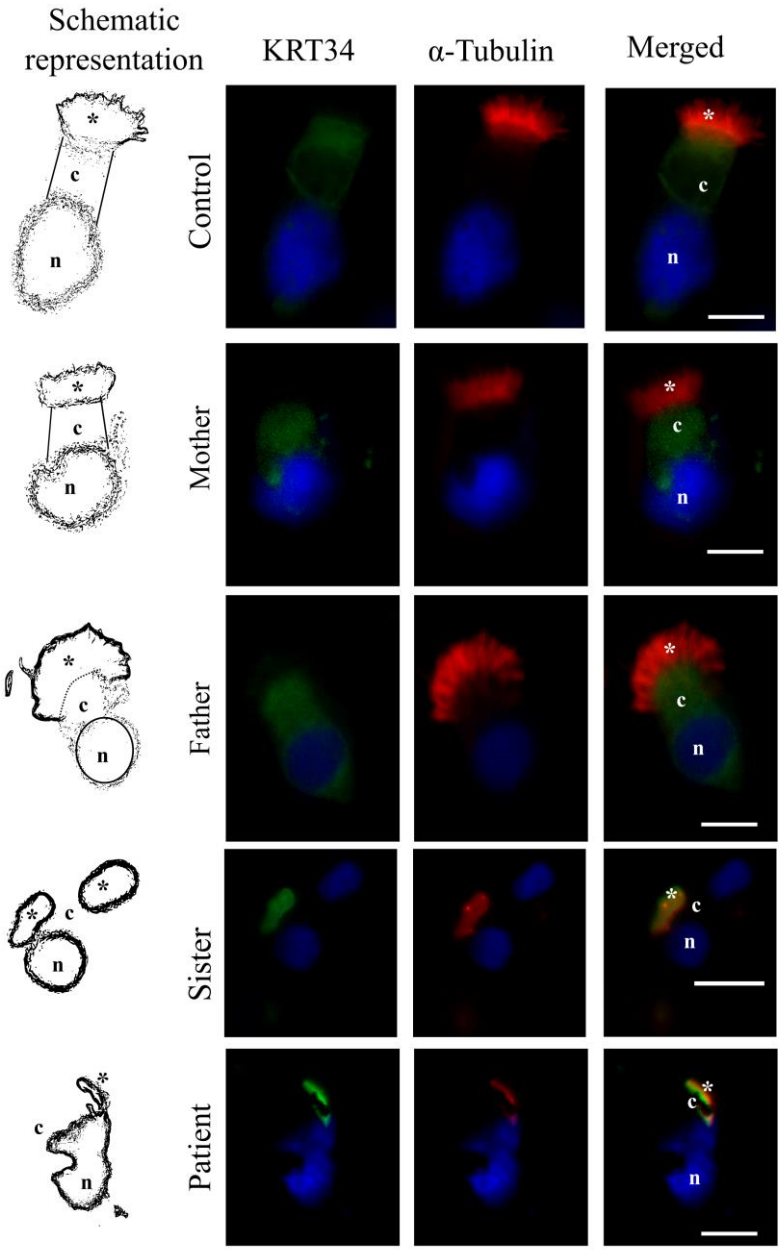


Figure 3.2-16. Expression and localization of specific KTR34 protein by immunofluorescence in respiratory epithelial cells from healthy volunteers (Control), both parents (Mother and father), the sister and the patient. Staining with anti-KTR34 antibodies (green) and with antibodies against axoneme-specific acetylated α-tubulin (red). Nuclei stained with DAPI (blue). * = cilia; c = cytoplasm; n = nuclei.

Although no direct association to PCD was reported, label-free quantitative proteomics demonstrated that USP11 interacts with Syndesmos (SDOS) protein. SDOS was showed to regulate several mRNAs involved in the biogenesis of the primary cilium and to control the cilia formation (Avolio *et al.*, 2018). Here, we present a synonymous variant in *USP11*, which leads to a reduction of mRNA and protein expression in the studied PCD patient. Our data, together with the data from the report of Avolio and co-workers may suggest *USP11* as a novel PCD gene candidate. Further studies are needed to validate this association and understand the role of USP11 in PCD.

The CNV loss founded in *CRHR1* gene lead to a significative reduction in gene expression and, apparently, in protein expression as well. This gene was previously associated with an improve in lung function in asthmatic patients (Tantisira *et al.*, 2004), the exact mechanism by which this improvement is still unknown. If additional studies validate our proposed association of CRHR1 to PCD, it perhaps could also open a research for development of newer therapies for PCD patients as opened in the asthmatic.

The CNV gain observed in *KTR34* gene was corroborated by an increase gene expression that was also verified in patient's sister, who presents respiratory complications compatible with Bronchiectasis, but TEM analysis do not diagnosed as a PCD patient. Interestingly, in both cases KTR34 protein was not detected in cytoplasm but in cilia, suggesting that the overexpression of this gene lead to a dislocation of the protein in respiratory cells. We do not have data to justify why this occurs neither what is the functional role of this.

In conclusion, here we show three candidate genes (*USP11, CRHR1* and *KTR34*) that can possibly contribute to the PCD. None of these genes were previously reported in PCD patients and were here firstly showed its presence and its immunolocalization in respiratory cells. This study is limited to a small pedigree and thus further studies are needed to confirm this hypothesis and to understand the mechanism of action of these genes in these cells and in PCD. Another limitation that we found was due to the rarity of ciliated cells and its fragility, likely due to the age of both patient and his sister. Culture of ciliated cells, by ALI-culture or by submerged-culture of primary epithelial cells, is of upmost importance and has already proved good results in research into PCD (Jorissen *et al.*, 2000; Hirst *et al.*, 2010) particularly in this cases where the cells are rare and ask for a repetition nasal brushing should be avoid as it is traumatic for children.

Besides propose three possible candidates as novel PCD genes, more importantly this work highlights that unveiling the genetic etiology of PCD patients is not straightforward and several approaches may be needed.

3.3. Next-generation sequencing & bioinformatics

The results that will be presented in this section show preliminary work that we started in bioinformatic field. These results were already published elsewhere (Oliveira *et al.*, 2018)

3.3.1. Evaluating Runs of Homozygosity in Exome Sequencing Data - Utility in Disease Inheritance Model Selection and Variant Filtering.

Background

Monogenic Mendelian diseases are a group of clinical entities caused by genetic defects present in a single *locus*, that mostly following the inheritance laws originally proposed by Gregor Mendel(Miko, 2008). They represent an opportunity to learn about gene functions and the pathophysiological mechanisms of diseases. The disease-causing gene may be in an autosome or in sex-chromosomes and of dominant or recessive nature. Individually these diseases are rare, which makes the genotype-phenotype association more complex than initially thought. But collectively they occur at a high incidence, with an estimated 7.9 million children being born annually with a serious birth defect of genetic origin(Christianson *et al.*, 2006).

Over the last two decades, most of the Mendelian genes have been identified by ***linkage analysis studies,*** a strategy to map genes (i.e. find their location in chromosomes) that predispose to disease/traits. This method is based on the premise that the closer the two genes are on a chromosome, the lower is the likelihood of recombination between them and the more likely they are to be inherited together (Lobo and Shaw, 2008). This linkage disequilibrium (LD) is thus defined as a non-random association between alleles at two or more *loci*. Besides a physical proximity, in LD the genes may be linked due to non-physical factors. Population subdivision, inbreeding events, changes in population size and the migration of individuals between populations affect LD throughout the genome

In addition to gene mapping, another important technique in clinical genetics is Sanger Sequencing, a method of determining DNA sequence developed by Frederick Sanger and co-workers in 1975 (Sanger *et al.*, 1977). Sanger sequencing is considered the "gold standard" in genetic diseases diagnostics and also the first choice to confirm/validate sequence variants. Technological advances over the past decade led to the development of high-throughput sequencing platforms. The so-called "next generation sequencing". With this new technology, the human genome can be completely sequenced within a short time, allowing the simultaneous analysis of multiple genes and consequently the attainment of new important findings (see Mardis (2008); Boycott *et al.* (2013); Koboldt *et al.* (2013) for further reading).

Due to the extreme rarity of some Mendelian diseases, it is very difficult to find enough individuals and affected families with the same phenotype, which is usually a prerequisite for gene discovery. Population history and geographical events, such as genetic bottlenecks (i.e. a sharp reduction in the size of a population), and cultural factors such as consanguineous marriage (related as second cousins or closer), increases the probability of the offspring inheriting two deleterious copies of a recessive gene. These at-risk kindreds represent a particular context where the incidence of autosomal recessive disorders and its transmission to offspring is considerably higher (Bittles, 2001). In 2013, 44 marriages between related individuals were reported in Portugal, which represents only 0.1% of total marriages(INE, 2002). The affected offspring will have not only two identical copies of the ancestral allele, but the surrounding DNA segments will also be homozygous. Thus, the affected individual carries long stretch of DNA segments that are **identical by descent (IBD)**, i.e. homozygous segments inherited from each parent. These homozygous DNA stretches are called '**runs of homozygosity'** (ROH) (Fig. 3.3-1).

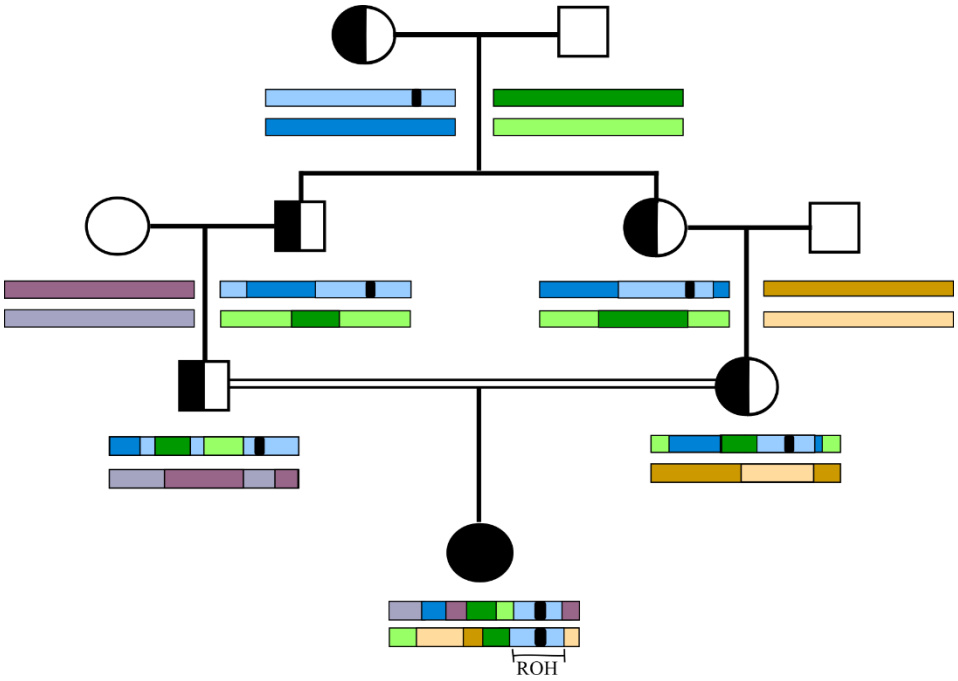


Figure 3.3-1. Schematic representation of a small consanguineous pedigree and ROH with a causative mutation (black box).

The ROH length is dependent on the degree of shared parental ancestry and on the number of generations that share the DNA region. Therefore, longer ROH (measuring tens of Mb) can be identified in individuals from geographically isolated populations or belonging to ethnical groups with a higher consanguinity rate, due to cultural or religious reasons. In contrast, in larger and older populations with reduced consanguinity rates, ROH length is generally much shorter, as homozygous stretches have been broken down over successive generations by repeated recombination events during meiosis (McQuillan *et al.*, 2008). Nevertheless, there are small regions of our genome in which the recombination rates are considerably lower and thus IBD regions are always observed.

Homozygosity mapping

Homozygosity mapping (HM, also known as autozygosity mapping), is a positional mapping strategy used for the first time by Lander and Bostein in 1987 (Lander and Botstein, 1987). This powerful technique for autosomal recessive disease mapping relies on the assumption that an affected individual inherits two identical alleles of a disease gene from a common ancestor. In cases of consanguinity or population bottlenecks, the

IBD regions have longer ROH. Accordingly, to perform HM it is necessary to search for consistently homozygous regions in affected individuals from inbred families or from geographically confined populations. As autosomal recessive diseases can be extremely heterogeneous, HM is more robust when searching for a mutation segregating within a small, closed/ consanguineous population. In these populations, each allele assumes a high frequency such that two individuals taken at random from the reproductive pool are likely to have alleles in common, hence likely to be carriers of the same disease-causing allele, and at risk of having clinically affected offspring(Goodship *et al.*, 2000; Alkuraya, 2010a). Detecting ROH requires the use of markers that both span the region in question and are sufficiently polymorphic to punctuate it as a distinct haplotype (i.e. a distinct profile for consecutive markers). Scanning for blocks of homozygosity initially resorted to microsatellite markers (stretches of short tandem repeats of 1-6 bases pairs) found throughout the human genome. However, although these are highly informative (polymorphic), PCR reactions can be difficult to optimize and prone to artifacts. Microsatellites analysis has therefore by the analysis of single nucleotide polymorphisms (SNPs) (Alkuraya, 2010b). SNPs are less informative, due to their lower heterozygosity scores but are much more abundant in the human genome (~1 per Kb) and allow the use of high-throughput assays, such array chips that allow the simultaneous calling of hundreds of thousands of SNPs (Syvänen, 2005; Gibbs and Singleton, 2006).

Homozygosity Mapping Bioinformatics

SOFTWARE FOR SNP-ARRAY DATA

Most of the disease-causing genes have been identified in the past by linkage analysis. This involves finding an association between the inheritance pattern of the phenotypic trait (usually a disease state) and a genetic marker, or a series of adjacent markers. Hence, genetic linkage analyses are useful to detect chromosomal regions containing disease genes, by examining patterns of inheritance in families (Xu *et al.*, 2012). Two main types of genetic linkage analysis are commonly used, namely model-based linkage analysis (or parametric analysis) and model-free linkage analysis (or non-parametric analysis). In parametric linkage analysis, it is assumed that models de-scribing both the trait and

marker loci are known. In contrast, in non-parametric methods some assumptions are made about the trait model, thus is more prone to errors (Bailey-Wilson, 2006). For the analysis of simple Mendelian disorders with an extended pedigree, with a well-established mode of inheritance and where the family members are un-ambiguously clinically characterized as affected or unaffected, the parametric ***logarithm of odds (LOD) score*** method is one of the most popular statistical tool. It is based on an assessment of the recombination fraction, denoted by theta (θ), which is the probability of a recombination event between the two loci of interest and a function of distance. Two unlinked loci have a theoretical θ = 0.5 (null hypothesis of no linkage) and the closer a pair of loci, the lower their recombination fraction (Pulst *et al.*, 1999).

The LOD may be expressed as follows, (L denotes likelihood):

$$LOD = \log10(L(\theta = \theta)/L(\theta = 0.5) \quad \text{Equation (1)}$$

However, for complex traits with incomplete penetrance, phenocopies, multiple trait *loci* and possibly incorrectly specified dominance relationships, a precise genetic model cannot be specified, which makes LOD score analysis inadequate. Here, non-parametric linkage (NPL) analysis are commonly used(Bailey-Wilson, 2006).

As previously mentioned, two alleles can be identical by state (IBS), if they have the same DNA sequence, or IBD. If IBS alleles are derived from the same ancestral allele. A statistical test is performed to compare the observed degree of sharing to that expected when assuming that the marker and the trait are not linked. NPL analysis often examines IBD or IBS allele sharing in cases where siblings exhibit the same trait of interest(Bailey-Wilson, 2006).

Most of the genetic linkage analyses have been performed with the help of bioinformatic tools. The first generally available computer program for linkage analysis was LIPED (Ott, 1974). It estimated the recombination fraction by calculating pedigree likelihoods for various assumed values of the recombination fraction, using only two-point parametric LOD scores and the Elston-Stewart algorithm (Elston and Stewart,

1971). Thereafter, several software solutions have been developed. Here we will briefly describe of some of the most popular software used to perform gene mapping.

GENEHUNTER, developed by Kruglyak and co-workers is among the more frequently used software for parametric multipoint analysis of pedigrees, mainly because of its user-friendly interface and rapid operation. It was the first to calculate the exact multipoint linkage scores involving many markers, using the **hidden Markov model (HMM)** (Kruglyak *et al.*, 1996). **HMM** is a probabilistic sequence model in which, given a sequence of units (e.g. DNA sequence), a probability distribution over possible sequences of labels is computed and the best label sequence is chosen (Ghahramani, 2001). This computer program is written in C language, usable in UNIX based systems through a command-line interface. It has a complete multipoint algorithm to determine the probability distribution over possible inheritance patterns at each point in the genome. The software then applies those concepts to define a unified multipoint framework for both parametric (the traditional LOD-score calculations) and non-parametric analysis (Kruglyak *et al.*, 1996). However, it has the drawback of being limited to relatively small pedigrees (approximately 12 non-founders, i.e. individuals whose parents are in the pedigree) and not be feasible for large multigenerational pedigrees. Splitting pedigrees or discarding outlier individuals (trimming) is not a good option since thus can result in a greater loss of information or invalidate the heterogeneity LOD score (Goedken *et al.*, 2000).

SIMWALK2 is another statistical tool, first described by Sobel and Lange to perform haplotype, parametric linkage, non-parametric linkage, IBD and mistyping analyses on a pedigree (Sobel and Lange, 1996). It uses Markov chain Monte Carlo: an algorithm that is used to characterize a particular distribution without the knowledge about all of the distribution's mathematical properties, just by sampling values out of the distribution at random. The Markov chain Monte Carlo method can be employed to draw samples from distributions even when all that is known about the distribution is how to calculate the density for different samples(van Ravenzwaaij *et al.*, 2016).

MERLIN is another software available for mapping genes by linkage (Abecasis *et al.*, 2002). It is written in C++ language and carries out single point and multipoint analyses

of pedigree data, including IBD calculations, non-parametric and variance component linkage analyses and information content mapping. MERLIN does not calculate parametric LOD scores, which are available in GENEHUNTER and ALLEGRO, but does this the fastest (Dudbridge, 2003). Finally, MERLIN can execute gene-dropping simulations for estimating empirical significance levels. Gene dropping simulation is a computer simulation in which founder genotypes are first simulated according to population allele frequencies. The genes are "dropped" down the pedigree by simulating gene-flow according to Mendel's laws. Finally, the phenotypes are simulated from the genotypes according to the penetrance. Simulated phenotypes are then compared to the observed, and outcomes inconsistent with the observed phenotypes are discarded. The result is a random sample of genotypes given the observed phenotypes(MacCluer *et al.*, 1986).

ALLEGRO is similar to GENEHUNTER; it also uses an HMM to calculate multipoint parametric LOD scores, NPL scores and allele-sharing LOD scores, reconstruction of haplotypes, estimated recombination count between markers, and entropy information, but it is about 20-100 times faster (Gudbjartsson *et al.*, 2000). In contrast to GENEHUNTER, for NPL analysis (which gives a p-value computed by comparing the observed NPL score with its complete data distribution), besides providing this value, ALLEGRO calculates an additional p-value by comparing the observed allele-sharing LOD score with its complete data distribution. This value increases the accuracy of the results.

SOFTWARE FOR WES DATA

Recent studies have clearly demonstrated the power and the effectiveness of applying HM to WES data, to identify causative genes for Mendelian disorders (Alkuraya, 2013). Krawitz and co-workers in 2010, developed a statistical model that allowed to infer IBD regions from the exome sequencing data of an affected child of a non-consanguineous couple, in which the disease followed the autosomal recessive inheritance mode (Krawitz *et al.*, 2010). In non-consanguineous families, the affected children do not share two equal haplotypes inherited from a single common ancestor, at the disease *locus*, but

inherit identical maternal and paternal haplotypes in a region surrounding the disease gene, which is not necessarily from the same ancestor (IBD = 2). They used an algorithm based on an HMM, to identify chromosomal regions with IBD = 2 in the presence of sequencing data, to find the causative gene (Krawitz *et al.*, 2010) Another pioneering work from Becker et al, studied four patients from consanguineous families, extracting genotypes of all dbSNP130-annotated SNPs from the exome sequencing data, and using these 299,494 genotypes as markers for genome-wide identification of homozygous regions. From the analysis of these regions, the authors were able to identify a single homozygous truncating pathogenic variant(Becker *et al.*, 2011).

The combination of exome sequencing with HM is so powerful to unveiling the cause of autosomal-recessive disorders, special in consanguineous families, that finding the mutated gene might only require a single affected individual. The HM allows narrowing down the target data sets; examination at the base-pair level then enables identification of candidate causative variants. This approach would start by mapping reads from exome sequencing against a reference genome (human hg19 in our examples). Data from whole-genome and RNA sequencing approaches can also be used. This step is usually performed through a bioinformatic pipeline that generates SAM/BAM (Sequence /Binary Alignment Map) and, at the end, VCF files. These files are then used as input for the HM analysis tools. The position and zygosity of resulted sequence variants can be used to retrieve/infer ROH regions. Longer ROH are indicative of homozygous regions. Table 3.3-1 lists some of the available tools to perform this analysis, supported by different algorithms.

The web-based tool HomozygosityMapper allows users to interactively analyses NGS data for HM and is freely available at http://www.homozygositymapper.org/ (Seelow *et al.*, 2009; Seelow and Schuelke, 2012). It functions entirely on web-based software using the HTML interface, thus is user-friendly and does not require any local installation or specific data format (Seelow *et al.*, 2009). It is independent of parameters such as family structure or allelic frequencies. The algorithm slides along an array of markers/SNPs (genotypes) inspecting zygosity and filling ***block length array***. If the SNP is **heterozygous the block length is 0**, whereas if the SNO is homozygous it will

determine the size of the homozygous block using the function *DetectBlockEnd* or until reaching the end of genotypes array. Through the use of this function, the algorithm determines the end of each homozygous block, ignoring single heterozygous genotypes with seven or more homozygous/unknown genotypes on either side. HomozygosityMapper calculates a "homozygosity score", and it is very robust against genotyping errors due to the permissivity of the *DetectBlockEnd* function as mentioned above. HomozygosityMapper is not intended to replace other linkage tools, such a GENEHUNTER, but, as it is much faster it may be used in combination therewith. The software can rapidly identify the possible disease regions and then subsequently generate LOD scores and haplotypes with conventional software. HomozygosityMapper is link with GeneDistiller (Seelow *et al.*, 2008), which is a database that includes information from various data sources such as gene-phenotype associations, gene expression pattern, and protein-protein interactions, allowing researchers to easily search information for the genes within a candidate interval, for instance with a high homozygosity score (Seelow *et al.*, 2009; Seelow and Schuelke, 2012). Therefore, the **HomozygosityMapper** can be used as an indicator of the type of inheritance pattern and/or be used as a filter for the analysis of NGS data.

Table 3.3-1. Main algorithms and bioinformatics tools used for ROH detection

Algorithm	Software [Ref.]	UI	OS	Input data files	Range ROH size (Mb)
Sliding blocks	Homozygosity-Mapper Seelow *et al.* (2009); Seelow and Schuelke (2012)	GUI	Unix/Linux web server [a]	VCF files + SNP genotypes	> 1.5
Sliding-window	PLINK Purcell *et al.* (2007)	CLI & GUI	Unix/Linux, Mac OS, Windows	BED, PED, and FAM files	[0.5-1.5]; >1.5
	GERMLINE Gusev *et al.* (2009)	CLI	Unix/Linux	PED, MAP and Hapmap files	> 1.5
	HomSI Görmez *et al.* (2013)	GUI	Unix/Linux, Mac OS, Windows	VCF files	> 1.5
Heterogeneous hidden Markov model	H^3M^2 Magi *et al.* (2014)	CLI	Unix/Linux	BAF profiles	<0.5; [0.5-1.5]; >1.5
Frequentistic genotype assigment	Agile- Genotyper and VariantMapper Carr *et al.* (2013)	GUI	Windows	SAM + tab-delimited text files	> 1.5

[a]- http://www.homozygositymapper.org; BAM- Binary Alignment Map; BED- Browser Extensible Data; CLI- command-line interface; GUI- graphical user interface; OS- operative system; Ref.- references; SAM- Sequence Alignment Map; SNP- Single Nucleotide Polymorphism; UI- user interface; VCF- Variant Call Format.

PLINK (Purcell *et al.*, 2007), GERMLINE (Gusev *et al.*, 2009) and EXome-HOMozygosity (Pippucci *et al.*, 2011) are examples in which a sliding-window algorithm is applied for WES-based ROH detection. In a sliding window analysis, the statistics are calculated for a small frame of the data. The window incrementally advances across the region of interest and, at each new position, the reported statistics are calculated. In this way, chromosomes are scanned by moving a window of a fixed size along their entire length and variation in genetic markers across the region of interest can be measured. In practical terms, a sliding window is a sub-list that runs over an underlying collection of data (Fig. 3.3-2). This type of analysis allows investigation into the way patterns of variation change across a surveyed genomic segment (Tang *et al.*, 2009).

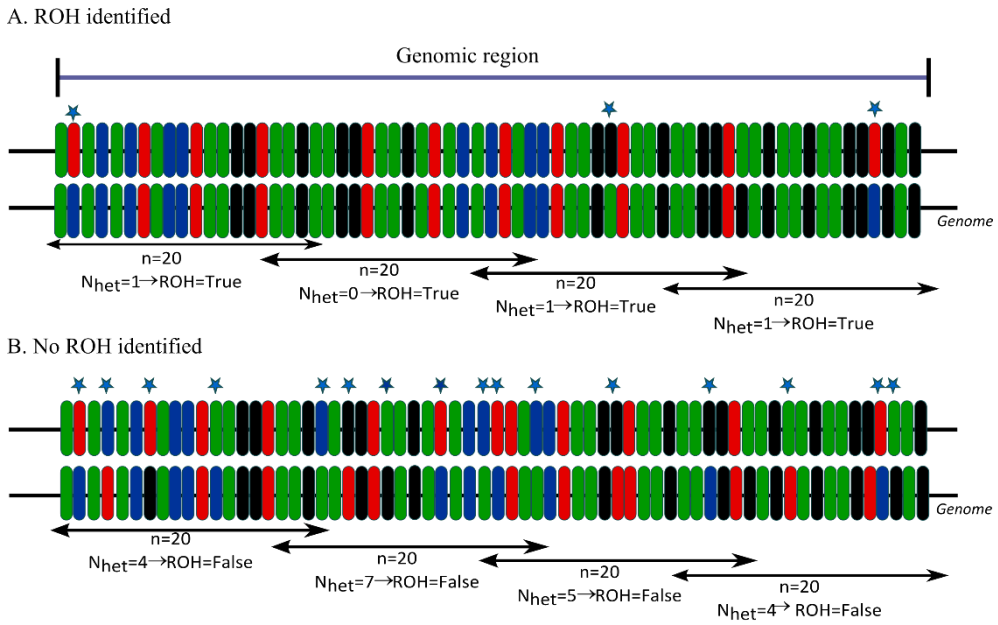


Figure 3.3-2. Schematic representation of the sliding window algorithm. A. Representation of genomic region that includes an ROH. B. Representation of a genomic segment with no ROH. Stares represent heterozygous SNP.

In summary, all the presented methods have proven to be extremely valuable for the identification of ROH from WES data, with the possibility of performing HM mapping without resorting to SNP arrays. Still, although powerful, there is an error rate associated with this strategy that is difficult to estimate, since it is highly dependent on the exome metrics, sequencing platform and the bioinformatic tool used to infer ROH. Moreover, they have to take on the technical limitations of WES, namely the inadequate coverage of some exonic regions or sequencing errors which may be a source of false positive

and/or false negative calls. Another issue is related with the incomplete annotation of the human genome, which can affect the accuracy of the mapping and annotation of variants; for instance, a deep intronic (pathogenic) variant, could indeed be the causative mutation in a non-annotated exon. With the improvement of NGS technologies, these limitations will be surpassed, and this method will be even more powerful.

ROH quantification in WES to support variant filtering and disease inheritance model selection

Previously we presented HM data of two patients with distinct recessive conditions: a congenital neuromuscular disease and PCD, caused by homozygous pathogenic variants identified by WES (Oliveira *et al.*, 2017). Although, only in one of the cases the disease-causing variant was located in a recognizable ROH, in both the overall HM data was compatible/suggestive with that of a homozygous autosomal recessive (AR) disease model.

With genetic conditions considered to be rare conditions, the patients are often found to represent sporadic, rather than familial cases. This means that, in particular kindred, no other individuals with the same phenotype are found. In this context, variant filtering of WES data is more difficult, as most disease models are still applicable. This scenario is even more complex when the WES is restricted to the patient (singleton), as is the case of adult patients where parental samples are not available for trio analysis. We postulate that ROH quantification could assist the choice of the suitable disease inheritance model. More specifically, if the ROH regions are above a predetermined threshold value this would indicate that the homozygous AR model could contribute to identify the disease-causing variant. A lower ROH overall score, on the other hand, would indicate that this model is unsuitable for variant filtering (Fig. 3.3-3).

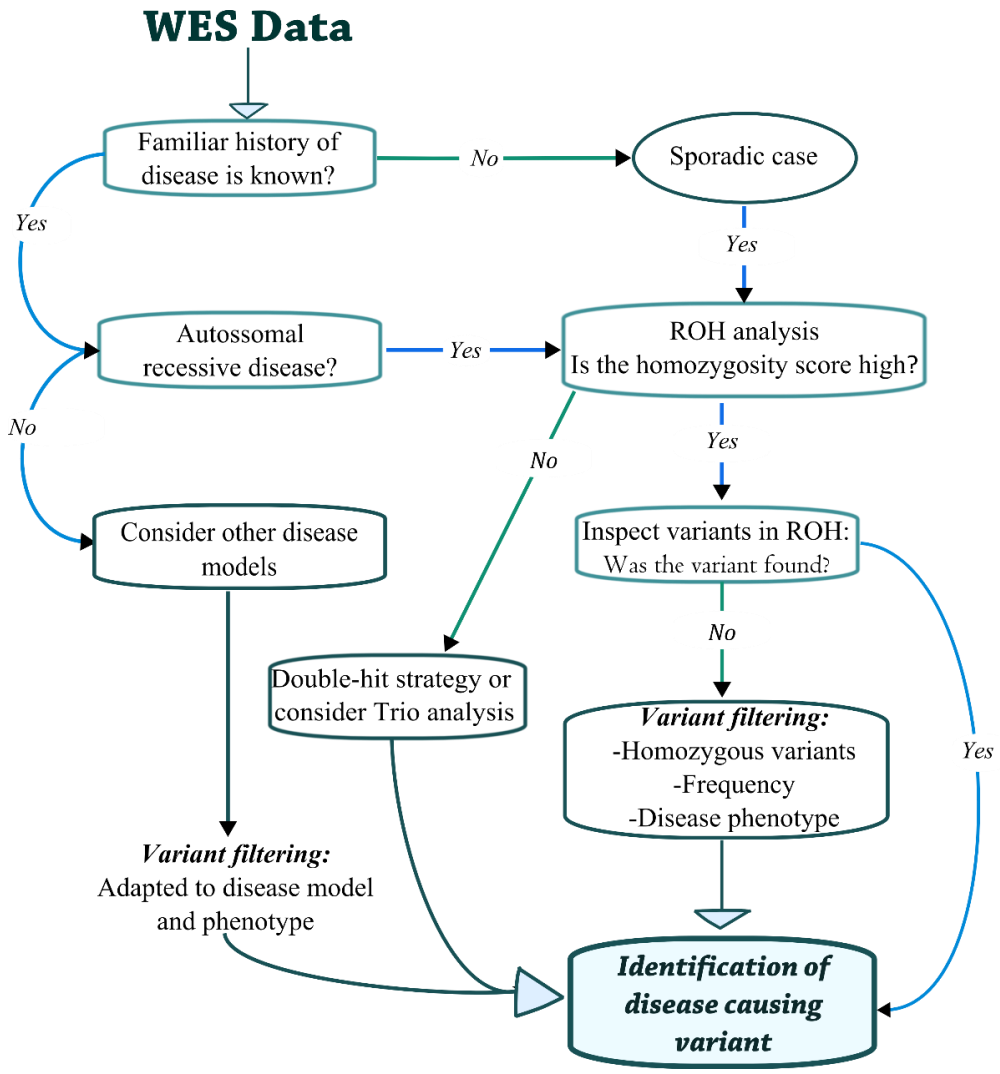


Figure 3.3-3. Proposed workflow to identify disease-causing variants from WES data, with and without family history

As a preliminary study, we analyzed ROH in WES data of twelve unrelated patients. In six cases the data was compatible with an ***autosomal recessive (AR)*** disease model and further six cases whose disease model (and disease-causing variants) are ***still unknown (UKN)***, used as test group. Among the AR disease model, we further subdivide the patients in two groups: the first where the disease-causing variants were in homozygous state, thus we called this group "**homozygous-AR"** (n=3, example in Fig. 3.3-4A); and the other group where the disease-causing variant was found in com-pound heterozygosity, known as "**heterozygous AR"** (n=3). To systemize this analysis, ROH were determined using HomozygosityMapper. In this software, we varied the block length (40, 60, 80, 100, 150, 200 and 250) and collected several parameters (highest homozygosity score, number of ROH, total size of ROH, average size of ROH, and the

size of the largest ROH). With the generated data we could then evaluate how each parameter changed ac-cording to block length.

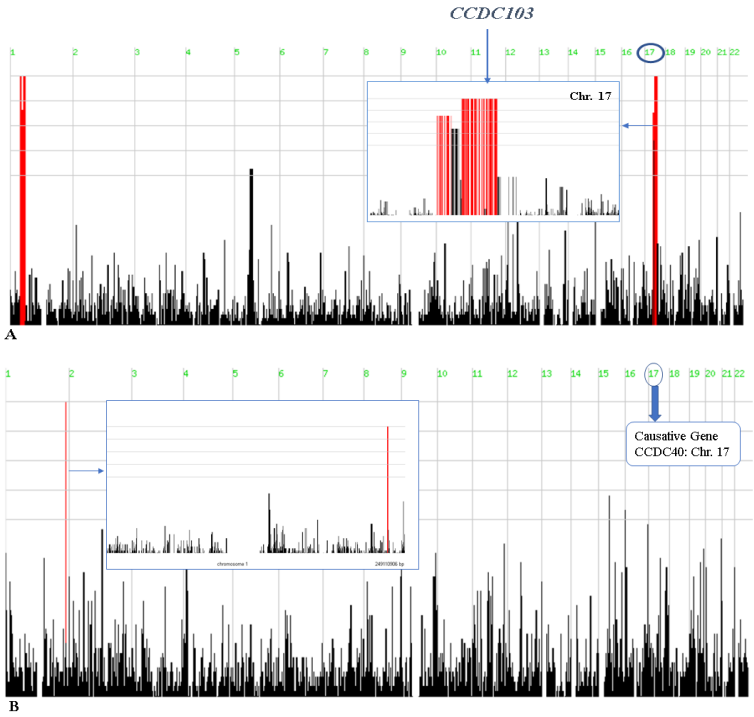


Figure 3.3-4. Genome-wide homozygosity mapping using WES and the HomozygosityMapper software. The longest ROH are highlighted in red. **A**. An example of a homozygous autosomal recessive case is shown, as previously presented in(Oliveira *et al.*, 2017). The patient has two long stretches of homozygous SNPs in chromosomes 1 and 17. The disease-causing gene is included in the second ROH (inserted blue box). **B**. An example of a compound heterozygous autosomal recessive, longer stretches of homozygous SNP were not identified, just a small homozygous region in chromosome 1 is presented (inserted blue box).

A clear cut-off between the two groups (hom-AR and het-AR) for homozygosity score was identified (Fig. 3.3-5). In addition, for cases with longer ROH the block length of 80 was not enough and should be optimized (increased) to determine the highest score accurately (up to a value of 250 in the samples tested). It was interesting to note that scores above 60 were only detected in homozygous-AR cases (Fig.3.3-5). We found statistically significant differences between the two groups using Kruskal–Wallis test (p=0.0146). Further, we applied a Mann-Whitney U test, between hom-AR and het-AR groups which was found to be statistically significant, for the following parameters: highest score (Z = -3.302, p =0.001), ROH average size (Z = -5.211, p =0.000) and ROH largest size (Z=-3,429, p=0,001). To further verify if the differences empirically observed in our cases could be statistically supported, we correlated the highest homozygosity

score obtained in HomozygosityMapper with the total size in Mb of the respective ROH for each case. Next, we applied k-means clustering algorithm, a method of cluster analysis that allows the partitation of n observations into k clusters, in which each observation is part of the cluster with the nearest mean.

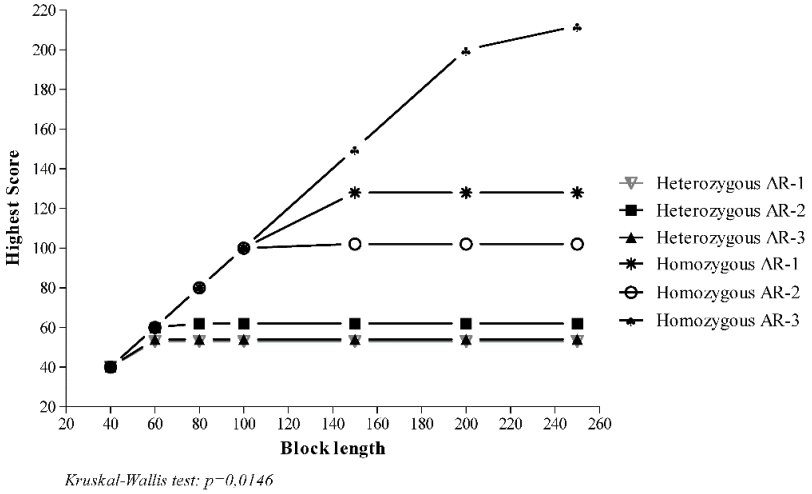


Figure 3.3-5. Correlation between highest homozygosity score and block length and comparison between heterozygous and homozygous autosomal recessive (AR) groups. Statistical analysis carried out with IBM SPSS statistics v.24; graphic in GraphPad Prism v.5. There is a clear cut-off between the two groups.

Two different clusters were identified (Fig.3.3-6). The first cluster corresponds to the heterozygous AR samples (k1) whose values seem to be more homogeneous compared with second cluster (k2, homozygous-AR) which has more dispersed data points. Kruskal-Wallis test demonstrated statistically significant differences between the two study groups ($\chi^{2(2)} = 6.2$ and $p = 0.044$). No statistically significant differences were found between the UKN group and the other two (homozygous- and heterozygous AR), showing that it does not form a specific cluster. In the test group (unknown in Fig.3.3-6), four samples are located in the vicinity of cluster k1 whereas the other two are nearer k2.

It should be noted that the number of cases analyzed are still very low, and these figures should be increased at least at 50 to 100 fold so that we can draw further conclusions. Nonetheless, we were able to show that, by computing the homozygosity highest score and ROH total size, it is possible to infer whether a homozygous or a heterozygous variant would be the cause of a particular disease.

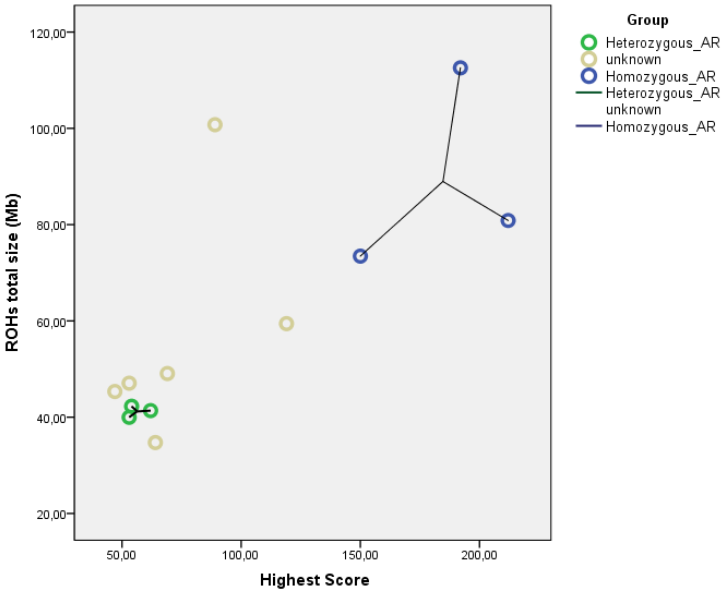


Figure 3.3-6. Scatter plot presenting data analysis resorting to k-means clustering algorithm. Two clusters were identified for two groups: heterozygous AR (k1, green) and homozygous AR (k2, blue). Data analysis and plot performed with IBM SPSS statistics v.24.

The existence of longer stretches of homozygous SNP (ROH) may indeed provide a dual role: i) indicate the pathogenic variant zygosity status (homozygous or heterozygous), and ii) point towards the location of the disease-causing gene, since is likely to be located within these longer homozygous stretches. For instance, the case presented in Fig. 3.3-3A, represents a patient diagnosed with PCD. This case exemplifies how two longer ROH were identified, one in chromosome 1 and other on the 17, with the causative variant ultimately being found in the gene *CCDC103*, located in chromosome 17 (Oliveira *et al.*, 2017).

Overall, this work reviewed HM as an approach for gene discovery and the identification of disease-causing variants. The application of HM should be considered, not only in research but also in routine clinical genetics setting, in cases where an autosomal recessive disease is suspected. The currently available algorithms and bioinformatic tools designed for ROH detection from array-SNP and WES data were also reviewed, highlighting their main features, advantages and limitations. Selection of appropriate algorithms should mainly consider the technology used to generate the data, but also specific features of the patient/ family under study, such as genetic context, ROH average size and the number of relatives affected by the same condition. As depicted in Fig. 3.3-3, ROH analysis should be an option to include in the variant filtering strategies, in cases with familial AR transmission pattern or in sporadic cases.

As for identifying new associations between a gene and a particular phenotype, the presence of consanguinity in specific kindred has its particularities; not only is the disease inheritance model self-evident (even in sporadic cases), but also concentrating the analysis of variants located in longer ROH regions reduces considerably the analytical burden. The term 'consanguinous' is somewhat ambiguous, as there are different levels, from the first to the 4th degree (e.g., first cousins). So, the average percentage of DNA shared ranges from 50 to 12.5%. It should be noted that these percentages may influences the number and size of ROH. However, ROH analysis should not be confined to consanguineous cases as it may be found to be useful for cases with unknown parental consanguinity, especially if parents are from the same remote/confined village, where mutation founder effects or inbreeding events might be an issue.

This work has also shown some preliminary data that show how the analysis of ROH regions may also be extremely useful for variant filtering and for the selection of a disease inheritance model. More specifically, by correlating the highest homozygosity score and total size of ROH obtained with HomozygosityMapper two distinct clusters were obtained. If our assumption is correct, in the presence of a particular subject whose WES data clusters in k1 heterozygous variants, thus should be verified first as a possible cause for the phenotype, corresponding to autosomal dominant (*de novo*) or compound heterozygous AR disease models. If the patient is a male, we cannot exclude an X-linked recessive inheritance, thus hemizygous variants should also be inspected. However, it should be noted that the number of cases analyzed is still limited, and further data needs to be collected and processed to verify these preliminary results.

In conclusion, the combination of WES and HM allow researchers to identify candidate *loci* and underlying genetic defect itself (at the nucleotide level) in a single step. Nonetheless, there are still several limitations and further bioinformatic developments are required. Considering the data presented, there are sensitivity issues that require addressing, especially if the genetic defect is located in a small ROH or if the pathogenic variant is *de novo* in an individual born to consanguineous parents. Finally, we consider

that it would be useful to develop a tool that combines variant filtering and homozygosity mapping, which currently can only be performed individually.

Chapter 4

Main conclusions and Future perspectives

4.Main conclusions and Future perspectives

PCD is a ciliopathy caused by dysfunction of motile cilia and a rare autosomal recessive disease with a reported prevalence ranging from 1:4,000 to 1:40,000 in the general population (Lucas *et al.*, 2014). As described in Introduction, motile cilia are specialized structures that occur at very specific regions, namely in the upper airways (respiratory cilia), brain (ependymal cilia), female reproductive system (Fallopian tube cilia), male reproductive system (sperm flagella) and embryo (Nodal cilia). It comprises the axoneme, composed of nine peripheral microtubule doublets (A, B) that are linked by the DRC and present DA. The peripheral microtubules are connected, by RS, to a single pair of central microtubules (the CPC). The axoneme is the major responsible for the cilia motility that is essential for various tissue-specific functions, like the mucociliary clearance and sperm flagellar motility. PCD patients can show a panoply of clinical features but the most prevalent are chronic respiratory infection due to an ineffective mucociliary clearance (Rubbo and Lucas, 2017), as well as heterotaxy syndrome, which in combination with PCD is designated as KS. Moreover, reproductive issues leading to subfertility or infertility, such as sperm immotility, are often described too (Pereira *et al.*, 2015; Pereira *et al.*, 2017).

Due to the heterogeneity of the disease, the laboratorial diagnosis of PCD cannot rely on a single test. Usually, these include the combination of nasal nitric oxide measurements, ultrastructural analysis of the ciliary axoneme, high-speed video microscopy analysis and genetic screening (Jackson *et al.*, 2016).

This Thesis highlighted the need for high-throughput techniques to unveil the genetic etiology of PCD. We have shown that the classical gene by gene approach is not cost-effective and demonstrated that WES is a valuable tool in PCD diagnosis (Chapter 3.1.1; Pereira *et al.* (2015)). We also demonstrated that the use of WES in PCD diagnosis can be a delicate and prolonged process, as each patient is unique and rarely the same gene variant(s) is harboured by distinct patients. Even for the most mutated genes in PCD patients, such as *DNAH5,* the estimated prevalence ranges from only 15 to 30 % of PCD

patients. Therefore, is not possible to have a faster screening genetic test, has occur for other genetic diseases such as cystic fibrosis, which is caused by dysfunction of a single gene (*CFTR* gene) and with two variants (p.F508del and p.G551D) accounting for the majority of the cystic fibrosis cases (Cutting, 2014). Commercial genetic panels for PCD have been proposed to overcome the WES limitations in terms of data analysis, management, and storage, but has demonstrated here, some PCD patients (about 30%) may do not present pathologic variants in the most prevalent genes thus commercial genetic panels, will give an inconclusive result.

Within this doctoral thesis, the genetic background of patients presenting KS, with the absence of axoneme DA, was studied by searching for gene variants that could justify the patient's phenotype, which was performed mainly through WES analysis. Additionally, the genotype-phenotype correlations in these patients were also accomplished.

Initially, it was presented a genotype-phenotype correlation of *CCDC103* pathogenic variants in PCD patients with *situs-inversus totalis* and absence of DA. CCDC103 is a small protein with interesting biochemical properties that was shown to be required for ODA assembly, but, it remained, until now poorly characterized. Here, the expression profiles of CCDC103 were firstly characterized in human cells, thus increasing the knowledge regarding its expression and subcellular localization. We showed that the *CCDC103* gene is expressed differently in reproductive and somatic cells, both at the protein and RNA level. This data corroborates the involvement of CCDC103 in PCD and suggests a role in infertility.

Subsequently, through the analysis of KS children's' exome, it was demonstrated how genetic diagnostic is not trivial. Firstly, by showing that even in the most prevalent PCD genes, new variants are still to be reported. Secondly, by presenting a case that evidences that WES is not a perfect solution and that finding a disease-causing variant may require the combination of multiple approaches and tools. By combining WES and CNV analysis, here was proposed potential novel candidate genes for PCD.

Additionally, we proposed a model, which was not previously reported in PCD but was already observed in other complex diseases, where variants in different genes may

act together and contribute to the same phenotype. Particularly, we had proposed that an interaction between *DNAH5* and *DNAH7* gene variants may have led to the absence of both DA as observed in the studied child-2.

In the last part of this Thesis, preliminary data were presented for the development of a bioinformatic workflow that combines WES and HM, which aims to allow researchers to identify the disease inheritance mode, namely for sporadic cases, as well as the candidate loci and underlying genetic defect itself in a single step. Specifically, it was proposed that by the homozygosity score it is possible to infer if the variant is homozygous or not, and, if so, it is possible to directly have a hint about the disease-causing variant *locus*. Additional studies with a higher number of individuals are needed to further improve this preliminary workflow, which is a valuable tool for assisting the genetic analysis and could be applied to other diseases than PDC.

The differentiation of human induced pluripotent stem cells into the airway epithelium has recently been achieved (Wong *et al.*, 2012; Firth *et al.*, 2014; Konishi *et al.*, 2016), opening the way for advances towards modelling PCD *in vitro*. Several advances towards PCD gene therapy have also been recently described. For instance, minicircles containing the 14 kb *DNAH5* cDNA were produced and showed to provide an enhanced transgene expression in airway cells in comparison to plasmid DNA, with no toxicity differences. This is also the largest minicircle produced to date (Munye *et al.*, 2016) Efforts to restore ciliary function with lentiviral vectors presented also positive results (Ostrowski *et al.*, 2014). Therefore, we believe that the identification of genetic factors involved in the pathophysiology of PCD/KS, as well as in sperm motility will be of utmost interest in the future development of newer therapies that may arise in a couple of years and the emerge of novel forms of assisted reproductive technology. Nevertheless, having a genetic cause identified is for now of great value for personalized medicine and to assist the diagnosis, which enables the selection of the best current possible treatment.

Although several reports have been published regarding pathogenic gene variants in PCD patients, collective efforts must be implemented to uniformize the diagnostic methodology or gather PCD patients from several centres and study them together, as

differences in diagnosing could affect the genetic analysis. Moreover, having a higher number of PCD patients, which could further be divided in different groups depending on age, sex or ultrastructural ciliary defects, is needed in order to understand if is possible to obtain a genetic cluster, i.e. if possible, to associate a given gene/variant to a given age group or if differences are evident among sexes. If so, this analysis would be of upmost importance to improve the genotype-phenotype correlation and development of new diagnostic panels.

A major limitation of the works performed during this thesis was the small number of patients and, in certain cases, the unavailability of biological samples for further analysis. However, since all patients included here display KS whose incidence is estimated to be 1 in every 25,000 individuals, it is reasonable to understand the difficult that is to increase the number of studied individuals in a single thesis. In addition, as already discussed, in PCD, it is rare to find non-consanguineous individuals carrying the same genetic variant(s), which also hinders the genetic analysis and the inclusion of a higher number of individuals for genotype-phenotype correlation. To overcome this limitation, further studies with animal models are of upmost importance to validate the hypothesis proposed in this Thesis and to understand the molecular mechanism underlying the phenotype-genotype correlations observed. Particularly, genome editing studies with CRISPR-Cas9 (short for clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9), which have been shown as a valuable tool for genome editing (Komor *et al.*, 2017). For instance, in genotype-phenotype of *CCDC103* pathogenic variants, it would help to understand if differences observed in both patients were related to mutation type (missense vs frameshift) or to mutation location (n terminal region vs c-terminal region). Moreover, those models would be of upmost importance to better understand the role of *CCDC103* pathogenic variants in fertility, particularly if they affect only the male fertility but not the female fertility, as our data suggested. Nevertheless, construction of these animal models can be tricky and even of unviable realization. For instance, as part of the present Thesis, one of the initial tasks was the analysis of a zebrafish *knock-in* for the *CCDC103* pathogenic variant observed in the male patient 1, however, we wait for 2 years to develop a Zebra-fish mutant. At the final time, our American collaborators gave up trying to get any plausible result.

The comprehensive work performed in PCD has allowed the identification of several genes and the main pathophysiological mechanisms underlying this condition. However, this puzzle is still unfinished as about 30% of PCD patients remain with an inconclusive genetic diagnosis. Moreover, the knowledge about the complex cellular modifications in terms of proteomics and metabolomics is still scarce. Given the complexity of cilia and of the flagellum structure, to have a full picture of the molecular components and mechanisms that are subjacent to the formation, assemble and function of motile cilia and flagella, an integrative approach is necessary. Metabolites are definable as the intermediate end product of metabolism. Ultimately, they provide an easier way to assess the cellular state. As for metabolomics, it has been evolving as a powerful tool to identify biomarkers (Clish, 2015). Promising results in various types of airway diseases and in human fertility were achieved (Monteiro *et al.*, 2013; Nobakht *et al.*, 2014; Clish, 2015; Muhlebach and Sha, 2015; Maniscalco and Motta, 2017). Yet, as for PCD, only limited data have been reported (Paff *et al.*, 2013; Joensen *et al.*, 2014; Montuschi *et al.*, 2014). Additionally, the proteome has also been shown valuable in the identification of novel cilia proteins (Ostrowski *et al.*, 2002; Blackburn *et al.*, 2017). In the future, the proteomic and metabolomic signature of PCD patients should be performed and combined with the genetic analysis. This combination of omics (genomic, transcriptomic, proteomic and metabolomic) would be critical to the identification of novel biomarkers and the development of better diagnostics and/or therapies.

Overall, the findings from this Thesis may have an impact on the analysis of genetic data in PCD and thus in the genetic counselling of patients. Moreover, the results here presented aimed to open doors for future research, as they added an additional piece of information to the unfinished puzzle of understanding the genetic bases of PCD/KS with DA absence and ultimately may help to further understand the PCD pathophysiology.

Chapter 5

References

5. References

Abecasis, G.R., Cherny, S.S., Cookson, W.O., Cardon, L.R., 2002. **Merlin—rapid analysis of dense genetic maps using sparse gene flow trees**. *Nature Genetics* 30, 97-101. https://doi.org/10.1038/ng786

Afzelius, B.A., 1976. **A human syndrome caused by immotile cilia**. *Science* 193, 317-319. https://doi.org/10.1126/science.1084576

Afzelius, B.A., Dallai, R., Lanzavecchia, S., Bellon, P.L., 1995. **Flagellar structure in normal human spermatozoa and in spermatozoa that lack dynein arms**. *Tissue&Cell* 27, 241-247. https://doi.org/10.1016/S0040-8166(95)80044-1

Afzelius, B.A., Eliasson, R., Johnsen, O., Lindholmer, C., 1975. **Lack of dynein arms in immotile human spermatozoa.** *The Journal of cell biology* 66, 225-232. https://doi.org/10.1083/jcb.66.2.225

Afzelius, B.A., Srurgess, J.M., 1985. **The immotile-cilia syndrome: a microtubule-associated defect**. *Critical Reviews in Biochemistry and Molecular Biology* 19, 63-87. https://doi.org/10.3109/10409238509086788

Alkuraya, F.S., 2010a. **Autozygome decoded**. *Genetics in Medicine* 12, 765-771. https://doi.org/10.1097/GIM.0b013e3181fbfcc4

Alkuraya, F.S., 2010b. **Homozygosity mapping: One more tool in the clinical geneticist's toolbox**. *Genetics in Medicine* 12, 236-239. https://doi.org/10.1097/GIM.0b013e3181ceb95d

Alkuraya, F.S., 2013. **The application of next-generation sequencing in the autozygosity mapping of human recessive diseases**. *Human genetics* 132, 1197-1211. https://doi.org/10.1007/s00439-013-1344-x.

Anand-Ivell, R., Dai, Y., Ivell, R., 2013. **Neohormones as biomarkers of reproductive health**. *Fertility and sterility* 99, 1153-1160. https://doi.org/10.1016/j.fertnstert.2012.12.023

Anand-Ivell, R., Ivell, R., 2014. **Regulation of the reproductive cycle and early pregnancy by relaxin family peptides**. *Molecular and cellular endocrinology* 382, 472-479. https://doi.org/10.1016/j.mce.2013.08.010

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., *et al.*, 2014. **An atlas of active enhancers across human cell types and tissues**. *Nature* 507, 455. http://doi.org/10.1038/nature12787

Anna, A., Monika, G., 2018. **Splicing mutations in human genetic disorders: examples, detection, and confirmation**. *Journal of applied genetics* 59, 253-268. http://doi.org/10.1007/s13353-018-0444-7

Antonarakis, S.E., Krawczak, M., Cooper, D.N., 2000. **Disease-causing mutations in the human genome.** *European journal of pediatrics* 159 Suppl S173-178.

Antony, D., Becker-Heck, A., Zariwala, M.A., Schmidts, M., Onoufriadis, A., *et al.*, 2013. **Mutations in CCDC39 and CCDC40 are the Major Cause of Primary Ciliary Dyskinesia with Axonemal Disorganization and Absent Inner Dynein Arms**. *Human Mutation* 34, 462-472. https://doi.org/10.1002/humu.22261

Asai, D.J., Koonce, M.P., 2001. **The dynein heavy chain: structure, mechanics and evolution**. *Trends in Cell Biology* 11, 196-202. http://dx.doi.org/10.1016/S0962-8924(01)01970-5

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., *et al.*, 2000. **Gene Ontology: tool for the unification of biology**. *Nat Genet* 25, 25-29. http://doi.org/10.1038/75556

Austin-Tse, C., Halbritter, J., Zariwala, M.A., Gilberti, R.M., Gee, H.Y., *et al.*, 2013. **Zebrafish Ciliopathy Screen Plus Human Mutational Analysis Identifies< i> C21orf59</i> and< i> CCDC65</i> Defects as Causing Primary Ciliary Dyskinesia**. *The American Journal of Human Genetics* 93, 672-686.

Avidan, N., Tamary, H., Dgany, O., Cattan, D., Pariente, A., *et al.*, 2003. **CATSPER2 , a human autosomal nonsyndromic male infertility gene**. 497-502. http://doi.org/10.1038/sj.ejhg.5200991

Avolio, R., Järvelin, A.I., Mohammed, S., Agliarulo, I., Condelli, V., *et al.*, 2018. **Protein Syndesmos is a novel RNA-binding protein that regulates primary cilia formation**. *Nucleic Acids Research* 46, 12067-12086. https://doi.org/10.1093/nar/gky873

Baccetti, B., Collodel, G., Estenoz, M., Manca, D., Moretti, E., *et al.*, 2005a. **Gene deletions in an infertile man with sperm fibrous sheath dysplasia.** *Human reproduction (Oxford, England)* 20, 2790-2794. https://doi.org/10.1093/humrep/dei126

Baccetti, B., Collodel, G., Gambera, L., Moretti, E., Serafini, F., *et al.*, 2005b. **Fluorescence in situ hybridization and molecular studies in infertile men with dysplasia of the fibrous sheath.** *Fertility and sterility* 84, 123-129. https://doi.org/10.1016/j.fertnstert.2005.01.128

Bailey-Wilson, J.E., 2006. Parametric and Nonparametric Linkage Analysis, Encyclopedia of Life Sciences. John Wiley & Sons, Ltd, Chichester, UK.

Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., *et al.*, 2011. **Exome sequencing as a tool for Mendelian disease gene discovery**. *Nature Reviews Genetics* 12, 745-755. https://doi.org/10.1038/nrg3031

Barbato, A., Frischer, T., Kuehni, C.E., Snijders, D., Azevedo, I., *et al.*, 2009. **Primary ciliary dyskinesia: a consensus statement on diagnostic and treatment approaches in children**. *European Respiratory Journal* 34, 1264-1276. https://doi.org/10.1183/09031936.00176608

Bartoloni, L., Blouin, J.-l., Pan, Y., Gehrig, C., Maiti, A.K., *et al.*, 2002. **Mutations in the DNAH11 (axonemal heavy chain dynein type 11) gene cause one form of situs inversus totalis and most likely primary ciliary dyskinesia**. 11.

Bayly, P.V., Lewis, B.L., Ranz, E.C., Okamoto, R.J., Pless, R.B., *et al.*, 2011. **Propulsive forces on the flagellum during locomotion of Chlamydomonas reinhardtii**. *Biophysical journal* 100, 2716-2725. http://doi.org/10.1016/j.bpj.2011.05.001

Becker-Heck, A., Zohn, I.E., Okabe, N., Pollock, A., Lenhart, K.B., *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. https://doi.org/10.1038/ng.727

Becker, J., Semler, O., Gilissen, C., Li, Y., Bolz, Hanno J., *et al.*, 2011. **Exome Sequencing Identifies Truncating Mutations in Human SERPINF1 in Autosomal-Recessive Osteogenesis Imperfecta**. *The American Journal of Human Genetics* 88, 362-371. https://doi.org/10.1016/j.ajhg.2011.01.015

Behan, L., Dimitrov, B.D., Kuehni, C.E., Hogg, C., Carroll, M., *et al.*, 2016. **PICADAR: a diagnostic predictive tool for primary ciliary dyskinesia**. *European Respiratory Journal* 53, ERJ-01551-02015. http://doi.org/10.1183/13993003.01551-2015

Bekker, J.M., Colantonio, J.R., Stephens, A.D., Clarke, W.T., King, S.J., *et al.*, 2007. **Direct interaction of Gas11 with microtubules: implications for the dynein regulatory complex**. *Cell motility and the cytoskeleton* 64, 461-473. https://doi.org/10.1002/cm.20196

Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q.B., *et al.*, 2016. **Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants**. *Human genetics* 135, 359-362. https://doi.org/10.1073/pnas.1418631112

Berg, J.S., Evans, J.P., Leigh, M.W., Omran, H., Bizon, C., *et al.*, 2011. **Next generation massively parallel sequencing of targeted exomes to identify genetic mutations in primary ciliary dyskinesia: implications for application to clinical testing**. *Genetics in Medicine* 13, 218-229. https://doi.org/10.1097/GIM.0b013e318203cff2

Bernardino, R.L., Alves, M.G., Oliveira, P.F., 2018. **Establishment of Primary Culture of Sertoli Cells**, in: Alves, M.G., Oliveira, P.F. (Eds.), *Sertoli Cells*. Springer, Humana Press, New York, NY, pp. 1-8.https://doi.org/10.1007/978-1-4939-7698-0_1

Bettencourt-Dias, M., Glover, D.M., 2007. **Centrosome biogenesis and function: centrosomics brings new understanding**. *Nature Reviews Molecular Cell Biology* 8, 451-463. https://doi.org/10.1038/nrm2180

Betts, M.J., Russell, R.B., 2003. **Amino Acid Properties and Consequences of Substitutions**, in: Barnes, M.R., Gray, I.C. (Eds.), *Bioinformatics for Geneticists*. John Wiley & Sons, Ltd, New York, pp. 289-316.https://doi.org/10.1002/0470867302.ch14

Biesecker, L.G., Green, R.C., 2014. **Diagnostic clinical genome and exome sequencing**. *New England Journal of Medicine* 370, 2418-2425. https://doi.org/10.1056/NEJMra1312543

Bittles, A.H., 2001. **Consanguinity and its relevance to clinical genetics**. *Clinical genetics* 60, 89-98. https://doi.org/10.1034/j.1399-0004.2001.600201.x

Blackburn, K., Bustamante-Marin, X., Yin, W., Goshe, M.B., Ostrowski, L.E., 2017. **Quantitative Proteomic Analysis of Human Airway Cilia Identifies Previously Uncharacterized Proteins of High Abundance**. *Journal of Proteome Research* 16, 1579-1592. https://doi.org/10.1021/acs.jproteome.6b00972

Blanchon, S., Legendre, M., Copin, B., Duquesnoy, P., Montantin, G., *et al.*, 2012. **Delineation of CCDC39/CCDC40 mutation spectrum and associated phenotypes in primary ciliary dyskinesia**. *Journal of medical genetics* 49, 410-416. 10.1136/jmedgenet-2012-100867

Boldt, K., Van Reeuwijk, J., Lu, Q., Koutroumpas, K., Nguyen, T.-M.T., *et al.*, 2016. **An organelle-specific protein landscape identifies novel diseases and molecular mechanisms**. *Nature communications* 7, 11491. https://doi.org/10.1038/ncomms11491

Boon, M., Jorissen, M., Proesmans, M., Boeck, K., 2013. **Primary ciliary dyskinesia, an orphan disease**. *European Journal of Pediatrics* 172, 151-162. https://doi.org/10.1007/s00431-012-1785-6

Boycott, K.M., Vanstone, M.R., Bulman, D.E., MacKenzie, A.E., 2013. **Rare-disease genetics in the era of next-generation sequencing: discovery to translation**. *Nature Review Genetics* 14, 681-691. https://doi.org/10.1038/nrg3555

Burgess, S.A., Walker, M.L., Sakakibara, H., Knight, P.J., Oiwa, K., 2003. **Dynein structure and power stroke**. *Nature* 421, 715-718. https://doi.org/10.1038/nature01377

Burnicka-Turek, O., Shirneshan, K., Paprotta, I., Grzmil, P., Meinhardt, A., *et al.*, 2009. **Inactivation of insulin-like factor 6 disrupts the progression of spermatogenesis at late meiotic prophase**. *Endocrinology* 150, 4348-4357. https://doi.org/10.1210/en.2009-0201

Bustamante-Marin, X.M., Ostrowski, L.E., 2017. **Cilia and Mucociliary Clearance**. *Cold Spring Harbor perspectives in biology* 9, a028241. https://doi.org/10.1101/cshperspect.a028241

Bustamante, J., Arias, A.A., Vogt, G., Picard, C., Galicia, L.B., *et al.*, 2011. **Germline CYBB mutations that selectively affect macrophages in kindreds with X-linked predisposition to tuberculous mycobacterial disease**. *Nature Immunology* 12, 213. https://doi.org/10.1038/ni.1992

Bylander, A., Nutu, M., Wellander, R., Goksör, M., Billig, H., *et al.*, 2010. **Rapid effects of progesterone on ciliary beat frequency in the mouse fallopian tube**. *Reproductive biology and endocrinology : RB&E* 8, 48-48. http://doi.org/10.1186/1477-7827-8-48

Carr, I.M., Bhaskar, S., O' Sullivan, J., Aldahmesh, M.A., Shamseldin, H.E., *et al.*, 2013. **Autozygosity Mapping with Exome Sequence Data**. *Human Mutation* 34, 50-56. https://doi.org/10.1002/humu.22220

Cartegni, L., Chew, S.L., Krainer, A.R., 2002. **Listening to silence and understanding nonsense: exonic mutations that affect splicing.** *Nature Reviews Genetics* 3, 285-298. http://doi.org/10.1038/nrg775

Cartwright Julyan, H.E., Piro, N., Piro, O., Tuval, I., 2007. **Embryonic nodal flow and the dynamics of nodal vesicular parcels**. *Journal of The Royal Society Interface* 4, 49-56. http://doi.org/10.1098/rsif.2006.0155

Casey, D.M., Inaba, K., Pazour, G.J., Takada, S., Wakabayashi, K.-i., *et al.*, 2003. **DC3, the 21-kDa Subunit of the Outer Dynein Arm-Docking Complex (ODA-DC), Is a Novel EF-Hand Protein Important for Assembly of Both the Outer Arm and the ODA-DC**. *Molecular Biology of the Cell* 14 3650-3663. https://doi.org/10.1091/mbc.E03-01-0057

Casey, J.P., Goggin, P., McDaid, J., White, M., Ennis, S., *et al.*, 2015. **A case report of primary ciliary dyskinesia, laterality defects and developmental delay caused by the co-existence of a single gene and chromosome disorder**. *BMC medical genetics* 16, 45. https://doi.org/10.1186/s12881-015-0192-z

Caspary, T., Larkins, C.E., Anderson, K.V., 2007. **The Graded Response to Sonic Hedgehog Depends on Cilia Architecture**. *Developmental Cell* 12, 767-778. http://doi.org/10.1016/j.devcel.2007.03.004

Castel, S.E., Cervera, A., Mohammadi, P., Aguet, F., Reverter, F., *et al.*, 2018. **Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk**. *Nature genetics* 50, 1327. https://doi.org/10.1038/s41588-018-0192-y

Castleman, V.H., Romio, L., Chodhari, R., Hirst, R.A., Castro, S.C.P.D., *et al.*, 2009. **Mutations in Radial Spoke Head Protein Genes RSPH9 and RSPH4A Cause Primary Ciliary Dyskinesia with Central-Microtubular-Pair Abnormalities**. *The American Journal of Human Genetics* 84, 197-209. https://doi.org/10.1016/j.ajhg.2009.01.011

Chemes, H.E., 2000. **Phenotypes of Sperm Pathology : Genetic and Acquired Forms in Infertile Men**. *Journal of andrology* 21, 799-808. https://doi.org/10.1002/j.1939-4640.2000.tb03409.x

Chemes, H.E., Olmedo, S.B., Carrere, C., Oses, R., Carizza, C., *et al.*, 1998. **Ultrastructural pathology of the sperm flagellum : association between flagellar pathology and fertility prognosis in severely asthenozoospermic men**. *Human reproduction* 13, 2521-2526. https://doi.org/10.1093/humrep/13.9.2521

Chemes, H.E., Rawe, V.Y., 2003. **Sperm pathology: a step beyond descriptive morphology. Origin, characterization and fertility potential of abnormal sperm phenotypes in infertile men**. *Human Reproduction Update* 9, 405-428. https://doi.org/10.1093/humupd/dmg034

Chen, G.-W., Luo, X., Liu, Y.-L., Jiang, Q., Qian, X.-M., *et al.*, 2011. **R171H missense mutation of< i> INSL6</i> in a patient with spermatogenic failure**. *European journal of medical genetics* 54, e455-e457.

Cheng, J., Randall, A., Baldi, P., 2005. **Prediction of protein stability changes for single-site mutations using support vector machines**. *Proteins: Structure, Function, and Bioinformatics* 62, 1125-1132. 10.1002/prot.20810

Chien, A., Shih, S.M., Bower, R., Tritschler, D., Porter, M.E., *et al.*, 2017. **Dynamics of the IFT machinery at the ciliary tip**. *eLife* 6, e28606. https://doi.org/10.7554/eLife.28606.001

Chilvers, M.A., O'Callaghan, C., 2000. **Analysis of ciliary beat pattern and beat frequency using digital high speed imaging: comparison with the photomultiplier and photodiode methods**. *Thorax* 55, 314-317. http://doi.org/10.1136/thorax.55.4.314

Chilvers, M.A., Rutman, A., O'Callaghan, C., 2003. **Functional analysis of cilia and ciliated epithelial ultrastructure in healthy children and young adults**. *Thorax* 58, 333-338. http://doi.org/10.1136/thorax.58.4.333

Choksi, S.P., Lauter, G., Swoboda, P., Roy, S., 2014. **Switching on cilia: transcriptional networks regulating ciliogenesis**. *Development* 141, 1427 - 1441. https://doi.org/10.1242/dev.074666

Christianson, A., Howson, C.P., Modell, B., 2006. Global report on birth defects: the hidden toll of dying and disabled children, New York.

Clark, H., Clark, L.S., 2005. **The genetics of neonatal respiratory disease**. *Seminars in Fetal and Neonatal Medicine* 10, 271-282. https://doi.org/10.1016/j.siny.2005.02.004

Clish, C.B., 2015. **Metabolomics: an emerging but powerful tool for precision medicine**. *Molecular Case Studies* 1, a000588. https://doi.org/10.1101/mcs.a000588

Colantonio, J.R., Vermot, J., Wu, D., Langenbacher, A.D., Fraser, S., *et al.*, 2008. **The dynein regulatory complex is required for ciliary motility and otolith biogenesis in the inner ear**. *Nature* 457, 205-209.

Cole, D.G., Diener, D.R., Himelblau, A.L., Beech, P.L., Fuster, J.C., *et al.*, 1998. **Chlamydomonas Kinesin-II–dependent Intraflagellar Transport (IFT): IFT Particles Contain Proteins Required for Ciliary Assembly in Caenorhabditis elegans Sensory Neurons**. *The Journal of Cell Biology* 141, 993-1008. https://doi.org/10.1083/jcb.141.4.993

Collins, S.A., Behan, L., Harris, A., Gove, K., Lucas, J.S., 2016. **The dangers of widespread nitric oxide screening for primary ciliary dyskinesia**. *Thorax* 71, 560-561. http://doi.org/10.1136/thoraxjnl-2015-208056

Corbelli, R., Hammer, J., 2007. **Measurement of nasal nitric oxide**. *Paediatric Respiratory Reviews* 8, 269-272. https://doi.org/10.1016/j.prrv.2007.07.007

Cutting, G.R., 2014. **Cystic fibrosis genetics: from molecular understanding to clinical application**. *Nature Reviews Genetics* 16, 45. https://doi.org/10.1038/nrg3849

Czarnecki, P.G., Shah, J.V., 2012. **The ciliary transition zone: from morphology and molecules to medicine**. *Trends in Cell Biology* 22, 201-210. https://doi.org/10.1016/j.tcb.2012.02.001

Daniels, M., Leigh, M.W., Davis, S.D., Armstrong, M.C., Carson, J.L., *et al.*, 2013. **Founder mutation in RSPH4A identified in patients of Hispanic descent with primary ciliary dyskinesia**. *Human mutation* 34, 1352-1356. https://doi.org/10.1002/humu.22371

Davis, S.D., Ferkol, T.W., Rosenfeld, M., Lee, H.-S., Dell, S.D., *et al.*, 2015. **Clinical features of childhood primary ciliary dyskinesia by genotype and ultrastructural phenotype**. *American journal of respiratory and critical care medicine* 191, 316-324. https://doi.org/10.1164/rccm.201409-1672OC.

Dawe, H.R., Farr, H., Gull, K., 2007. **Centriole/basal body morphogenesis and migration during ciliogenesis in animal cells**. *Journal of Cell Science* 120, 7-15. http://doi.org/10.1242/jcs.03305

De Iongh, R., Rutland, J., 1989. **Orientation of respiratory tract cilia in patients with primary ciliary dyskinesia, bronchiectasis, and in normal subjects.** *Journal of clinical pathology* 42, 613-619. http://dx.doi.org/10.1136/jcp.42.6.613

De Paul, A.L., Mukdsi, J.H., Petiti, J.P., Gutiérrez, S., Quintar, A.A., *et al.*, 2012. **Immunoelectron microscopy: a reliable tool for the analysis of cellular processes**, in: Dehghani, D.H. (Ed.), *Applications of Immunocytochemistry*. InTech, InTech.https://doi.org/10.5772/33108

Deane, J.A., Cole, D.G., Seeley, E.S., Diener, D.R., Rosenbaum, J.L., 2001. **Localization of intraflagellar transport protein IFT52 identifies basal body transitional fibers as the docking site for IFT particles**. *Current Biology* 11, 1586-1590. https://doi.org/10.1016/S0960-9822(01)00484-5

Deng, H., Xia, H., Deng, S., 2014. **Genetic basis of human left–right asymmetry disorders**. *Expert Reviews in Molecular Medicine* 16, e19. http://doi.org/10.1017/erm.2014.22

DiPetrillo, C.G., Smith, E.F., 2010. **Pcdp1 is a central apparatus protein that binds Ca2+-calmodulin and regulates ciliary motility**. *The Journal of cell biology* 189, 601-612. https://doi.org/10.1083/jcb.200912009

Djakow, J., Svobodová, T., Hrach, K., Uhlík, J., Cinek, O., *et al.*, 2012. **Effectiveness of sequencing selected exons of DNAH5 and DNAI1 in diagnosis of primary ciliary dyskinesia**. *Pediatric pulmonology* 47, 864-875. https://doi.org/10.1002/ppul.22520

Dudbridge, F., 2003. **A survey of current software for linkage analysis**. *Human genomics* 1, 63.

Duriez, B., Duquesnoy, P., Escudier, E., Bridoux, A.-M., Escalier, D., *et al.*, 2007. **A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia**. *Proceedings of the National Academy of Sciences* 104, 3336-3341. https://doi.org/10.1073/pnas.0611405104

Eddy, E.M., Toshimori, K., O'Brien, D.a., 2003. **Fibrous sheath of mammalian spermatozoa.** *Microscopy research and technique* 61, 103-115. https://doi.org/10.1002/jemt.10320

Elston, R.C., Stewart, J., 1971. **A General Model for the Genetic Analysis of Pedigree Datal**. *Human Hereditary* 21, 523-542.

Escalier, D., Albert, M., 2006. **New fibrous sheath anomaly in spermatozoa of men with consanguinity**. *Fertility and Sterility* 86, 219-e211.

Failly, M., Bartoloni, L., Letourneau, A., Munoz, A., Falconnet, E., *et al.*, 2009. **Mutations in DNAH5 account for only 15% of a non-preselected cohort of patients with primary ciliary dyskinesia**. *Journal of medical genetics* 46, 281-286. https://doi.org/10.1136/jmg.2008.061176

Failly, M., Saitta, A., Munoz, A., Falconnet, E., Rossier, C., *et al.*, 2008. **DNAI1 mutations explain only 2% of primary ciliary dykinesia**. *Respiration* 76, 198-204. https://doi.org/10.1159/000128567

Feistel, K., Blum, M., 2006. **Three types of cilia including a novel 9+4 axoneme on the notochordal plate of the rabbit embryo**. *Developmental Dynamics* 235, 3348-3358. http://doi.org/10.1002/dvdy.20986

Ferkol, T., Leigh, M., 2006. **Primary Ciliary Dyskinesia and Newborn Respiratory Distress**. *Seminars in Perinatology* 30, 335-340. https://doi.org/10.1053/j.semperi.2005.11.001

Fernandez-Gonzalez, A., Kourembanas, S., Wyatt, T.A., Mitsialis, S.A., 2009. **Mutation of Murine Adenylate Kinase 7 Underlies a Primary Ciliary Dyskinesia Phenotype**. *American Journal of Respiratory Cell and Molecular Biology* 40, 305-313. https://doi.org/10.1165/rcmb.2008-0102OC

Firth, A.L., Dargitz, C.T., Qualls, S.J., Menon, T., Wright, R., *et al.*, 2014. **Generation of multiciliated cells in functional airway epithelia from human induced pluripotent stem cells**. *Proceedings of the National Academy of Sciences* 111, E1723-E1730. https://doi.org/10.1073/pnas.1403470111

Fliegauf, M., Benzing, T., Omran, H., 2007. **When cilia go bad: cilia defects and ciliopathies**. *Nature Reviews Molecular Cell Biology* 8, 880-893. https://doi.org/10.1038/nrm2278

Fliegauf, M., Olbrich, H., Horvath, J., Wildhaber, J.H., Zariwala, M.A., *et al.*, 2005. **Mislocalization of DNAH5 and DNAH9 in respiratory cells from patients with primary ciliary dyskinesia**. *American journal of respiratory and critical care medicine* 171, 1343-1349. 10.1164/rccm.200411-1583OC

Gaillard, A.R., Diener, D.R., Rosenbaum, J.L., Sale, W.S., 2001. **Flagellar Radial Spoke Protein 3 Is an a-Kinase Anchoring Protein (Akap)**. *The Journal of Cell Biology* 153 443-448. https://doi.org/10.1083/jcb.153.2.443

Garcia, G., III, Raleigh, D.R., Reiter, J.F., 2018. **How the Ciliary Membrane Is Organized Inside-Out to Communicate Outside-In**. *Current Biology* 28, R421-R434. http://doi.org/10.1016/j.cub.2018.03.010

Gardner, L.C., O'Toole, E., Perrone, C.A., Giddings, T., Porter, M.E., 1994. **Components of a" dynein regulatory complex" are located at the junction between the radial spokes and the dynein arms in Chlamydomonas flagella.** *The Journal of cell biology* 127, 1311-1325. https://doi.org/10.1083/jcb.127.5.1311

Gee, M.A., Heuser, J.E., Vallee, R.B., 1997. **An extended microtubule-binding structure within the dynein motor domain**. *Nature* 390, 636-639. http://doi.org/10.1038/37663

Ghahramani, Z., 2001. **An introduction to hidden Markov models and Bayesian networks**. *International journal of pattern recognition and artificial intelligence* 15, 9-42.

Gibbons, I.R., Garbarino, J.E., Tan, C.E., Reck-Peterson, S.L., Vale, R.D., *et al.*, 2005. **The Affinity of the Dynein Microtubule-binding Domain Is Modulated by the Conformation of Its Coiled-coil Stalk**. *Journal of Biological Chemistry* 280, 23960-23965. https://doi.org/10.1074/jbc.M501636200

Gibbs, J.R., Singleton, A., 2006. **Application of Genome-Wide Single Nucleotide Polymorphism Typing: Simple Association and Beyond**. *PLOS Genetics* 2, e150.

Gilbert, S., 2000. **Early Mammalian Development**, in: Sunderland (Ed.), *Developmental Biology*, 6 ed. Sinauer Associates.https://www.ncbi.nlm.nih.gov/books/NBK10052/

Goedken, R., Ludington, E., Crowe, R., Fyer, A.J., Hodge, S.E., *et al.*, 2000. **Drawbacks of GENEHUNTER for larger pedigrees: Application to panic disorder**. *American Journal of Medical Genetics* 96, 781-783. 10.1002/1096-8628(20001204)96:6<781::AID-AJMG17>3.0.CO;2-J

Goetz, S.C., Anderson, K.V., 2010. **The primary cilium: a signalling centre during vertebrate development**. *Nature Reviews Genetics* 11, 331-344. https://doi.org/10.1038/nrg2774

Gold, M.G., Lygren, B., Dokurno, P., Hoshi, N., McConnachie, G., *et al.*, 2006. **Molecular basis of AKAP specificity for PKA regulatory subunits**. *Molecular cell* 24, 383-395. https://doi.org/10.1016/j.molcel.2006.09.006

Gonçalves, J., Pelletier, L., 2017. **The Ciliary Transition Zone: Finding the Pieces and Assembling the Gate**. *Molecules and cells* 40, 243-253. http://doi.org/10.14348/molcells.2017.0054

Goodenough, U.W., Heuser, J.E., 1985. **Substructure of inner dynein arms, radial spokes, and the central pair/projection complex of cilia and flagella.** *The Journal of cell biology* 100, 2008-2018. https://doi.org/10.1083/jcb.100.6.2008

Goodship, J., Gill, H., Carter, J., Jackson, A., Splitt, M., *et al.*, 2000. **Report Autozygosity Mapping of a Seckel Syndrome Locus to Chromosome 3q22.1-q24**. *Am. J. Hum. Genet* 67, 498-503.

Görmez, Z., Bakir-Gungor, B., Sağıroğlu, M.Ş., Gormez, Z., Bakir-Gungor, B., *et al.*, 2013. **HomSI: a homozygous stretch identifier from next-generation sequencing data**. *Bioinformatics* 30, 445-447. 10.1093/bioinformatics/btt686

Goutaki, M., Meier, A.B., Halbeisen, F.S., Lucas, J.S., Dell, S.D., *et al.*, 2016. **Clinical manifestations in primary ciliary dyskinesia: systematic review and meta-analysis**. *European respiratory journal* 48, 1081-1095. https://doi.org/10.1183/13993003.00736-2016

Greenstone, M.A., Jones, R.W., Dewar, A., Neville, B.G., Cole, P.J., 1984. **Hydrocephalus and primary ciliary dyskinesia**. *Archives of Disease in Childhood* 59, 481-482. https://doi.org/10.1136/adc.59.5.481

Gudbjartsson, D.F., Jonasson, K., Frigge, M.L., Kong, A., 2000. **Allegro, a new computer program for multipoint linkage analysis**. *Nature Genetics* 25, 12-13. 10.1038/75514

Guichard, C., Harricane, M.-C., Lafitte, J.-J., Godard, P., Zaegel, M., *et al.*, 2001. **Axonemal Dynein Intermediate-Chain Gene (< i> DNAI1</i>) Mutations Result in Situs Inversus and Primary Ciliary Dyskinesia (Kartagener Syndrome)**. *The American Journal of Human Genetics* 68, 1030-1035. https://doi.org/10.1086/319511

Gusev, A., Lowe, J.K., Stoffel, M., Daly, M.J., Altshuler, D., *et al.*, 2009. **Whole population, genome-wide mapping of hidden relatedness**. *Genome Research* 19, 318-326. 10.1101/gr.081398.108

Hamada, H., Meno, C., Watanabe, D., Saijoh, Y., 2002. **Establishment of vertebrate left–right asymmetry**. *Nature Reviews Genetics* 3, 103-113. http://doi.org/10.1038/nrg732

Heuser, T., Raytchev, M., Krell, J., Porter, M.E., Nicastro, D., 2009. **The dynein regulatory complex is the nexin link and a major regulatory node in cilia and flagella**. *The Journal of cell biology* 187, 921-933. https://doi.org/10.1083/jcb.200908067

Hildebrand, M.S., Avenarius, M.R., Fellous, M., Zhang, Y., Meyer, N.C., *et al.*, 2010. **Genetic male infertility and mutation of CATSPER ion channels**. *European Journal of Human Genetics* 18, 1178-1184. https://doi.org/10.1038/ejhg.2010.108

Hirokawa, N., Tanaka, Y., Okada, Y., Takeda, S., 2006. **Nodal Flow and the Generation of Left-Right Asymmetry**. *Cell* 125, 33-45. http://doi.org/10.1016/j.cell.2006.03.002

Hirst, R.A., Rutman, A., Williams, G., O'Callaghan, C., 2010. **Ciliated air-liquid cultures as an aid to diagnostic testing of primary ciliary dyskinesia**. *Chest* 138, 1441-1447. https://doi.org/10.1378/chest.10-0175

Hjeij, R., Lindstrand, A., Francis, R., Zariwala, M.A., Liu, X., *et al.*, 2013. **< i> ARMC4</i> Mutations Cause Primary Ciliary Dyskinesia with Randomization of**

Left/Right Body Asymmetry. *The American Journal of Human Genetics* 93, 357-367. https://doi.org/10.1016/j.ajhg.2013.06.009

Hjeij, R., Onoufriadis, A., Watson, C.M., Slagle, C.E., Klena, N.T., *et al.*, 2014. **CCDC151 mutations cause primary ciliary dyskinesia by disruption of the outer dynein arm docking complex formation**. *The American Journal of Human Genetics* 95, 257-274. https://doi.org/10.1016/j.ajhg.2014.08.005

Homer, N., 2010. TMAP: The torrent mapping program.

Horani, A., Brody, S.L., Ferkol, T.W., Shoseyov, D., Wasserman, M.G., *et al.*, 2013. **CCDC65 mutation causes primary ciliary dyskinesia with normal ultrastructure and hyperkinetic cilia.** *PloS one* 8, e72299. https://doi.org/10.1371/journal.pone.0072299

Horani, A., Druley, T.E., Zariwala, M.A., Patel, A.C., Levinson, B.T., *et al.*, 2012. **Whole-Exome Capture and Sequencing Identifies< i> HEATR2</i> Mutation as a Cause of Primary Ciliary Dyskinesia**. *The American Journal of Human Genetics* 91, 685-693. https://doi.org/10.1016/j.ajhg.2012.08.022

Hornef, N., Olbrich, H., Horvath, J., Zariwala, M.a., Fliegauf, M., *et al.*, 2006. **DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arm defects.** *American journal of respiratory and critical care medicine* 174, 120-126. https://doi.org/10.1164/rccm.200601-084OC

Horváth, J., Fliegauf, M., Olbrich, H., Kispert, A., King, S.M., *et al.*, 2005. **Identification and analysis of axonemal dynein light chain 1 in primary ciliary dyskinesia patients**. *American journal of respiratory cell and molecular biology* 33, 41-47. https://doi.org/10.1165/rcmb.2004-0335OC

Hung, M.-C., Link, W., 2011. **Protein localization in disease and therapy**. *J Cell Sci* 124, 3381-3392. 10.1242/jcs.089110

Hwang, J., Hunter, E.L., Sale, W.S., Wirschell, M., 2018. **Control of axonemal inner dynein arms**, in: King, S.M. (Ed.), *Dyneins* Second Edition ed. Academic Press, pp. 270-297.https://doi.org/10.1016/B978-0-12-809471-6.00009-7

Ibañez-Tallon, I., Gorokhova, S., Heintz, N., 2002. **Loss of function of axonemal dynein Mdnah5 causes primary ciliary dyskinesia and hydrocephalus**. *Human Molecular Genetics* 11, 715-721. https://doi.org/10.1093/hmg/11.6.715

Ideguchi, H., Ueda, A., Tanaka, M., Yang, J., Tsuji, T., *et al.*, 2002. **Structural and functional characterization of the USP11 deubiquitinating enzyme, which interacts with the RanGTP-associated protein RanBPM**. *The Biochemical journal* 367, 87-95. http://doi.org/10.1042/BJ20011851

Inaba, K., 2007. **Molecular basis of sperm flagellar axonemes**. *Annals of the New York Academy of Sciences* 1101, 506-526. https://doi.org/10.1196/annals.1389.017

Inaba, K., 2011. **Sperm flagella: comparative and phylogenetic perspectives of protein components Unicellular algae Chlamydomonas**. *Molecular Human Reproduction* 17, 524-538. https://doi.org/10.1093/molehr/gar034

Inaba, K., Mizuno, K., 2016. **Sperm dysfunction and ciliopathy**. *Reproductive medicine and biology* 15, 77-94. https://doi.org/10.1007/s12522-015-0225-5

INE, I.N.d.E., 2002. Marriages (Between persons of the opposite sex - No.) by Place of registration (NUTS - 2002), Sex, Relationship or affinity between the spouses and Spouse previous marital status; Annual.

Ishikawa, H., Marshall, W.F., 2011. **Ciliogenesis: building the cell's antenna**. *Nature reviews Molecular cell biology* 12, 222-234. https://doi.org/10.1038/nrm3085

Ishikawa, T., 2017. **Axoneme Structure from Motile Cilia**. *Cold Spring Harbor Perspectives in Biology* 9, 1-18. https://doi.org/10.1101/cshperspect.a028076

Jackson, C.L., Behan, L., Collins, S.A., Goggin, P.M., Adam, E.C., *et al.*, 2016. **Accuracy of diagnostic testing in primary ciliary dyskinesia**. *European respiratory journal* 47, 837-848. https://doi.org/10.1183/13993003.00749-2015

Jackson, D.A., Symonst, R.H., Berg, P., 1972. **Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of Escherichia coli**. *Proceedings of the National Academy of Sciences of the United States of America* 69, 2904-2909. https://doi.org/10.1073/pnas.69.10.2904

Jamuar, S.S., Tan, E.-C., 2015. **Clinical application of next-generation sequencing for Mendelian diseases**. *Human Genomics* 9, 10. https://doi.org/10.1186/s40246-015-0031-5

Jensen, V.L., Lambacher, N.J., Li, C., Mohan, S., Williams, C.L., *et al.*, 2018. **Role for intraflagellar transport in building a functional transition zone**. *EMBO reports* 19, e45862. https://doi.org/10.15252/embr.201845862

Joensen, O., Paff, T., Haarman, E.G., Skovgaard, I.M., Jensen, P.Ø., *et al.*, 2014. **Exhaled breath analysis using electronic nose in cystic fibrosis and primary ciliary dyskinesia patients with chronic pulmonary infections**. *PLoS One* 9, e115584. https://doi.org/10.1371/journal.pone.0115584

Jorissen, M., Willems, T., Van der Schueren, B., 2000. **Ciliary function analysis for the diagnosis of primary ciliary dyskinesia: advantages of ciliogenesis in culture**. *Acta oto-laryngologica* 120, 291-295. https://doi.org/10.1080/000164800750001116

Kann, M.-L., Fouquet, J.-P., 1989. **Comparison of LR white resin, lowicryl K4M and Epon postembedding procedures for immunogold staining of actin in the testis**. *Histochemistry* 91, 221-226. https://doi.org/10.1007/bf00490136

Kano, G., Tsujii, H., Takeuchi, K., Nakatani, K., Ikejiri, M., *et al.*, 2016. **Whole-exome sequencing identification of novel DNAH5 mutations in a young patient with primary ciliary dyskinesia**. *Molecular medicine reports* 14, 5077-5083. https://doi.org/10.3892/mmr.2016.5871

Kim, R.H., A Hall, D., Cutz, E., Knowles, M.R., Nelligan, K.A., *et al.*, 2014. **The role of molecular genetic analysis in the diagnosis of primary ciliary dyskinesia**. *Annals of the American Thoracic Society* 11, 351-359. http://doi.org/10.1513/AnnalsATS.201306-194OC

King, S.M., 2000. **The dynein microtubule motor**. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1496, 60-75. http://dx.doi.org/10.1016/S0167-4889(00)00009-4

King, S.M., 2016. **Axonemal Dynein Arms**. *Cold Spring Harbor perspectives in biology* 8, a028100. https://doi.org/10.1101/cshperspect.a028100

King, S.M., 2018a. **Composition and assembly of axonemal dyneins**, in: King, S.M. (Ed.), *Dyneins (Second Edition)*. Academic Press, Elsevier pp. 162-201.https://doi.org/10.1016/B978-0-12-809471-6.00005-X

King, S.M., 2018b. **Regulatory mechanics of outer-arm dynein motors**, in: King, S.M. (Ed.), *Dyneins* 2ed. Academic Press, pp. 250-269.https://doi.org/10.1016/B978-0-12-809471-6.00008-5

King, S.M., Patel-King, R.S., 2015. **The oligomeric outer dynein arm assembly factor CCDC103 is tightly integrated within the ciliary axoneme and exhibits periodic binding to microtubules**. *Journal of Biological Chemistry* 290, 7388-7401. http://dx.doi.org/10.1074/jbc.M114.616425

Kispert, A., Petry, M., Olbrich, H., Volz, A., Ketelsen, U.P., *et al.*, 2003. **Genotype-phenotype correlations in PCD patients carrying DNAH5 mutations**. *Thorax* 58, 552-554. http://dx.doi.org/10.1136/thorax.58.6.552-b

Knowles, M.R., Daniels, L.A., Davis, S.D., Zariwala, M.A., Leigh, M.W., 2013a. **Primary Ciliary Dyskinesia Recent Advances in Diagnostics, Genetics, and Characterization of Clinical Disease**. *American Journal of Respiratory and Critical Care Medicine* 188, 913-922. https://doi.org/10.1164/rccm.201301-0059CI

Knowles, M.R., Leigh, M.W., Carson, J.L., Davis, S.D., Dell, S.D., *et al.*, 2012. **Mutations of DNAH11 in patients with primary ciliary dyskinesia with normal ciliary ultrastructure**. *Thorax* 67, 433-441. https://doi.org/10.1136/thoraxjnl-2011-200301

Knowles, M.R., Leigh, M.W., Ostrowski, L.E., Huang, L., Carson, J.L., *et al.*, 2013b. **Exome Sequencing Identifies Mutations in< i> CCDC114</i> as a Cause of Primary Ciliary Dyskinesia**. *The American Journal of Human Genetics* 92, 99-106. https://doi.org/10.1016/j.ajhg.2012.11.003

Knowles, M.R., Ostrowski, L.E., Leigh, M.W., Sears, P.R., Davis, S.D., *et al.*, 2014. **Mutations in RSPH1 cause primary ciliary dyskinesia with a unique clinical and ciliary phenotype**. *American journal of respiratory and critical care medicine* 189, 707-717. https://doi.org/10.1164/rccm.201311-2047OC

Knowles, M.R., Ostrowski, L.E., Loges, N.T., Hurd, T., Leigh, M.W., *et al.*, 2013c. **Mutations in< i> SPAG1</i> Cause Primary Ciliary Dyskinesia Associated with Defective Outer and Inner Dynein Arms**. *The American Journal of Human Genetics* 93, 711-720. https://doi.org/10.1016/j.ajhg.2013.07.025

Koboldt, Daniel C., Steinberg, Karyn M., Larson, David E., Wilson, Richard K., Mardis, E.R., 2013. **The Next-Generation Sequencing Revolution and Its Impact on Genomics**. *Cell* 155, 27-38. 10.1016/j.cell.2013.09.006

Komor, A.C., Badran, A.H., Liu, D.R., 2017. **CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes**. *Cell* 168, 20-36. https://doi.org/10.1016/j.cell.2016.10.044

Kon, T., Oyama, T., Shimo-Kon, R., Imamula, K., Shima, T., *et al.*, 2012. **The 2.8 Å crystal structure of the dynein motor domain**. *Nature* 484, 345. https://doi.org/10.1038/nature10955

Konishi, S., Gotoh, S., Tateishi, K., Yamamoto, Y., Korogi, Y., *et al.*, 2016. **Directed Induction of Functional Multi-ciliated Cells in Proximal Airway Epithelial Spheroids from Human Pluripotent Stem Cells**. *Stem Cell Reports* 6, 18-25. https://doi.org/10.1016/j.stemcr.2015.11.010

Koonce, M.P., Tikhonenko, I., 2000. **Functional elements within the dynein microtubule-binding domain.** *Molecular biology of the cell* 11, 523-529. 10.1091/mbc.11.2.523

Kott, E., Legendre, M., Copin, B., Papon, J.-F., Dastot-Le Moal, F., *et al.*, 2013. **Loss-of-Function Mutations in< i> RSPH1</i> Cause Primary Ciliary Dyskinesia with Central-Complex and Radial-Spoke Defects**. *The American Journal of Human Genetics* 93, 561-570. https://doi.org/10.1016/j.ajhg.2013.07.013

Krawitz, P.M., Schweiger, M.R., Rödelsperger, C., Marcelis, C., Kölsch, U., *et al.*, 2010. **Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome**. *Nature genetics* 42, 827-829.

Kruglyak, L., Daly, M.J., Reeve-Daly, M.P., Lander ', E.S., Lander, E.S., 1996. **Parametric and Nonparametric Linkage Analysis: A Unified Multipoint Approach**. *Am. J. Hum. Genet* 58, 1347-1363.

Lander, E.S., Botstein, D., 1987. **Homozygosity Mapping: A Way to Map Human Recessive Traits with the DNA of Inbred Children**. *Science* 236, 1568-1570.

Lee, H., Deignan, J.L., Dorrani, N., Strom, S.P., Kantarci, S., *et al.*, 2014. **Clinical exome sequencing for genetic identification of rare Mendelian disorders.** *JAMA* 312, 1880-1887. https://doi.org/10.1001/jama.2014.14604

Lee, L., 2013. **Riding the wave of ependymal cilia: Genetic susceptibility to hydrocephalus in primary ciliary dyskinesia**. *Journal of Neuroscience Research* 91, 1117-1132. https://doi.org/10.1002/jnr.23238

Lefebvre, R., Chenoweth, P.J., Drost, M., LeClear, C.T., MacCubbin, M., *et al.*, 1995. **Characterization of the oviductal sperm reservoir in cattle**. *Biology of reproduction* 53, 1066-1074. https://doi.org/10.1095/biolreprod53.5.1066

Leigh, M.W., Ferkol, T.W., Davis, S.D., Lee, H.-S., Rosenfeld, M., *et al.*, 2016. **Clinical Features and Associated Likelihood of Primary Ciliary Dyskinesia in Children and Adolescents**. *Annals of the American Thoracic Society* 13, 1305-1313. http://doi.org/10.1513/AnnalsATS.201511-748OC

Leigh, M.W., Pittman, J.E., Carson, J.L., Ferkol, T.W., Dell, S.D., *et al.*, 2009. **Clinical and genetic aspects of primary ciliary dyskinesia/Kartagener syndrome.** *Genetics in medicine* 11, 473-487. https://doi.org/10.1097/GIM.0b013e3181a53562

Lewandowska, M.A., 2013. **The missing puzzle piece: splicing mutations**. *International journal of clinical and experimental pathology* 6, 2675.

Li, Y., Liu, H., Li, J., Zhang, Q., Gong, S., *et al.*, 2014. **Morphology and Ciliary Motion of Mucosa in the Eustachian Tube of Neonatal and Adult Gerbils**. *PLOS ONE* 9, e99840. https://doi.org/10.1371/journal.pone.0099840

Li, Y., Tollefsbol, T.O., 2011. DNA methylation detection: bisulfite genomic sequencing analysis., in: Tollefsbol, T. (Ed.), Epigenetics Protocols. Methods in Molecular Biology (Methods and Protocols). Humana Press, pp. 11-21.

Lin, J., Tritschler, D., Song, K., Barber, C.F., Cobb, J.S., *et al.*, 2011. **Building Blocks of the Nexin-Dynein Regulatory Complex in Chlamydomonas Flagella**. *Journal of Biological Chemistry* 286, 29175-29191. http://doi.org/10.1074/jbc.M111.241760

Lishko, P.V., Kirichok, Y., 2010. **The role of Hv1 and CatSper channels in sperm activation**. *The Journal of physiology* 588, 4667-4672. https://doi.org/10.1113/jphysiol.2010.194142

Lobo, I., Shaw, K., 2008. **Discovery and types of genetic linkage**. *Nature Education* 1, 139.

Loges, N.T., Antony, D., Maver, A., Deardorff, M.A., Güleç, E.Y., *et al.*, 2018. **Recessive DNAH9 Loss-of-Function Mutations Cause Laterality Defects and Subtle Respiratory Ciliary-Beating Defects**. *The American Journal of Human Genetics* 103, 995-1008. https://doi.org/10.1016/j.ajhg.2018.10.020

Loges, N.T., Olbrich, H., Becker-Heck, A., Häffner, K., Heer, A., *et al.*, 2009. **Deletions and point mutations of LRRC50 cause primary ciliary dyskinesia due to dynein arm defects.** *American journal of human genetics* 85, 883-889. http://doi.org/10.1016/j.ajhg.2009.10.018

Loges, N.T., Olbrich, H., Fenske, L., Mussaffi, H., Horvath, J., *et al.*, 2008. **DNAI2 mutations cause primary ciliary dyskinesia with defects in the outer dynein arm.** *American journal of human genetics* 83, 547-558. https://doi.org/10.1016/j.ajhg.2008.10.001

Lok, S., Johnston, D.S., Conklin, D., Lofton-Day, C.E., Adams, R.L., *et al.*, 2000. **Identification of INSL6, a new member of the insulin family that is expressed in the testis of the human and rat**. *Biology of reproduction* 62, 1593-1599.

Lucas, J.S., Adam, E.C., Goggin, P.M., Jackson, C.L., Powles-Glover, N., *et al.*, 2012. **Static respiratory cilia associated with mutations in Dnahc11/DNAH11: a mouse model of PCD**. *Human mutation* 33, 495-503. https://doi.org/10.1002/humu.22001

Lucas, J.S., Burgess, A., Mitchison, H.M., Moya, E., Williamson, M., *et al.*, 2014. **Diagnosis and management of primary ciliary dyskinesia**. *Archives of disease in childhood* 99, 850-856. https://doi.org/10.1136/archdischild-2013-304831

Luconi, M., Cantini, G., Baldi, E., Forti, G., 2010. **Role of a-kinase anchoring proteins (AKAPs) in reproduction.** *Frontiers in Bioscience* 16, 1315-1330. https://doi.org/10.2741/3791

Luconi, M., Forti, G., Baldi, E., 2006. **Pathophysiology of sperm motility**. *Frontiers in Bioscience* 11, 1433-1447. https://doi.org/10.2741/1894

Lurie, M., Tur-Kaspa, I., Weill, S., Katz, I., Rabinovici, J., *et al.*, 1989. **Ciliary Ultrastructure of Respiratory and Fallopian Tube Epithelium in a Sterile Woman with Kartagener's Syndrome: A Quantitative Estimation**. *CHEST* 95, 578-581. https://doi.org/10.1378/chest.95.3.578

Lyons, R.A., Saridogan, E., Djahanbakhch, O., 2006. **The reproductive significance of human Fallopian tube cilia**. *Human Reproduction Update* 12, 363-372. http://doi.org/10.1093/humupd/dml012

MacCluer, J.W., VandeBerg, J.L., Read, B., Ryder, O.A., 1986. **Pedigree analysis by computer simulation**. *Zoo Biology* 5, 147-160. 10.1002/zoo.1430050209

Magi, A., Tattini, L., Palombo, F., Benelli, M., Gialluisi, A., *et al.*, 2014. **H3M2: detection of runs of homozygosity from whole-exome sequencing data.** *Bioinformatics (Oxford, England)* 30, 2852-2859. 10.1093/bioinformatics/btu401

Mali, G.R., Yeyati, P.L., Mizuno, S., Dodd, D.O., Tennant, P.A., *et al.*, 2018. **ZMYND10 functions in a chaperone relay during axonemal dynein assembly**. *eLife* 7. https://doi.org/10.7554/eLife.34389

Maniscalco, M., Motta, A., 2017. **Metabolomics of exhaled breath condensate: a means for phenotyping respiratory diseases?** *Future Medicine*. https://doi.org/10.2217/bmm-2017-0068

Mardis, E.R., 2008. **The impact of next-generation sequencing technology on genetics**. *Trends in Genetics* 24, 133-141. 10.1016/j.tig.2007.12.007

Marshall, C.R., Scherer, S.W., Zariwala, M.A., Lau, L., Paton, T.A., *et al.*, 2015. **Whole-Exome Sequencing and Targeted Copy Number Analysis in Primary Ciliary Dyskinesia**. *G3: Genes|Genomes|Genetics* 5, 1775-1781. https://doi.org/10.1534/g3.115.019851

Marshall, W.F., 2008. **Basal Bodies: Platforms for Building Cilia**, in: Yoder, B.K. (Ed.), *Current Topics in Developmental Biology:Ciliary Function in Mammalian Development*. Academic Press, Elsevier, pp. 1-22.https://doi.org/10.1016/S0070-2153(08)00801-6

Marshall, W.F., Rosenbaum, J.L., 2001. **Intraflagellar transport balances continuous turnover of outer doublet microtubules: implications for flagellar length control**. *Journal of Cellular Biology* 155, 405-414. http://doi.org/10.1083/jcb.200106141

Matlin, A.J., Clark, F., Smith, C.W.J., 2005. **UNDERSTANDING ALTERNATIVE SPLICING : TOWARDS A CELLULAR CODE**. *Nature Reviews Molecular Cell Biology* 6, 386-399. 10.1038/nrm1645

Matoulkova, E., Michalova, E., Vojtesek, B., Hrstka, R., 2012. **The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells**. *RNA Biology* 9, 563-576. https://doi.org/10.4161/rna.20231

Maurizy, C., Quinternet, M., Abel, Y., Verheggen, C., Santo, P.E., *et al.*, 2018. **The RPAP3-Cterminal domain identifies R2TP-like quaternary chaperones**. *Nature Communications* 9, 2093. https://doi.org/10.1038/s41467-018-04431-1

Maxam, A.M., Gilbert, W., 1977. **A new method for sequencing DNA** *Proceedings of the National Academy of Sciences of the United States of America* 74, 560-564. https://doi.org/10.1073/pnas.74.2.560

Mazor, M., Alkrinawi, S., Chalifa-Caspi, V., Manor, E., Sheffield, V.C., *et al.*, 2011. **Primary Ciliary Dyskinesia Caused by Homozygous Mutation in< i> DNAL1</i>, Encoding Dynein Light Chain 1**. *The American Journal of Human Genetics* 88, 599-607. https://doi.org/10.1016/j.ajhg.2011.03.018

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., *et al.*, 2008. **Runs of homozygosity in European populations.** *American journal of human genetics* 83, 359-372. 10.1016/j.ajhg.2008.08.007

Meienberg, J., Bruggmann, R., Oexle, K., Matyas, G., 2016. **Clinical sequencing: is WGS the better WES?** *Human genetics* 135, 359-362. https://doi.org/10.1007/s00439-015-1631-9

Mendeliome Group, S., 2015. **Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases.** *Genome biology* 16, 134. https://doi.org/10.1186/s13059-015-0693-2

Merveille, A.-C., Davis, E.E., Becker-Heck, A., Legendre, M., Amirav, I., *et al.*, 2011. **CCDC39 is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs.** *Nature genetics* 43, 72-78. https://doi.org/10.1038/ng.726

Michel, J.J.C., Scott, J.D., 2002. **AKAP mediated signal transduction**. *Annual review of pharmacology and toxicology* 42, 235-257. https://doi.org/10.1146/annurev.pharmtox.42.083101.135801

Miki, K., Willis, W.D., Brown, P.R., Goulding, E.H., Fulcher, K.D., *et al.*, 2002. **Targeted Disruption of the< i> Akap4</i> Gene Causes Defects in Sperm Flagellum and Motility**. *Developmental biology* 248, 331-342.

Miko, I., 2008. **Gregor Mendel and the Principles of Inheritance**. *Nature Education* 1, 134.

Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. **A simple salting out procedure for extracting DNA from human nucleated cells.** *Nucleic acids research* 16, 1215. https://doi.org/10.1093/nar/16.3.1215

Mirvis, M., Stearns, T., James Nelson, W., 2018. **Cilium structure, assembly, and disassembly regulated by the cytoskeleton**. *Biochemical Journal* 475, 2329-2353. https://doi.org/10.1042/BCJ20170453

Mitchison, H.M., Schmidts, M., Loges, N.T., Freshour, J., Dritsoula, A., *et al.*, 2012. **Mutations in axonemal dynein assembly factor DNAAF3 cause primary ciliary dyskinesia**. *Nature genetics* 44, 381-389. https://doi.org/10.1038/ng.1106

Mocz, G., Gibbons, I.R., 2001. **Model for the motor component of dynein heavy chain based on homology to the AAA family of oligomeric ATPases**. *Structure* 9, 93-103. 10.1016/S0969-2126(00)00557-8

Monteiro, M.S., Carvalho, M., Bastos, M.L., Guedes de Pinho, P., 2013. **Metabolomics analysis for biomarker discovery: advances and challenges**. *Current medicinal chemistry* 20, 257-271. https://doi.org/10.2174/0929867311320020006

Montuschi, P., Paris, D., Montella, S., Melck, D., Mirra, V., *et al.*, 2014. **Nuclear magnetic resonance–based metabolomics discriminates primary ciliary dyskinesia from cystic fibrosis**. *American journal of respiratory and critical care medicine* 190, 229-233. https://doi.org/10.1164/rccm.201402-0249LE.

Moore, D.J., Onoufriadis, A., Shoemark, A., Simpson, M.A., Zur Lage, P.I., *et al.*, 2013. **Mutations in< i> ZMYND10</i>, a Gene Essential for Proper Axonemal Assembly of Inner and Outer Dynein Arms in Humans and Flies, Cause Primary Ciliary Dyskinesia**. *The American Journal of Human Genetics* 93, 346-356. https://doi.org/10.1016/j.ajhg.2013.07.009

Moretti, E., Scapigliati, G., Pascarelli, N.A., Baccetti, B., Collodel, G., 2007. **Localization of AKAP4 and tubulin proteins in sperm with reduced motility.** *Asian journal of andrology* 9, 641-649. 10.1111/j.1745-7262.2007.00267.x

Morillas, H.N., Zariwala, M., Knowles, M.R., 2007. **Genetic causes of bronchiectasis: primary ciliary dyskinesia**. *Respiration* 74, 252-263. https://doi.org/10.1159/000101783

Muhlebach, M.S., Sha, W., 2015. **Lessons learned from metabolomics in cystic fibrosis**. *Molecular and cellular pediatrics* 2, 9. https://doi.org/10.1186/s40348-015-0020-8

Mulugeta, S., Gray, J.M., Notarfrancesco, K.L., Gonzales, L.W., Koval, M., *et al.*, 2002. **Identification of LBM180, a lamellar body limiting membrane protein of alveolar type II cells, as the ABC transporter protein ABCA3**. *Journal of Biological Chemistry* 277, 22147-22155. https://doi.org/10.1074/jbc.M201812200

Munkholm, M., Mortensen, J., 2014. **Mucociliary clearance: pathophysiological aspects**. *Clinical Physiology and Functional Imaging* 34, 171-177. http://doi.org/10.1111/cpf.12085

Munro, N.C., Currie, D.C., Lindsay, K.S., Ryder, T.A., Rutman, A., *et al.*, 1994. **Fertility in men with primary ciliary dyskinesia presenting with respiratory infection.** *Thorax* 49, 684-687. http://dx.doi.org/10.1136/thx.49.7.684

Munye, M.M., Tagalakis, A.D., Barnes, J.L., Brown, R.E., McAnulty, R.J., *et al.*, 2016. **Minicircle DNA Provides Enhanced and Prolonged Transgene Expression Following Airway Gene Transfer**. *Scientific Reports* 6, 23125. http://doi.org/10.1038/srep23125

Nakamura, T., Hamada, H., 2012. **Left-right patterning: conserved and divergent mechanisms**. *Development* 139, 3257-3262. https://doi.org/10.1242/dev.061606

Narita, K., Kozuka-Hata, H., Nonami, Y., Ao-Kondo, H., Suzuki, T., *et al.*, 2012. **Proteomic analysis of multiple primary cilia reveals a novel mode of ciliary development in mammals**. *Biology Open* 1, 815-825. https://doi.org/10.1242/bio.20121081

Neely, M.D., Boekelheide, K., 1988. **Sertoli cell processes have axoplasmic features: an ordered microtubule distribution and an abundant high molecular weight microtubule-associated protein (cytoplasmic dynein).** *The Journal of cell biology* 107, 1767-1776. https://doi.org/10.1083/JCB.107.5.1767

Nicastro, D., Schwartz, C., Pierson, J., Gaudette, R., Porter, M.E., *et al.*, 2006. **The molecular architecture of axonemes revealed by cryoelectron tomography**. *Science* 313, 944-948. https://doi.org/10.1126/science.1128618

Nobakht, F., Aliannejad, R., Rezaei tavirani, M., Taheri, S., Arefi Oskouie, A., 2014. **The metabolomics of airway diseases, including COPD, asthma and cystic fibrosis**. *Biomarkers* 20. https://doi.org/10.3109/1354750X.2014.983167

Nonaka, S., Tanaka, Y., Okada, Y., Takeda, S., Harada, A., *et al.*, 1998. **Randomization of Left–Right Asymmetry due to Loss of Nodal Cilia Generating Leftward Flow of Extraembryonic Fluid in Mice Lacking KIF3B Motor Protein**. *Cell* 95, 829-837. https://doi.org/10.1016/S0092-8674(00)81705-5

Nonaka, S., Yoshiba, S., Watanabe, D., Ikeuchi, S., Goto, T., *et al.*, 2005. **De Novo Formation of Left–Right Asymmetry by Posterior Tilt of Nodal Cilia**. *PLOS Biology* 3, e268. http://doi.org/10.1371/journal.pbio.0030268

Noone, P.G., Leigh, M.W., Sannuti, A., Minnix, S.L., Carson, J.L., *et al.*, 2004. **Primary Ciliary Dyskinesia**. *American Journal of Respiratory and Critical Care Medicine* 169, 459-467. https://doi.org/10.1164/rccm.200303-365OC

Norwood, J., Hein, C., Halbert, S., Anderson, R., 1978. **Polycationic macromolecules inhibit cilia-mediated ovum transport in the rabbit oviduct**. *Proceedings of the National Academy of Sciences* 75, 4413-4416. https://doi.org/10.1073/pnas.75.9.4413

Oakley, B.R., 2000. **An abundance of tubulins.** *Trends in cell biology* 10, 537-542. https://doi.org/10.1016/S0962-8924(00)01857-2

Ohashi, H., Hasegawa, M., Wakimoto, K., Miyamoto-Sato, E., 2015. **Next-Generation Technologies for Multiomics Approaches Including Interactome Sequencing**. *BioMed Research International* 2015, 1-9. https://doi.org/10.1155/2015/104209

Ohashi, Y., Nakai, Y., Kihara, S., Maruoka, K., Ikeoka, H., *et al.*, 1985. **The Ciliary Activity of the Middle Ear Lining-functional and Morphological Observation**. *Auris Nasus Larynx* 12, S123-S125. https://doi.org/10.1016/S0385-8146(85)80122-X

Okada, Y., Takeda, S., Tanaka, Y., Belmonte, J.-C.I., Hirokawa, N., 2005. **Mechanism of Nodal Flow: A Conserved Symmetry Breaking Event in Left-Right Axis Determination**. *Cell* 121, 633-644. https://doi.org/10.1016/j.cell.2005.04.008

Olbrich, H., Häffner, K., Kispert, A., Völkel, A., Volz, A., *et al.*, 2002. **Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry.** *Nature genetics* 30, 143-144. https://doi.org/10.1038/ng817

Olbrich, H., Schmidts, M., Werner, C., Onoufriadis, A., Loges, N.T., *et al.*, 2012. **Recessive< i> HYDIN</i> Mutations Cause Primary Ciliary Dyskinesia without Randomization of Left-Right Body Asymmetry**. *The American Journal of Human Genetics* 91, 672-684. https://doi.org/10.1016/j.ajhg.2012.08.016

Oliveira, J., Negrao, L., Fineza, I., Taipa, R., Melo-Pires, M., *et al.*, 2015. **New splicing mutation in the choline kinase beta (CHKB) gene causing a muscular dystrophy detected by whole-exome sequencing**. *Journal of Human Genetics* 60, 305-312. https://doi.org/10.1038/jhg.2015.20.

Oliveira, J., Pereira, R., Santos, R., Sousa, M., 2017. Homozygosity Mapping using Whole-Exome Sequencing: A Valuable Approach for Pathogenic Variant Identification in Genetic Diseases, Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 3: BIOINFORMATICS, (BIOSTEC 2017), pp. 2010-2016.

Oliveira, J., Pereira, R., Santos, R., Sousa, M., 2018. Evaluating Runs of Homozygosity in Exome Sequencing Data - Utility in Disease Inheritance Model Selection and Variant Filtering, Biomedical Engineering Systems and Technologies, pp. 268-288.

Omoto, C.K., Gibbons, I.R., Kamiya, R., Shingyoji, C., Takahashi, K., *et al.*, 1999. **Rotation of the central pair microtubules in eukaryotic flagella**. *Molecular biology of the cell* 10, 1-4. https://doi.org/10.1091/mbc.10.1.1

Omran, H., Kobayashi, D., Olbrich, H., Tsukahara, T., Loges, N.T., *et al.*, 2008. **Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins**. *Nature* 456, 611-616. https://doi.org/10.1038/nature07471

Onoufriadis, A., Paff, T., Antony, D., Shoemark, A., Micha, D., *et al.*, 2013a. **Splice-site mutations in the axonemal outer dynein arm docking complex gene CCDC114 cause primary ciliary dyskinesia.** *American journal of human genetics* 92, 88-98. https://doi.org/10.1016/j.ajhg.2012.11.002

Onoufriadis, A., Shoemark, A., Munye, M.M., James, C.T., Schmidts, M., *et al.*, 2013b. **Combined exome and whole-genome sequencing identifies mutations in ARMC4 as a cause of primary ciliary dyskinesia with defects in the outer dynein arm**. *Journal of medical genetics* 51, 61-67. http://dx.doi.org/10.1136/jmedgenet-2013-101938

Onoufriadis, A., Shoemark, A., Schmidts, M., Patel, M., Jimenez, G., *et al.*, 2014. **Targeted NGS gene panel identifies mutations in RSPH1 causing primary ciliary dyskinesia and a common mechanism for ciliary central-pair agenesis due to radial spoke defects**. *Human molecular genetics*, ddu046. https://doi.org/10.1093/hmg/ddu046

Ostrowski, L.E., Blackburn, K., Radde, K.M., Moyer, M.B., Schlatzer, D.M., *et al.*, 2002. **A Proteomic Analysis of Human Cilia Identification of Novel Components**. *Molecular & Cellular Proteomics* 1, 451-465. https://doi.org/10.1074/mcp.M200037-MCP200

Ostrowski, L.E., Yin, W., Patel, M., Sechelski, J., Rogers, T., *et al.*, 2014. **Restoring ciliary function to differentiated primary ciliary dyskinesia cells with a lentiviral vector**. *Gene Therapy* 21, 253. https://doi.org/10.1038/gt.2013.79

Ott, J., 1974. **Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies.** *American journal of human genetics* 26, 588.

Paff, T., Kooi, I.E., Moutaouakil, Y., Riesebos, E., Sistermans, E.A., *et al.*, 2018. **Diagnostic yield of a targeted gene panel in primary ciliary dyskinesia patients**. *Human mutation* 39, 653-665. https://doi.org/10.1002/humu.23403.

Paff, T., van der Schee, M.P., Daniels, J.M.A., Pals, G., Postmus, P.E., *et al.*, 2013. **Exhaled molecular profiles in the assessment of cystic fibrosis and primary ciliary dyskinesia**. *Journal of Cystic Fibrosis* 12, 454-460. https://doi.org/10.1016/j.jcf.2012.12.010.

Pagani, F., Baralle, F.E., 2004. **Genomic variants in exons and introns: identifying the splicing spoilers**. *Nature Reviews Genetics* 5, 389. 10.1038/nrg1327

Panizzi, J.R., Becker-heck, A., Castleman, V.H., Al-mutairi, D., Liu, Y., *et al.*, 2012. **CCDC103 mutations cause primary ciliary dyskinesia by disrupting assembly of ciliary dynein arms**. *Nature genetics* 44, 714-719. https://doi.org/10.1038/ng.2277

Park, P.J., 2009. **ChIP-seq: advantages and challenges of a maturing technology.** *Nature reviews. Genetics* 10, 669-680. https://doi.org/10.1038/nrg2641

Pazour, G.J., Bloodgood, R.A., 2008. **Targeting Proteins to the Ciliary Membrane**, *Current Topics in Developmental Biology*. Academic Press, pp. 115-149.https://doi.org/10.1016/S0070-2153(08)00805-3

Pazour, G.J., Wilkerson, C.G., Witman, G.B., 1998. **A dynein light chain is essential for the retrograde particle movement of intraflagellar transport (IFT)**. *The Journal of cell biology* 141, 979-992. https://doi.org/10.1083/jcb.141.4.979

Pedersen, L.B., Rosenbaum, J.L., 2008. **Intraflagellar Transport (IFT): Role in Ciliary Assembly, Resorption and Signalling**, in: Yoder, B.K. (Ed.), *Current Topics in Developmental Biology: Ciliary Function in Mammalian Development*. Academic Press, pp. 23-61.https://doi.org/10.1016/S0070-2153(08)00802-8

Pennarun, G., Escudier, E., Chapelin, C., Bridoux, a.M., Cacheux, V., *et al.*, 1999. **Loss-of-function mutations in a human gene related to Chlamydomonas reinhardtii dynein IC78 result in primary ciliary dyskinesia.** *American journal of human genetics* 65, 1508-1519. https://doi.org/10.1086/302683

Pereira, R., Barbosa, T., Gales, L., Oliveira, E., Oliveira, J., *et al.*, 2019. Clinical and genetic analysis of children with Kartagener syndrome European Journal of Medical Genetics.

Pereira, R., Oliveira, J., Ferraz, L., Barros, A., Santos, R., *et al.*, 2015. **Mutation analysis in patients with total sperm immotility**. *Journal of assisted reproduction and genetics* 32, 893-902. https://doi.org/10.1007/s10815-015-0474-6

Pereira, R., Oliveira, J., Sousa, M., 2014. **A molecular approach to sperm immotility in humans: A review**. *Medicina Reproductiva y Embriología Clínica* 01, 15-25. https://doi.org/10.1016/S2340-9320(15)30004-9

Pereira, R., Sá, R., Barros, A., Sousa, M., 2017. **Major regulatory mechanisms involved in sperm motility**. *Asian journal of andrology* 19, 5-14. https://doi.org/10.4103/1008-682X.167716.

Pérez-Fígares, J.M., Jimenez, A.J., Rodríguez, E.M., 2001. **Subcommissural organ, cerebrospinal fluid circulation, and hydrocephalus**. *Microscopy Research and Technique* 52, 591-607. https://doi.org/10.1002/1097-0029(20010301)52:5<591::AID-JEMT1043>3.0.CO;2-7

Pfaffl, M.W., 2001. **A new mathematical model for relative quantification in real-time RT-PCR**. *Nucleic Acids Research* 29, 2002-2007. http://dx.doi.org/10.1093/nar/29.9.e45

Pifferi, M., Michelucci, A., Conidi, M.E., Cangiotti, A.M., Simi, P., *et al.*, 2010. **New DNAH11 mutations in primary ciliary dyskinesia with normal axonemal ultrastructure**. *European Respiratory Journal* 35, 1413-1416. https://doi.org/10.1183/09031936.00186209.

Piperno, G., Mead, K., LeDizet, M., Moscatelli, A., 1994. **Mutations in the" dynein regulatory complex" alter the ATP-insensitive binding sites for inner arm dyneins in Chlamydomonas axonemes.** *The Journal of cell biology* 125, 1109-1117. https://doi.org/10.1083/jcb.125.5.1109

Piperno, G., Mead, K., Shestak, W., 1992. **The inner dynein arms I2 interact with a" dynein regulatory complex" in Chlamydomonas flagella.** *The Journal of cell biology* 118, 1455-1463. https://doi.org/10.1083/jcb.118.6.1455

Pippucci, T., Benelli, M., Magi, A., Martelli, P.L., Magini, P., *et al.*, 2011. **EX-HOM (EXome HOMozygosity): A Proof of Principle**. *Human Heredity* 72, 45-53. 10.1159/000330164

Pongor, L., Kormos, M., Hatzis, C., Pusztai, L., Szabó, A., *et al.*, 2015. **A genome-wide approach to link genotype to clinical outcome by utilizing next generation sequencing and gene chip data of 6,697 breast cancer patients**. *Genome medicine* 7, 104-115. http://doi.org/10.1186/s13073-015-0228-1

Porter, K.R., Kallman, F., 1953. **The properties and effects of osmium tetroxide as a tissue fixative with special reference to its use for electron microscopy**. *Experimental Cell Research* 4, 127-141. https://doi.org/10.1016/0014-4827(53)90195-5

Porter, M.E., 2018. **Ciliary and flagellar motility and the nexin-dynein regulatory complex**, in: King, S.M. (Ed.), *Dyneins* 2ed. Academic Press, pp. 298-335.https://doi.org/10.1016/B978-0-12-809471-6.00010-3

Porter, M.E., Sale, W.S., 2000. **The 9 + 2 axoneme anchors multiple inner arm dyneins and a network of kinases and phosphatases that control motility.** *The Journal of cell biology* 151, F37-42. https://doi.org/10.1083/jcb.151.5.F37

Pulst, S.M., GM, L., Available, R.N., EM, S., VC, S., 1999. **Genetic Linkage Analysis**. *Archives of Neurology* 56, 667. 10.1001/archneur.56.6.667

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., *et al.*, 2007. **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *American journal of human genetics* 81, 559-575. 10.1086/519795

Qi, H., Moran, M.M., Navarro, B., Chong, J.A., Krapivinsky, G., *et al.*, 2007. **All four CatSper ion channel proteins are required for male fertility and sperm cell hyperactivated motility**. *Proceedings of the National Academy of Sciences* 104 1219-1223. http://doi.org/10.1073/pnas.0610286104

Raidt, J., Wallmeier, J., Hjeij, R., Onnebrink, J.G., Pennekamp, P., *et al.*, 2014. **Ciliary beat pattern and frequency in genetic variants of primary ciliary dyskinesia**. *European Respiratory Journal*, erj00520-02014. https://doi.org/10.1183/09031936.00052014.

Raidt, J., Werner, C., Menchen, T., Dougherty, G.W., Olbrich, H., *et al.*, 2015. **Ciliary function and motor protein composition of human fallopian tubes**. *Human Reproduction* 30, 2871-2880. https://doi.org/10.1093/humrep/dev227

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., *et al.*, 2006. **Global variation in copy number in the human genome**. *Nature* 444, 444-454. https://doi.org/10.1038/nature05329

Redwine, W.B., Hernandez-Lopez, R., Zou, S., Huang, J., Reck-Peterson, S.L., *et al.*, 2012. **Structural Basis for Microtubule Binding and Release by Dynein**. *Science* 337, 1532-1536. https://doi.org/10.1126/science.1224151

Reiter, J.F., Blacque, O.E., Leroux, M.R., 2012. **The base of the cilium: roles for transition fibres and the transition zone in ciliary formation, maintenance and compartmentalization**. *EMBO reports* 13, 608-618. https://doi.org/10.1038/embor.2012.73

Ribeiro, C.M.P., Paradiso, A.M., Livraghi, A., Boucher, R.C., 2003. **The Mitochondrial Barriers Segregate Agonist-induced Calcium-dependent Functions in Human Airway Epithelia**. *The Journal of General Physiology* 122, 377-387. http://doi.org/10.1085/jgp.200308893

Roberts, A.J., Kon, T., Knight, P.J., Sutoh, K., Burgess, S.A., 2013. **Functions and mechanics of dynein motor proteins**. *Nature Reviews Molecular Cell Biology* 14, 713-726. https://doi.org/10.1038/nrm3667

Roberts, A.J., Numata, N., Walker, M.L., Kato, Y.S., Malkova, B., *et al.*, 2009. **AAA+ Ring and Linker Swing Mechanism in the Dynein Motor**. *Cell* 136, 485-495. https://doi.org/10.1016/j.cell.2008.11.049

Rohatgi, R., Snell, W.J., 2010. **The ciliary membrane**. *Current Opinion in Cell Biology* 22, 541-546. https://doi.org/10.1016/j.ceb.2010.03.010

Rosenbaum, J.L., Witman, G.B., 2002. **Intraflagellar transport**. *Nature Reviews Molecular Cell Biology* 3, 813. http://doi.org/10.1038/nrm952

Roubin, R., Acquaviva, C., Chevrier, V., Sedjaï, F., Zyss, D., *et al.*, 2013. **Myomegalin is necessary for the formation of centrosomal and Golgi-derived microtubules**. *Biology Open* 2, 238-250. http://doi.org/10.1242/bio.20123392

Rubbo, B., Lucas, J.S., 2017. **Clinical care for primary ciliary dyskinesia: current challenges and future directions.** *European respiratory review : an official journal of the European Respiratory Society* 26, 170023-170034. https://doi.org/10.1183/16000617.0023-2017

Rubin, B.K., 2010. **Mucus and Mucins**. *Otolaryngologic Clinics of North America* 43, 27-34. https://doi.org/10.1016/j.otc.2009.11.002

Rutland, J., Dewar, A., Cox, T., Cole, P., 1982. **Nasal brushing for the study of ciliary ultrastructure.** *Journal of Clinical Pathology* 35, 357-359. http://dx.doi.org/10.1136/jcp.35.3.357

Sabila, M., Kundu, N., Deana, S., Ullah, H., 2016. **Tyrosine Phosphorylation Based Homo-dimerization of Arabidopsis RACK1A Proteins Regulates Oxidative Stress Signaling Pathways in Yeast**. *Frontiers in Plant Science* 7, 176. http://dx.doi.org/10.3389/fpls.2016.00176

Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., *et al.*, 1985. **Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia**. *Science* 230, 1350-1354. https://doi.org/10.1126/science.2999980

Salathe, M., 2007. **Regulation of mammalian ciliary beating**. *Annual Review of Physiology* 69, 401-422. https://doi.org/10.1146/annurev.physiol.69.040705.141253

Sanger, F., Coulson, A.R., 1975. **A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase**. *Journal of Molecular Biology* 94, 441-448. https://doi.org/10.1016/0022-2836(75)90213-2

Sanger, F., Nicklen, S., Coulson, A.R., 1977. **DNA sequencing with chain-terminating inhibitors**. *Proceedings of the national academy of sciences* 74, 5463-5467.

Sapiro, R., Kostetskii, I., Olds-Clarke, P., Gerton, G.L., Radice, G.L., *et al.*, 2002. **Male infertility, impaired sperm motility, and hydrocephalus in mice deficient in sperm-associated antigen 6**. *Molecular and cellular biology* 22, 6298-6305. https://doi.org/10.1128/MCB.22.17.6298-6305.2002

Satir, P., 1995. **Landmarks in cilia research from leeuwenhoek to US**. *Cell Motility and the Cytoskeleton* 32, 90-94. https://doi.org/10.1002/cm.970320203

Satir, P., 2017. **CILIA: before and after**. *Cilia* 6, 1-11. https://doi.org/10.1186/s13630-017-0046-8

Sawamoto, K., Wichterle, H., Gonzalez-Perez, O., Cholfin, J.A., Yamada, M., *et al.*, 2006. **New Neurons Follow the Flow of Cerebrospinal Fluid in the Adult Brain**. *Science* 311, 629-632. http://doi.org/10.1126/science.1119133

Schmid, A., Salathe, M., 2011. **Ciliary beat co-ordination by calcium**. *Biology of the Cell* 103, 159-169. https://doi.org/10.1042/BC20100120

Schmidt, H., Carter, A.P., 2018. **Mechanism and regulation of dynein motors**, in: King, S.M. (Ed.), *Dyneins*, Second Edition ed. Academic Press, pp. 36-51.https://doi.org/10.1016/B978-0-12-809470-9.00002-3

Schmidt, H., Gleave, E.S., Carter, A.P., 2012. **Insights into dynein motor domain function from a 3.3-Å crystal structure**. *Nature Structural and Molecular Biology* 19, 492. https://doi.org/10.1038/nsmb.2272

Schmidt, H., Zalyte, R., Urnavicius, L., Carter, A.P., 2015. **Structure of human cytoplasmic dynein-2 primed for its power stroke**. *Nature* 518, 435. https://doi.org/10.1038/nature14023

Schwabe, G.C., Hoffmann, K., Loges, N.T., Birker, D., Rossier, C., *et al.*, 2008. **Primary ciliary dyskinesia associated with normal axoneme ultrastructure is caused by DNAH11 mutations**. *Human mutation* 29, 289-298. https://doi.org/10.1002/humu.20656

Sears, P.R., Yin, W.-N., Ostrowski, L.E., 2015. **Continuous mucociliary transport by primary human airway epithelial cells in vitro**. *American journal of physiology. Lung cellular and molecular physiology* 309, L99-L108. http://doi.org/10.1152/ajplung.00024.2015

Seelow, D., Schuelke, M., 2012. **HomozygosityMapper2012—bridging the gap between homozygosity mapping and deep sequencing**. *Nucleic acids research* 40, W516-W520.

Seelow, D., Schuelke, M., Hildebrandt, F., Nürnberg, P., 2009. **HomozygosityMapper--an interactive approach to homozygosity mapping.** *Nucleic acids research* 37, W593-W599. https://doi.org/10.1093/nar/gkp369

Seelow, D., Schwarz, J.M., Schuelke, M., Blatter, M., Estreicher, A., 2008. **GeneDistiller—Distilling Candidate Genes from Linkage Intervals**. *PLoS ONE* 3, e3874. 10.1371/journal.pone.0003874

Shah, A.S., Ben-Shahar, Y., Moninger, T.O., Kline, J.N., Welsh, M.J., 2009. **Motile Cilia of Human Airway Epithelia Are Chemosensory**. *Science* 325, 1131-1134. http://doi.org/10.1126/science.1173869

Shingyoji, C., 2012. **Regulation of Dynein in Ciliary and Flagellar Movement**, in: King, S.M. (Ed.), *Dyneins*, 1 ed. Academic Press, Boston, pp. 366-393.https://doi.org/10.1016/B978-0-12-382004-4.10013-5

Shinohara, K., Hamada, H., 2017. **Cilia in Left–Right Symmetry Breaking**. *Cold Spring Harbor Perspectives in Biology* 9, 1-9. https://doi.org/10.1101/cshperspect.a028282

Shoemark, A., Moya, E., Hirst, R.A., Patel, M.P., Robson, E., *et al.*, 2017. **A high prevalence CCDC103 p.His154Pro mutation causing primary ciliary dyskinesia is associated with normal diagnostic investigations**. *European Respiratory Journal* 50, PA1851. https://doi.org/10.1183/1393003.congress-2017.PA1851

Silvanovich, A., Li, M.-g., Serr, M., Mische, S., Hays, T.S., 2003. **The third P-loop domain in cytoplasmic dynein heavy chain is essential for dynein motor function and ATP-sensitive microtubule binding**. *Molecular biology of the cell* 14, 1355-1365.

Siyahhan, B., Knobloch, V., de Zélicourt, D., Asgari, M., Schmid Daners, M., *et al.*, 2014. **Flow induced by ependymal cilia dominates near-wall cerebrospinal fluid dynamics in the lateral ventricles**. *Journal of The Royal Society Interface* 11, 20131189. https://doi.org/10.1098/rsif.2013.1189

Skipor, J., Thiery, J.-C., 2008. **The choroid plexus–cerebrospinal fluid system: undervaluated pathway of neuroendocrine signaling into the brain**. *Acta Neurobiol Exp (Wars)* 68, 414-428. https://ane.pl/archive?vol=68&no=3&id=6843

Sloboda, R.D., 2005. **Intraflagellar transport and the flagellar tip complex**. *Journal of Cellular Biochemistry* 94, 266-272. https://doi.org/10.1002/jcb.20323

Smith, E.F., 2002. **Regulation of flagellar dynein by calcium and a role for an axonemal calmodulin and calmodulin-dependent kinase**. *Molecular biology of the cell* 13, 3303-3313. https://doi.org/10.1091/mbc.e02-04-0185

Smith, E.F., Lefebvre, P.A., 1997. **The role of central apparatus components in flagellar motility and microtubule assembly**. *Cell motility and the cytoskeleton* 38, 1-8. https://doi.org/10.1002/(SICI)1097-0169(1997)38:1<1::AID-CM1>3.0.CO;2-C

Smith, E.F., Yang, P., 2004. **The radial spokes and central apparatus: Mechano-chemical transducers that regulate flagellar motility**. *Cell Motility and the Cytoskeleton* 57, 8-17. https://doi.org/10.1002/cm.10155

Smith, K.G., Kamdar, A.A., Stark, J.M., 2019. **Lung Defenses: Intrinsic, Innate, and Adaptive**, in: Wilmott, R.W., Deterding, R., Li, A., Ratjen, F., Sly, P., Zar, H.J., Bush, A. (Eds.), *Kendig's Disorders of the Respiratory Tract in Children*, 9 ed. Content Repository Only!, Philadelphia, pp. 120-133.e122.https://doi.org/10.1016/B978-0-323-44887-1.00008-0

Sobel, E., Lange, K., 1996. **Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics.** *American journal of human genetics* 58, 1323-1337.

Sousa, M., Oliveira, E., Alves, Â., Gouveia, M., Figueiredo, H., *et al.*, 2015. **Ultrastructural analysis of five patients with total sperm immotility**. *Zygote* 23, 1-8. https://doi.org/10.1017/S0967199414000616

Spoon, C., Grant, W., 2011. **Biomechanics of hair cell kinocilia: experimental measurement of kinocilium shaft stiffness and base rotational stiffness with Euler-Bernoulli and Timoshenko beam analysis**. *The Journal of experimental biology* 214, 862-870. https://doi.org/10.1242/jeb.051151

Stankiewicz, P., Lupski, J.R., 2010. **Structural Variation in the Human Genome and its Role in Disease**. *Annual Review of Medicine* 61, 437-455. http://doi.org/10.1146/annurev-med-100708-204735

Suarez, S.S., 2008. **Control of hyperactivation in sperm**. *Human reproduction update* 14, 647-657. https://doi.org/10.1093/humupd/dmn029.

Suarez, S.S., 2010. **How do sperm get to the egg? Bioengineering expertise needed!** *Experimental mechanics* 50, 1267-1274. https://doi.org/10.1007/s11340-009-9312-z

Suarez, S.S., Pacey, A.A., 2006. **Sperm transport in the female reproductive tract**. *Human Reproduction Update* 12, 23-37. https://doi.org/10.1093/humupd/dmi047

Sui, W., Hou, X., Che, W., Ou, M., Sun, G., *et al.*, 2016. **CCDC40 mutation as a cause of primary ciliary dyskinesia: a case report and review of literature**. *The clinical respiratory journal* 10, 614-621. https://doi.org/10.1111/crj.12268

Sutto, Z., Conner, G.E., Salathe, M., 2004. **Regulation of human airway ciliary beat frequency by intracellular pH**. *The Journal of Physiology* 560, 519-532. http://doi.org/10.1113/jphysiol.2004.068171

Sweeney, H.L., Holzbaur, E.L.F., 2018. **Motor Proteins**. *Cold Spring Harbor Perspectives in Biology* 10, a021931. http://doi.org/10.1101/cshperspect.a021931

Syvänen, A.-C., 2005. **Toward genome-wide SNP genotyping**. *Nature genetics* 37, S5.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., *et al.*, 2019. **STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets**. *Nucleic Acids Research* 47, D607-D613. https://doi.org/10.1093/nar/gky1131

Takada, S., Wilkerson, C.G., Wakabayashi, K.-i., Kamiya, R., Witman, G.B., 2002. **The Outer Dynein Arm-Docking Complex : Composition and Characterization of a Subunit (Oda1) Necessary for Outer Arm Assembly**. *Molecular Biology of the Cell* 13, 1015-1029. https://doi.org/10.1091/mbc.01

Talbot, P., Shur, B.D., Myles, D.G., 2003. **Cell Adhesion and Fertilization: Steps in Oocyte Transport, Sperm-Zona Pellucida Interactions, and Sperm-Egg Fusion1**. *Biology of Reproduction* 68, 1-9. https://doi.org/10.1095/biolreprod.102.007856

Tang, R., Feng, T., Sha, Q., Zhang, S., 2009. **A variable-sized sliding-window approach for genetic association studies via principal component analysis.** *Annals of human genetics* 73, 631-637. https://doi.org/10.1111/j.1469-1809.2009.00543.x

Tantisira, K.G., Lake, S., Silverman, E.S., Palmer, L.J., Lazarus, R., *et al.*, 2004. **Corticosteroid pharmacogenetics: association of sequence variants in CRHR1 with improved lung function in asthmatics treated with inhaled corticosteroids**. *Human Molecular Genetics* 13, 1353-1359. https://doi.org/10.1093/hmg/ddh149

Tarkar, A., Loges, N.T., Slagle, C.E., Francis, R., Dougherty, G.W., *et al.*, 2013. **DYX1C1 is required for axonemal dynein assembly and ciliary motility**. *Nature genetics* 45, 995-1003. https://doi.org/10.1038/ng.2707

Taschner, M., Lorentzen, E., 2016. **The intraflagellar transport machinery**. *Cold Spring Harbor perspectives in biology* 8, a028092. http://doi.org/10.1101/cshperspect.a028092

Tebani, A., Afonso, C., Marret, S., Bekri, S., 2016. **Omics-Based Strategies in Precision Medicine: Toward a Paradigm Shift in Inborn Errors of Metabolism Investigations**. *International Journal of Molecular Sciences* 17, 1555. https://doi.org/10.3390/ijms17091555

Turner, R.M., 2003. **Tales from the tail: what do we really know about sperm motility?** *Journal of andrology* 24, 790-803. https://doi.org/10.1002/j.1939-4640.2003.tb03123.x

Turner, R.M.O., Musse, M.P., Mandal, A., Klotz, K.E.N., Friederike, C., *et al.*, 2001. **Molecular Genetic Analysis of Two Human Sperm Fibrous**. *Journal of andrology* 22, 302-315.

Van Der Baan, S., 1991. **Primary ciliary dyskinesia and the middle ear**. *The Laryngoscope* 101, 751-754. https://doi.org/10.1288/00005537-199107000-00010

van Ravenzwaaij, D., Cassey, P., Brown, S.D., 2016. **A simple introduction to Markov Chain Monte–Carlo sampling**. *Psychonomic Bulletin & Review*, 1-12. 10.3758/s13423-016-1015-8

Venselaar, H., Ah Te Beek, T., Kuipers, R.K., Hekkelman, M.L., Vriend, G., 2010. **Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces**. *BMC Bioinformatics* 11, 548. https://doi.org/10.1186/1471-2105-11-548

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., *et al.*, 2001. **The sequence of the human genome**. *science* 291, 1304-1351. https://doi.org/10.1126/science.1058040

Vladar, E.K., Stratton, M.B., Saal, M.L., Salazar-De Simone, G., Wang, X., *et al.*, 2018. **Cyclin-dependent kinase control of motile ciliogenesis**. *eLife* 7, e36375. http://doi.org/10.7554/eLife.36375

Walker, W.T., Jackson, C.L., Lackie, P.M., Hogg, C., Lucas, J.S., 2012. **Nitric oxide in primary ciliary dyskinesia**. *European Respiratory Journal* 40, 1024-1032. http://doi.org/10.1183/09031936.00176111

Wang, Z., Gerstein, M., Snyder, M., 2009. **RNA-Seq: a revolutionary tool for transcriptomics.** *Nature reviews. Genetics* 10, 57-63. https://doi.org/10.1038/nrg2484

Watson, J.D., Crick, F.H.C., 1953. **Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid**. *Nature* 171, 737-738. https://doi.org/10.1038/171737a0

Wayne, N., Mishra, P., Bolon, D.N., 2011. **Hsp90 and Client Protein Maturation**, in: Calderwood, S.K., Prince, T.L. (Eds.), *Molecular Chaperones: Methods and Protocols*. Humana Press, Totowa, NJ, pp. 33-44.https://doi.org/10.1007/978-1-61779-295-3_3

Werner, S., Pimenta-Marques, A., Bettencourt-Dias, M., 2017. **Maintaining centrosomes and cilia**. *Journal of Cell Science* 130, 3789-3800. http://dx.doi.org/10.1242/jcs.203505

Wheway, G., Mitchison, H., 2019. **Opportunities and Challenges for Molecular Understanding of Ciliopathies–The 100,000 Genomes Project**. *Frontiers in Genetics* 10. https://doi.org/10.3389/fgene.2019.00127

Wijers, C.D., Chmiel, J.F., Gaston, B.M., 2017. **Bacterial infections in patients with primary ciliary dyskinesia: Comparison with cystic fibrosis**. *Chronic respiratory disease* 14, 392-406. http://doi.org/10.1177/1479972317694621

Wirschell, M., Olbrich, H., Werner, C., Tritschler, D., Bower, R., *et al.*, 2013. **The nexin-dynein regulatory complex subunit DRC1 is essential for motile cilia function in algae and humans**. *Nature genetics* 45, 262-268. https://doi.org/10.1038/ng.2533

Wodehouse, T., Kharitonov, S.A., Mackay, I.S., Barnes, P.J., Wilson, R., *et al.*, 2003. **Nasal nitric oxide measurements for the screening of primary ciliary dyskinesia**. *European Respiratory Journal* 21, 43-47. http://doi.org/10.1183/09031936.03.00305503

Wong, A.P., Bear, C.E., Chin, S., Pasceri, P., Thompson, T.O., *et al.*, 2012. **Directed differentiation of human pluripotent stem cells into mature airway epithelia expressing functional CFTR protein**. *Nature Biotechnology* 30, 876. https://doi.org/10.1038/nbt.2328

Wright, C.V.E., 2001. **Mechanisms of Left-Right Asymmetry: What's Right and What's Left?** *Developmental Cell* 1, 179-186. https://doi.org/10.1016/S1534-5807(01)00036-3

Xu, W., Bull, S.B., Mirea, L., Greenwood, C.M.T., 2012. **Model-free linkage analysis of a binary trait**. *Statistical Human Genetics: Methods and Protocols* 850, 317-345.

Xu, X., Gong, P., Wen, J., 2017. **Clinical and genetic analysis of a family with Kartagener syndrome caused by novel DNAH5 mutations**. *Journal of assisted reproduction and genetics* 34, 275-281.

Xue, Y., Ankala, A., Wilcox, W.R., Hegde, M.R., 2015. **Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing**. *Genetics in Medicine* 17, 444-451. https://doi.org/10.1038/gim.2014.122

Yaghi, A., Dolovich, M.B., 2016. **Airway Epithelial Cell Cilia and Obstructive Lung Disease**. *Cells* 5, 40. http://doi.org/10.3390/cells5040040

Yang, L., Banerjee, S., Cao, J., Bai, X., Peng, Z., *et al.*, 2018. **Compound Heterozygous Variants in the Coiled-Coil Domain Containing 40 Gene in a Chinese Family with Primary Ciliary Dyskinesia Cause Extreme Phenotypic Diversity in Cilia Ultrastructure**. *Frontiers in Genetics* 9, 1-10. https://doi.org/10.3389/fgene.2018.00023

Yang, P., Diener, D.R., Yang, C., Kohno, T., Pazour, G.J., *et al.*, 2006. **Radial spoke proteins of Chlamydomonas flagella.** *Journal of cell science* 119, 1165-1174. https://doi.org/10.1242/jcs.02811

Yeh, S.-D., Chen, Y.-J., Chang, A.C.Y., Ray, R., She, B.-R., *et al.*, 2002. **Isolation and properties of Gas8, a growth arrest-specific gene regulated during male gametogenesis to produce a protein associated with the sperm motility apparatus**. *Journal of Biological Chemistry* 277, 6311-6317.

Yuan, S., Sun, Z., 2013. **Expanding horizons: ciliary proteins reach beyond cilia**. *Annual review of genetics* 47, 353-376. https://doi.org/10.1146/annurev-genet-111212-133243

Zariwala, M., Knowles, M., Leigh, M., 2013a. Primary Ciliary Dyskinesia, GeneReviews™ [Internet]. Seattle (WA): University of Washington, Seattle, NCBI Bookshelf.

Zariwala, Maimoona A., Gee, Heon Y., Kurkowiak, M., Al-Mutairi, Dalal A., Leigh, Margaret W., *et al.*, 2013b. **ZMYND10 Is Mutated in Primary Ciliary Dyskinesia and**

Interacts with LRRC6. *The American Journal of Human Genetics* 93, 336-345. http://dx.doi.org/10.1016/j.ajhg.2013.06.007

Zariwala, M.A., Leigh, M.W., Ceppa, F., Kennedy, M.P., Noone, P.G., *et al.*, 2006. **Mutations of DNAI1 in primary ciliary dyskinesia - Evidence of founder effect in a common mutation**. *AMERICAN JOURNAL OF RESPIRATORY AND CRITICAL CARE MEDICINE* 174, 858-866. https://doi.org/10.1164/rccm.200603-370OC

Zhang, F., Lupski, J.R., 2015. **Non-coding genetic variants in human disease**. *Human Molecular Genetics* 24, R102-R110. http://doi.org/10.1093/hmg/ddv259

Zhang, K., Foster, H.E., Rondelet, A., Lacey, S.E., Bahi-Buisson, N., *et al.*, 2017. **Cryo-EM reveals how human cytoplasmic dynein is auto-inhibited and activated**. *Cell* 169, 1303-1314. https://doi.org/10.1016/j.cell.2017.05.025

Zhang, Y.J., O'Neal, W.K., Randell, S.H., Blackburn, K., Moyer, M.B., *et al.*, 2002. **Identification of Dynein Heavy Chain 7 as an Inner Arm Component of Human Cilia That Is Synthesized but Not Assembled in a Case of Primary Ciliary Dyskinesia**. *Journal of Biological Chemistry* 277, 17906-17915. 10.1074/jbc.M200348200

Zielenski, J., Rozmahel, R., Bozon, D., Kerem, B.-s., Grzelczak, Z., *et al.*, 1991. **Genomic DNA sequence of the cystic fibrosis transmembrane conductance regulator (CFTR) gene**. *Genomics* 10, 214-228. https://doi.org/10.1016/0888-7543(91)90503-7

Zuccarello, D., Ferlin, A., Cazzadore, C., Pepe, A., Garolla, A., *et al.*, 2008. **Mutations in dynein genes in patients affected by isolated non-syndromic asthenozoospermia**. *Human reproduction* 23, 1957-1962. https://doi.org/10.1093/humrep/den193

Zukas, R., Chang, A.J., Rice, M., Springer, A.L., 2012. **Structural analysis of flagellar axonemes from inner arm dynein knockdown strains of Trypanosoma brucei**. *Biocell* 36, 133-142.

Attachments

Attachment 1

Table A1-1. New variants added to ClinVar database resulted from the works performed during this thesis.

Variant information	Gene(s)	Clinical Condition	Proposed clinical significance
NM_001258398.1 c.568_569dup p.Glu187Argfs*22	*CCDC103*	Kartagener syndrome	Pathogenic
NM_001258398.1 c.104G>C; p.Arg35Pro	*CCDC103*	Kartagener syndrome	Uncertain significance
NM_181426.1 c.2540A>G; p.Glu847Gly	*CCDC39*	Fibrous Sheath Dysplasia	Uncertain significance
NM_017950.3 c.1989+1G>A	*CCDC40*	Kartagener syndrome	Likely pathogenic
NM_017950.3: c.2824_2825insCTGT p.Arg942Thrfs	*CCDC40*	Kartagener syndrome	Pathogenic
NM_017950.3 c.2620-92C>T	*CCDC40*	Kartagener syndrome	Uncertain significance
NM_000397.3 c.1103C>T; p.Ala368Val	*CYBB*	Kartagener syndrome	Uncertain significance
NM_001369.2 c.4530delG; p.Asn1511Metfs	*DNAH5*	Kartagener syndrome	Likely pathogenic
NM_001369.2 c.6000C>A; p.Tyr2000Ter	*DNAH5*	Kartagener syndrome	Likely pathogenic
NM_001369.2: c.11570+124G>C	*DNAH5*	Kartagener syndrome	Likely benign
NM_001369.2 c.5882+133A>G	*DNAH5*	Kartagener syndrome	Uncertain significance
NM_001370.1 c.3167A>T; p.Asp1056Val	*DNAH6*	Kartagener syndrome	Uncertain significance
NM_018897.3 c.11947C>T; p.Arg3983Trp	*DNAH7*	Kartagener syndrome	Pathogenic
NM_018897.3 c.8209G>A; p.Gly2737Ser	*DNAH7*	Kartagener syndrome	Pathogenic
NM_001481.2 c.828C>G; p.Asn276Lys	*GAS8*	Kartagener syndrome	Uncertain significance
NM_007179.2 c.262_263delCC p.Pro88Glyfs	*INSL6*	Kartagener syndrome	Uncertain significance
NM_206996.3: c.4445G>T p.Arg1482Leu	*SPAG17*	Kartagener syndrome	Uncertain significance
NM_004651.3 c.1599G>A; p.p(=)	*USP11*	Kartagener syndrome	Likely pathogenic

Attachment 2

This attachment contains all the primers that were used in this thesis.

In this attachment are also listed the PCR conditions that were applied to each amplification:

1) **Sd-38c**, with an initial denaturation at 95°C for 5min, followed by thirty-eight cycles of 95°C for 45 s, the specific primer annealing temperature (AT) for 30 s, and 72°C for 1 min, with a final extension 72°C for 10 min;
2) **Stand-35c**, with an initial denaturation at 95°C for 10min, followed by thirty-five cycles of 95°C for 1min, the specific primer AT for 30 s, and 72°C for 2 min, with a final extension 72°C for 10 min;
3) **High-GC**, a PCR with an initial denaturation at 95°C for 5min, followed by forty cycles of 95°C for 45 s, the specific primer AT for 30 s, and 72°C for 1 min, with a final extension 72°C for 10 min.

The PCR the reaction mixture for the two first conditions (30 µl) contained: 15µl of PCR Master Mix [(produced by Promega, Madison, USA) that includes 50 units/ml of *Taq* DNA polymerase supplied in a proprietary reaction buffer (pH 8.5), 400 µM dATPs, 400 µM dTTPs , 400 µM dGTPs, 400 µM dCTPs and 3mM $MgCl_2$]; 12 µl of sterile bidistilled water, 1 µl of each primer (10 pmol/µl each- produced from Thermo Fisher Scientific (Einsteinstrasse, Germany) or Nzytech(Lisboa, Portugal)) and 1 µl of DNA (at about 50 ng/µl) .

For the High-GC PCR condition, the reaction mixture (30µl) contained: 12,5 µl of PCR Master Mix (Promega), 1,5 µl of sterile bidistilled water, 5 µl of betaine 5M (Bioline), 1,5 µl of DMSO (Bioline), 1 µl of each primer and 2,5 µl of DNA (at about 50 ng/µl).

To amplify difficult genomic regions, were used alternatively an **EmeraldAmp** Max HS PCR Master Mix (from Takara Bio, distributed by ENZIfarma S.A, Lisboa, Portugal), a **ImmoMix™** Red (from Bioline), or **AmpliTaq Gold® PCR Master Mix** (Thermo Fisher Scientific).

1. Primers used during gene by gene approach with Sanger Sequencing

Table A2-1 List of primers used to amplify the gene CCDC39 (NM_181426.1), and the PCR conditions.

Exon	Primer Forward	Primer Reverse	PCR Condition/MM
1	CTCAACCGGAAGTTTCGCC	TGAAATAGGCAGAGAGGGTAAGGG	Stand-38c;Ta=58ºC PCR MM
2	AGCCTTCAGAACTGTGAGGAATTT	CAGCCTAGGAAACAGAGCGAGA	Stand-35c;Ta=63ºC PCR MM
3	TTTCTTACAACTAGTGAGCAACACC	CCATGTGCCTTTCACTGTCC	Stand-38c;Ta=58ºC PCR MM
4	GCATATTATAAACATCTGGAAGGC	GAAACAGTCAGCAATATTTTATGGG	Stand-38c;Ta=55ºC PCR MM
5	ATTGAGTTTTTATGCAGCTATATGC	CAATTCTAACTGTCAAACAGAGAGC	Stand-38c;Ta=55ºC Emerald
6	CTTACAGAGACTATAAGCGCACAGG	ATCTACTGACTCTTTCTTGTTCCGC	Stand-38c;Ta=55ºC PCR MM
7.1	TCTACGACCAGTGAAAGGCC	TGACATACCTCACCCTTCAGC	Stand-35c;Ta=55ºC Emerald
7.2	CAAGTTTTTGGAAAGTGAGATTGG	AATGAAAAGGTAACCTACAGAATGG	Stand-35c;Ta=55ºC Emerald
8	GGTTAGAACTTAAAGTGGGCAGAA	GATAGAAAACGTGCCTAAAAAGC	Stand-35c;Ta=63ºC PCR MM
9	TTTTCATCACAAAGACTCAGAAGCC	AGCAATGCATGAGAAGAGGTCC	Stand-38c;Ta=58ºC PCR MM
10	TGTATGACACGGGCGTAAAAGC	CTGTCTGTCCTCCACAAAAGGG	Stand-38c;Ta=58ºC PCR MM
11	CATGGTAGTATGATAAACATCCCCC	TGTGTGTAGAGTGGCAGTCACC	Stand-35c;Ta=61ºC Amplitaq Gold
12	TCCCCAGACAATACTGAGGG	AAATTTCTTTACAAGACTGGAGGG	Stand-38c;Ta=58ºC PCR MM
13	ATGGCCATTCATAGGGCTT	CAGCTAGTTCTCCTGACATCATC	Stand-38c;Ta=58ºC PCR MM
14	AGACACAATGAAACTGTGGCAGG	CCTTTGTTAAGTGTAATAGTGGCCC	Stand-38c;Ta=58ºC PCR MM
15	TTGAAGGCCTGATAATTTGGGC	TCTTCCTCATGTCTCACAGTGTTCC	Stand-38c;Ta=58ºC PCR MM
16	AACAAGAAATGGCTACTACATGCCC	ACTCTGAGCACTTGTTTTTGTCTCC	Stand-38c;Ta=58ºC PCR MM
17	CATATATACAGTCCATCATTGACCG	AATAGATGAAGGAGGATTTGGG	Stand-38c;Ta=55ºC PCR MM
18-19	AAATTAGAAAGAGTGACCAAACAGG	GCGGTGATGTAGAAGTGGC	Stand-38c;Ta=55ºC PCR MM
20	TTTGCCAATAATGAGAGGGTC	TGTAGCCTTGTCAAGAATCTGC	Stand-38c;Ta=58ºC PCR MM

Abbreviations used: Ta-annealing temperature; PCR MM-PCR Master Mix Promega; Emerald-EmeraldAmp Max HS PCR Master Mix; ImoMix- ImmoMix™ Red master mix; AmpliTaq Gold-AmpliTaq Gold® PCR Master Mix

Table A2-2 List of primers used to amplify the gene CCDC40 (NM 017950.3), and the PCR conditions.

Exon	Primer Forward	Primer Reverse	PCR Condition/MM
1	GTGGAGATGGCACAAAGGCC	GAGGACAAAGGGACAGGAGGG	High GC; PCR MM
2	AGGACCAAGAAAAAGAAAGCTG	TTGGGATTACAGGCGTGAG	Stand-38c;Ta=58ºC PCR MM
3	ACAAGGTCCTTTATACTTTGTTTCC	CCTTGCTAAGACATTTCTAACAGG	Stand-38c;Ta=58ºC PCR MM
4	ACTTCTCAGTGGCAATTAAGCCC	GCGTTCTGTCACTTAATCGGACC	Stand-38c;Ta=58ºC PCR MM
5	GCAGCTGTGCCATTGATG	GCATTTAGGGTCAGGAGTGGG	Stand-38c;Ta=58ºC PCR MM
6-7	GCAGGAGGGTAACCAGAAAGG	CTGGCCATGGTACATGGAGAC	Stand-38c;Ta=58ºC PCR MM
8	CCAGGAGGATCTCTGAGACTTGG	CTTCACGCAGGAGGCAAAGC	Stand-38c;Ta=58ºC PCR MM
9	GTGGCGTCAACTTGTATCAAGGG	GCTAAGGAACTTGCTGAGCATCC	Stand-38c;Ta=55ºC PCR MM
10	CTTGAAAGCTACATTCAGGCTCG	CTGAGCAGAAAAGCCTCTCCC	Stand-38c;Ta=58ºC PCR MM
11	TGGCTCTCTGATGTCTCACACG	ATGGCAGCATCCGTCTTCC	Stand-38c;Ta=58ºC PCR MM
12	AGAAGCTCACCACCCAGTGC	CAACCTGAACACCATGGAAAGG	Stand-38c;Ta=58ºC PCR MM
13	TCTGGGTTCCAACAAGTAGGTCC	TAGGGAAAGAGGCAGTTCTGTGC	Stand-38c;Ta=58ºC PCR MM
14	AATTACAGGCGTGAGTCACTGC	CATGCGCGTTTGTTATTTGC	Stand-38c;Ta=55ºC PCR MM
15	TAGCAGCCAAAAAGAGACCCC	TTACATCCTGGTGCAGAACGC	Stand-38c;Ta=58ºC PCR MM
16	GACAGAGAATGAGTTCGTGCGC	TGGTCTTCAACGGACAGAGGC	Stand-38c;Ta=58ºC PCR MM
17	AAACGTTTGCATAAGGAGCCC	CCTGGAAATCTCACCCATTGC	Stand-38c;Ta=58ºC PCR MM
18	CTCCATGTCCCTTGTTCCTCC	TGCAGTCTCCCAGACAGGC	High GC; PCR MM
19	CTGCTGAACTGTGCCAGTCTCC	TGGTGCTATTACTGGCCTGGG	Stand-38c;Ta=58ºC ImoMix
20	CTTTTCAGGCGTAGGTCTCGC	GTCACAGGCTCTGTCTTTGCG	High GC; PCR MM

Abbreviations used: Ta-annealing temperature; PCR MM-PCR Master Mix Promega; Emerald-EmeraldAmp Max HS PCR Master Mix; ImoMix ImmoMix™ Red master mix; AmpliTaq Gold-AmpliTaq Gold® PCR Master Mix.

Table A2-3 List of primers used to amplify the gene *DNAI1* (NM_001281428.1), and the PCR conditions.

Exon	Primer Forward	Primer Reverse	PCR Condition/MM
1	TGGATTCTATCCTGCAAGGGC	GGTAAGAAACGGTGGATTGGG	Stand-38c;Ta=58ºC PCR MM
2	CTCCTCAAATGCCACAAAGTG	TTATAAGGTGGTAGGTTTAAGGGTG	Stand-38c;Ta=55ºC Emerald
3_4	CAGGCAATGAGAACTCTGGCC	TACTGAGAGGCTTTTGCAATGTAGC	Stand-38c;Ta=58ºC PCR MM
5	TTGAGGTTGGAAAGGCAAGG	CAGAAGCTCTGAAAAGTTCTGGC	Stand-38c;Ta=58ºC PCR MM
6_7	CCTCCCTGCCACAGATTGG	AGGCTGAGGTCTGCTAGGGC	High GC; PCR MM
8	CAGAATGGACACAATATGGGC	CAGGGAGGCAAGTAGATGAGG	Stand-38c;Ta=58ºC PCR MM
9	TTGAATACCTAATTCCAACAGGCC	GGGAGGTATGGAATGCTAGAGGC	Stand-38c;Ta=58ºC PCR MM
10	CCCTTTGTGCCACTTGTGAGG	GGGATTATTTTTTCTCCGCTTGC	Stand-38c;Ta=58ºC PCR MM
11	ATGGCCGGTGTATTGTAATGTACC	CAAGGTGGTTAAGTGCAGAAGGG	Stand-38c;Ta=58ºC PCR MM
11_12	GGGTTTGCCATAAAGCGGC	TCTTCCCCATTAGACTGTGAGCC	Stand-38c;Ta=58ºC PCR MM
13	CTTGATTGGTTGTTGGATCTTTCC	ACCCAGATCCTGACCTGATGG	Stand-38c;Ta=58ºC PCR MM
14_15	TTCTGGGAAATGGGCTCCC	GGGTCTACAAACAGCCAAGGG	Stand-35c; Ta=63º Emerald
16	GACCCTGTTTAACTCAATCCCTGC	GCCCTTATTCCTATCACCCAAGC	Stand-38c;Ta=58ºC PCR MM
17_18	GCGATGTGGGTTAAGGACAGG	TGAATGATTCCTCTTTTGGCTGG	Stand-38c;Ta=58ºC PCR MM
19	AAACACAGAGACACAAAATTCCCC	TGGCTTCTCCCTTCTGTCATCC	Stand-38c;Ta=58ºC PCR MM
20	AGCTGGACTCCTTTTTGCAAGC	GGTTTTCTTCTCTTCCTGAGGGC	High GC; PCR MM

Abbreviations used: Ta-annealing temperature; PCR MM-PCR Master Mix Promega; Emerald-EmeraldAmp Max HS PCR Master Mix; ImoMix ImmoMix™ Red master mix; AmpliTaq Gold-AmpliTaq Gold® PCR Master Mix.

Table A2-3 List of primers used to amplify the gene *AKAP3* (NM_001278309.1), and the PCR conditions.

Exon	Primer Forward	Primer Reverse	PCR Condition/MM
1	CCATCTTTTCCCACAATCCCC	GCTCGACGCACATTTATTTACTTGG	Stand-38c;Ta=58ºC PCR MM
2	CCCTGGTGTGAAAAAGTTTGGG	CAGCTGGACAGAAGAACAACGC	Stand-38c;Ta=58ºC PCR MM
3	ACCTTTTTGTCATCTTGGAACTGC	CAAAGAATTGGAAGACTTGGAAGC	Stand-38c;Ta=58ºC PCR MM
4	ATTCCTTCCATTGAGAGCTGAGG	TTTGTTCTATAAGGGCTGATTTCCC	Stand-38c;Ta=58ºC PCR MM
5.1a	TGAGAGGGCTGATGACATTTGT	TGACTAGATTCGTGAGGCGGTT	Stand-35c;Ta=58ºC PCR MM
5.1b	GGGAAACGTCCAACTCAGGAG	CATTCCTGGAACATGCAGAGAC	Stand-35c;Ta=58ºC PCR MM
5.2	GGGAGTTCAGTAGATGAAGTTTCC	CAATGGTTGTGTCTTTCACTTGG	Stand-38c;Ta=58ºC PCR MM
5.3	AAAGAGGTTTCGAGGGCAGG	GGCTTACGCTGTGTGTTGGG	Stand-38c;Ta=58ºC PCR MM
5.4	GGGTGAGCACATTATCAAAGAGGG	CAACGGTCTTTCACACAACTTCC	Stand-38c;Ta=58ºC PCR MM
5.5	GTAGCTCCCGATGAATCTTGCC	ACTTGTCATCTCCCAGCTCTGCC	Stand-38c;Ta=58ºC PCR MM
5.6	GCTGTGTGTCATCATTGCTAAGTCC	GATCTTCAAATCCCACTGTTCCC	Stand-38c;Ta=58ºC PCR MM
6	TTTAAAGTCTGTGTTGCAGTTTCCC	TCTGGTTATTTTTTTCACTCCTGGG	Stand35c;Ta=61ºC; Emerald

Abbreviations used: Ta-annealing temperature; PCR MM-PCR Master Mix Promega; Emerald-EmeraldAmp Max HS PCR Master Mix; ImoMix ImmoMix™ Red master mix; AmpliTaq Gold-AmpliTaq Gold® PCR Master Mix.

Table A2-3 List of primers used to amplify the gene *AKAP4* (NM_003886.2), and the PCR conditions.

Exon	Primer Forward	Primer Reverse	PCR Condition/MM
1	TTAGAGCCCTCCATCTTTGTGC	TTCATTTCCTGCGAGACCTCC	Stand58_38c; PCR MM
2	CCCACCAGTTCCTAGCCTAAACC	TTACATCACACCACACCCTGCC	Stand58_38c; PCR MM
3	TGCTAATGCTAAGATTCTGTCTGCC	TTCAGAACCAGGGATCTTGGG	Stand58_38c; PCR MM
4	TTTGTGAGCATCTTCAGAGTGGG	TCTTACCATGTCTGATCCAGAGTCC	Stand58_38c; PCR MM
5.1	CAAACTCCAAAAGGTCAAATCCC	TTCTCCTTGATTTCCTTATGGGC	Stand58_38c; PCR MM
5.2	TCAGCTCCTCCAGCCAAACC	CAATCAAATCGGACACAATCTCC	Stand58_38c; PCR MM
5.3	AATTCCAGCATCTGTGGTCC	GCCAGTGAATCTGTGGAAGC	Stand58_38c; PCR MM
5.4	CACATTCTCAAAGAGGGCCTAACC	AATTGCCCGTTCGCTTGC	Stand58_38c; PCR MM
5.5	AGCAAAGCAGCTTCCATGTCC	TGAAACTCGAATGATCTTGGGC	Stand58_38c; PCR MM
6	AGATGTGTCAGCAGCCTCTAGTCC	CCAACACTGTATCACCATTCCTGG	Stand58_38c; PCR MM

Abbreviations used: Ta-annealing temperature; PCR MM-PCR Master Mix Promega; Emerald-EmeraldAmp Max HS PCR Master Mix; ImoMix ImmoMix™ Red master mix; AmpliTaq Gold-AmpliTaq Gold® PCR Master Mix.

Table A2-4 List of primers used to amplify the gene *DNAH5* (NM001369.2), and the PCR conditions.

Exon	Primer Forward	Primer Reverse	PCR Condition/MM
3	TCAACAGCTCAACACTTTAGGG	CAGGTCCGTTCTCACAGTAGC	Stand-35c; Ta=55ºC PCR MM
6	GGCTTTGGGTTGAATTGTGAGG	TGACAGTAAAGGAAAACCACATTGG	Stand-35c; Ta=61ºC PCR MM
7	ATTGGAAGCATGGAATACGC	AGCCTCCAAAGTGATGTGAGG	Stand-35c; Ta=58ºC PCR MM
8	TTTAAAGGGCGGTGCTGC	CAATTCAGCTAGTGCAAAAAGG	Stand-35c; Ta=58ºC PCR MM
9	TTAATGCAGATGGAATGGTTAGTC	CATAGGAAAGAAATTCCAAGAGC	Stand-35c; Ta=55ºC PCR MM
11	AATTTTGCTGTCGCCTTCAC	AGCACCATTACACAATTCTCAAAC	Stand-35c; Ta=55ºC PCR MM
12	TGATTTCTTTGAATGCAGTAGGGC	AATCATTGCTACACAGTGACAAACG	Stand-35c; Ta=58ºC PCR MM
13	AGTACCCGTGGAACCTGACG	TGACCCAAATTGCCAAGAGAG	Stand-35c; Ta=55ºC PCR MM
14	TGTCTTCCCAAGTTAATAGTTTTGC	GCTTAGATTCAACCCATCTGCC	Stand-35c; Ta=55ºC Emerald
17	GGATATTTTATTAGCGACAACCAAG	TGATCTGAAATGATGTGTTCCC	Stand-35c; Ta=55ºC Emerald
25	GCACAGTTAACCTCTTGCGG	TCTAAATTACCCAGTCTCGAGCAG	Stand-35c; Ta=61ºC PCR MM
27	CATCTCTGGCTTGCTTGTTTGC	CAATGGACAAGAACAATGCAGC	Stand-35c; Ta=61ºC PCR MM
28	AAATACTAGCAGACCGTCTTGGTCC	GCAGAATCTGTCCCATCTTAGGC	Stand-35c; Ta=61ºC PCR MM
30	AAATAGGCACAGAAGCTCAGTAGGG	GACAACAAAAAGATTGAAGGAAGGG	Stand-35c; Ta=61ºC PCR MM
31	GATGATATGGCACAGAAAGCCC	CAGTGATGGAAGCAACCTTAAGTCC	Stand-35c; Ta=61ºC PCR MM
32	CTCCCCTTTTCCCATCCAGG	CAGTTCTGAAGGTAAATTGGGCC	Stand-35c; Ta=61ºC PCR MM
33	TTACTTTCTTTGGGAAGGACAAG	AGATTGATTAGGGAAAAAACCG	Stand-35c; Ta=58ºC PCR MM
34	AATACAGGAAACAATGAGAAACGTG	TTCTACTGGGTAAATGCAGATAGTG	Stand-35c; Ta=58ºC PCR MM
35	TGAATAGCTTGGCTGGTGATACAG	AGGCAAAGGAAGCAAGGTCTG	Stand-35c; Ta=61ºC PCR MM
36	ATCTTGTGTGCGTTTCATGCC	GCATCAAGTGACCCAAAACAGC	Stand-35c; Ta=61ºC PCR MM
41	GGCCATCATTGTTGTGTTTCAA	CCTCTTGGGCATTCAAATGGG	Stand-35c; Ta=61ºC PCR MM
43	CAAGGTGCTCAATGAATATTTGTCC	GGGTTTGAATGTCCCAGTGC	Stand-35c; Ta=61ºC PCR MM
44	AGGCTACCTACACTTCCTGGGC	TATTCCCCCTCCCCAACAGC	Stand-35c; Ta=61ºC PCR MM

Abbreviations used: Ta-annealing temperature; PCR MM-PCR Master Mix Promega; Emerald-EmeraldAmp Max HS PCR Master Mix; ImoMix ImmoMix™ Red master mix; AmpliTaq Gold-AmpliTaq Gold® PCR Master Mix.

Table A2-4 (continuation)

Exon	Primer Forward	Primer Reverse	PCR Condition/MM
45	CCTATTATGTTTAAAGGCAGCTTGC	CGAACGTTTTCATATCTTACAGCC	Stand-35c; Ta=61ºC PCR MM
48	TTCATATCAGTTCAAAGGTTACCAG	CTAACTCCTTGAGTGTTTCCAAAG	Stand-35c; Ta=58ºC PCR MM
49	TCAAGGAGTTAGGCTCCGGG	TCCAGACCTCAAAACTATGTGCC	Stand-35c; Ta=61ºC PCR MM
50	TGGTGCAACCAGAGTCTTCG	AAAGCAGTAAAGAATACCCATGC	Stand-35c; Ta=61ºC PCR MM
51	CTGTCAACCTGAATTCCCAATTACG	TCTCCTACCGAATCCTGCCC	Stand-35c; Ta=61ºC PCR MM
53	CATTCCTTCAACCAGTGTGTCTC	TATAACATAAAAGTAGGGAGGTGGC	Stand-35c; Ta=58ºC PCR MM
54	ATAACAGGGCATCTCACTCACCC	ATCCCCAATAGCACTTATTCACACC	Stand-35c; Ta=61ºC PCR MM
60	TGGTCGTTCAGTAGTTTCAGTGGC	TCTATGCTCACTCTCCCCATG	Stand-35c; Ta=61ºC PCR MM
61	ATTTTGATTTTGCCTCTTGGCC	CAGCCATCTTGACTGACCATCC	Stand-35c; Ta=61ºC PCR MM
63	CTGTACAATCACACGGATGAACG	CCTGGACCCTGACAGAATAGCC	Stand-35c; Ta=61ºC PCR MM
67	TGTATCCCGTATCTGGTAGGCAGC	AGGGCATCATCACTCACAACG	Stand-35c; Ta=61ºC PCR MM
68	AGATATCGTTGTGAGTGATGATGCC	TTCCATGCCCAGGTATACAGTAGG	Stand-35c; Ta=61ºC PCR MM
72	GTGTATACTGGAACAAATGCATGG	GAGCTCTGATTCTTTCACACTGC	Stand-35c; Ta=61ºC PCR MM
73	TCTTGTTAGATTGGCAGGAAGTTG	ACGCCAAAGCTAGGAGGTCC	Stand-35c; Ta=61ºC PCR MM
75	AATTTAGACGGTGCTTATGTTGGC	GGAAGGTTCAGAAAAGCAGTCC	Stand-35c; Ta=58ºC Emerald
76	GCTGTTTGGAAGACTGCATGG	TGGTCTAAAAATACAGGGCTACAGG	Stand-35c; Ta=61ºC PCR MM
77	GAATAATTTGAAACACTTACGCCAG	GCTAGCAGCCTGAGAACAACTG	Stand-35c; Ta=61ºC PCR MM

Abbreviations used: Ta-annealing temperature; PCR MM-PCR Master Mix Promega; Emerald-EmeraldAmp Max HS PCR Master Mix; ImoMix ImmoMix™ Red master mix; AmpliTaq Gold-AmpliTaq Gold® PCR Master Mix.

2. Primers used for analysis validation by Sanger Sequencing

This part of the attachment contains the primers and PCD conditions used for validation by Sanger sequencing the candidate variants selected from the WES or copy number variation analysis. Here is also included the list of primers used for RT-PCR analysis

Table A2-5. Primer list used to validate WES analysis

Gene SR	Primer Forward	Primer Reverse	PCR Condition/ MM
CCDC103 **Exon 2**	AATGCCTCCTCGACTGAAGC	AAGATCAAACAGTAGACGGTGGG	Stand-35c; Ta=58ºC; PCR MM
CCDC103 **5'UTR**	CCAAAGAAAGCAAGGGGAC	GGTCCTTCCCTGAATAGTGTG	Stand-35c; Ta=60ºC; PCR MM
CCDC103 **Full cDNA-1**	ACCAAGGCCGGGACTGCTGGGG	TCTGTCTGGAAGAGGCAGCCGAGCC	Stand-35c; Ta=60ºC; PCR MM
CCDC103 **Full cDNA-2**	GACTTCTATCGTGATTGGCGAC	AGTTGAGAAGGGAACCAGGTGT	Stand-35c; Ta=62ºC; PCR MM
CCDC103 **3′UTR-1**	GGTTGACTGATGCGGCAAG	TCTAGAACACTGGCGCTCAC	Stand-35c; Ta=62ºC; PCR MM
CCDC103 **3′UTR-2**	CGCCCTTCTCATTCCGGT	CTCACAGGGTACTAAGCAGC	Stand-35c; Ta=64ºC; PCR MM
DNAH10 **Exon 47**	TCCATTGTACGGCTCTGCC	CACATCCTATAACCAACCGGC	Stand-35c; Ta=58ºC; PCR MM
DNAH6 **Exon 20**	GAGAGCAATAAGGAGGAAAATGGGA	ACTGGTCTAGGGAAAATGCTCTG	Stand-35c; Ta=58ºC; PCR MM
GAS8 **Exon 7**	CTGCGACAGGTGGATAGGTG	AGGAGAACACGAGGCAGAGA	Stand-35c; Ta=58ºC; PCR MM
SPAG17 **Exon 31**	AAGAACAGCAGGCCACACTCC	TGGCATGGGATGTAAAAAGGG	Stand-35c; Ta=58ºC; PCR MM
INSL6 **Exon 1**	CGCAGACAGGGAGCAGGGT	GAGCTGTGTACTAGGCACAAATTCC	High GC; Ta=63º; Dream
MYCBPAP Exon 1	CCGCAGGGCTTTCTAGGG	AGTGCGCTGTAGGCTCTTCCT	High GC; Ta=63º; Dream
CCDC40 **Exon 12**	AGAAGCTCACCACCCAGTGC	CAACCTGAACACCATGGAAAGG	
CCDC40 **Exon 17**	AAACGTTTGCATAAGGAGCCC	CCTGGAAATCTCACCCATTGC	
DNAH5 **Exon 28**	AAATACTAGCAGACCGTCTTGGTCC	GCAGAATCTGTCCCATCTTAGGC	Stand-35c; Ta=60ºC; PCR MM
DNAH5 **Exon 36**	ATCTTGTGTGCGTTTCATGCC	GCATCAAGTGACCCAAAACAGC	Stand-35c; Ta=60ºC; PCR MM

Accession numbers: *CCDC103*: NM_213607; *DNAH10*: NM_207437; *DNAH5*: NM_001369.2 *DNAH6*: NM_001370 *GAS8*: NM_001481; *SPAG17*: NM_206996; *INSL6*:NM_007179; *MYCBPAP*:NM_032133; *CCDC40*: *NM_017950.3*

SR: Sequenced region.

Table A2-5 (continuation)

Gene SR	Primer Forward	Primer Reverse	PCR Condition/ MM
***DNAH5* Exon 28**	ACTGTCACCATTTCTACCCAGTGA	AGTTGGCATAATCTTGCCTCCAGT	Stand-35c; Ta=58ºC; PCR MM
***DNAH5* Exon 36**	TCAGTGTGGCAATTCTCTCCTTTG	GCTTTCTCTACTCAGCCAGCCA	Stand-35c; Ta=58ºC; PCR MM
AHNAK Exon 5 (1)	TGCACCTGACATCAACATCG	GAACAGATGCATCCAGGTCT	Stand-35c; Ta=58ºC; PCR MM
AHNAK Exon 5 (2)	CTGCACCAAAGATAGAGGGT	CTCAGGCAAGGACACATC	Stand-35c; Ta=58ºC; PCR MM
AURKC Exon 5	ACTGGTGCAATTTAAAAGGAGG	CCCCAAGTTCGTATCTAGGTTTT	Stand-35c; Ta=58ºC; PCR MM
BTK Exon 14	TTCCTCTAACCAATGAATCCCG	TCCCTCGTCCCAAACCTCTC	Stand-35c; Ta=58ºC; PCR MM
***C19orf47* Exon 5**	TAGAGATAAGCGTCCCACTG	GGCTATGAGTTCCCGGTTAAT	Stand-35c; Ta=58ºC; PCR MM
***C19orf47* Exon 6**	CACGAGTACCATGCAGTTACA	ACGTATGGCACCTTGACC	Stand-35c; Ta=58ºC; PCR MM
***CYBB* Exon 9**	AGTGTGGAAATTGAGATGGCC	AGAACCTGTGTGTTTTGTTGG	Stand-35c; Ta=58ºC; PCR MM
CYBB cDNA	CCCTCCTATGACTTGGAAATGGAT	AGCTTCAGATTGGTGGCGTT	Stand-35c; Ta=58ºC; PCR MM
***USP11* Exon 11**	GTTCCACTGCCTATCAGCCAC	AAGACATCAGCCACCATCATCTAG	Stand-35c; Ta=58ºC; PCR MM
***USP11* cDNA**	AGTCCACGCTGGTGTGCCCCGA	TGCCCCCAGTTGAGGGCCCAGG	Stand-35c; Ta=58ºC; PCR MM

Accession numbers: *CCDC103*: NM_213607;*DNAH10*: NM_207437; *DNAH5*: NM_001369.2 *DNAH6*: NM_001370 DNAH7: *NM_018897.2; GAS8*: NM_001481; *SPAG17*: NM_206996; *INSL6*: NM_007179; *MYCBPAP*:NM_032133; *CCDC40*: *NM_017950.3; AHNAK: NM_001620.2; AURKC NM_001015878.1; BTK* NM_000061.2; C19orf47 NM_001256440.1; CYBB NM_000397.3; *USP11* NM_004651.3

SR: Sequenced region.

Table A2-6. Primer list used to RT-PCR analysis

Gene		Sequence	Annealing temperature
ABCA3 NM_001089	Forward	TCGTGGCCCCTCGGTACAACTGGA	60ºC
	Reverse	CCCAATGAGCTGGGCTCCCATTGCC	
ACTA1 NM_001100.3	Forward	ACAACGTGCCCATTTATGAGG	60ºC
	Reverse	TCAGGTAGTCGGTGAGATCG	
CCDC103 NM_213607	Forward	TGAGTAAGGGGACGGAAGGGCAC	60ºC
	Reverse	GCCCGTAACTTGGCAGCATTCTC	
DNAH5 NM_001369	Forward	GCTGGATGACGGTGCAAAACCT	60ºC
	Reverse	AACCGCTTGGCTTCCTTGGG	
GAPDH NM_001289745	Forward	AGGTCGGAGTCAACGGATTT	60ºC
	Reverse	TGGAATTTGCCATGGGTGGA	
***B2M* NM_004048**	Forward	ATGAGTATGCCTGCCGTGTG	60ºC
	Reverse	CAAACCTCCATGATGCTGCTTAC	
CATSPER2 NM_172095	Forward	GTCGTTTGGACTGGGAGACTCTTGTGC	60ºC
	Reverse	TGGGCCAAACACGGTCATCCTGA	
CRHR1 NM_001145146	Forward	AGCACGCATGTCCCTCCAAGGCTGT	60ºC
	Reverse	TCACGAGTTGCCCATGATGCCCCA	
CYBB NM_000397.3	Forward	*GTCTCAGGCCAATCACTTTGC*	60ºC
	Reverse	CATTATCCCAGTTGGGCCGT	
DMBT1 NM_007329.2	Forward	*TGGCAATGCCCCTCCCTGTGTGA*	60ºC
	Reverse	TGGAGTGTGGAAGGGGTGATTGGGCA	
KRT34 NM_021013	Forward	*GCTGACGGAGAGCGAGGCCCAC*	60ºC
	Reverse	CCGGGCACGCACGTCCAGCA	
PDE4DIP NM_014644.5	Forward	*GGTTCCCCTGGGAAGCACCAACACCA*	60ºC
	Reverse	*TGGGCATCCACTGTGAAGGTAGCCCCA*	
USP11 NM_004651.3	Forward	CCGCAAGCCAGAGCAGCACC	60ºC
	Reverse	TGCGCTCCCGCAGGTAGACA	
DNAH7 NM_018897.2	Forward	GTCTGCCGGTCACTCTTTGA	58ºC
	Reverse	ATCCAGTCCAATGCCACCAG	
EMC7 NM_020154.3	Forward	ATGAGACGGGAAATGGAGCA	58ªC
	Reverse	CCAGTGTTGCCGTGTTTGTG	

Attachment 3

Table A3-1 Compilation of published variants in the *CCDC40* gene.

Coding change	Protein change	Families (number)	Main Ultrastructural anomalies	*Situs inversus*	Method	Ref.
c.940–2A>G		1	IDA defects and Ax disorganisation	?	Sanger Sequencing	[1]
c.961C>T	p.Arg321Ter	2		?		
c.2440C>T	p.Arg814Ter	1		?		
c.687delA	p.Pro229Profs58Ter	2		?		
c.2712–1G>T		4	IDA defects	?	WES	
c.1415delG	p.Arg472fs3Ter	1	IDA defects	?	Sanger Sequencing	
c.1006C>T	p.Gln336Ter	1		?		
c.3175C>T	p.Arg1059Ter	1		?		
c.1464delC	p.Ile488Ilefs19Ter	1	IDA defects and Ax disorganisation	?	Sanger Sequencing	[1,2]
c.1345C>T	p.Arg449Ter	1		Yes		
c.344delC	p.Pro115Argfs52Ter	2		?		
c.3129delC	p.Asp1043Aspfs36Ter	2	IDA defects and Ax disorganisation	?	Sanger Sequencing	[1,3]
c.248delC	p.Ala83ValfsTer84	31	IDA and central pair defects	Yes	WES/ Sanger Sequencing/ Target NGS panel	[1–5]
c.1315C>T	p.Gln439Ter	1	IDA defects and Ax disorganisation	Yes	Sanger Sequencing	[3]
c.1527_1558del	p.Asp510Serfs22Ter	2	IDA defects	Yes		
c.1971C>T	p.Gln651Ter	1		Yes		
c.2440C>T	p.Arg814Ter	1		?		
c.1366C>T		1		Yes		
c.1810C>T	p.Gln604Ter	1		Yes		
c.2824_2825insTGT	p.Arg942MetinsW	2		?		
c.3129delC	p.Phe1044Serfs35Ter	1		Yes		
c.778del	p.Ala83Valfs82Ter	1		?		
c.960C>T	p.Arg321Ter	1		?		
ivs11-2A>G		1		Yes		

Axoneme = Ax; Inner dynein arms = IDA; Next-generation sequencing =NGS; WES = Whole exome sequence; ? = information not provided in the original paper.

Table A3-1 (continued)

Coding change	Protein change	Families (number)	Main Ultrastructural anomalies	*Situs inversus*	Method	Ref.
c.574C>T	p.Gln192Ter	1	IDA defects	?	Sanger Sequencing	[2]
c.2119G>T	p.Glu707Ter	1		Yes		
c.2712-1G>T		1		?		
c.2591_2592delCAinsACCG	p.Thr864AsnfsX10	1		Yes		
c.2920C>T	p.Gln974Ter	1		?		
c.3242_3245dupGGCG	p.Tyr1083AlafsX104	1		Yes		
c.2609G>A	p. Arg870His	1	IDA defects	?	WES / Sanger Sequencing	[6]
c.2527C>T	p.Leu843Phe	1	IDA defects and Ax disorganisation	?	NGS targeted panel	[7]
c.580-1G >C		1		?		
c.1259delA	p.Val421Trpfs*2	1	IDA defects and Ax disorganisation	Yes	NGS targeted panel	[8]
ex17-20del		1		Yes		
c.1416delG	p.Ile473Phefs*2	1	IDA defects	No	WES/ Sanger Sequencing	[4]
c.2441G>A	p.Arg814Ter	1		Yes		

Axoneme = Ax; Inner dynein arms = IDA; Next-generation sequencing = NGS; WES = Whole exome sequence; ? = information not provided in the original paper.

Table A3-2 Compilation of published variants in the *DNAH5* gene.

Coding change	Protein change	Families (number)	Main Ultrastructural anomalies	*Situs inversus*	Method	Ref.
c.8887C>G	p.Gln2949Glu	1	DA defects	?	Sanger Sequencing	[9]
c.8396G>C	p.Arg2799Pro	1	ODA defects	?		
c.10426C>T	p.Gln3462Ter	1	DA defects	?		
c.7429C>T	p.Gln2463Ter	1	ODA defects	?		
c.8440_8447 delGAACCA AA	p.2814fsTer1	1	ODA defects	?		
c.10574G>A	p.Arg3539His	1	ODA defects	?		
c.5172A>C	p.Lys1710Asn	2	DA defects	?		
c.12747G>T	p.Lys4235Asn	1	ODA defects	?		
c.4879C>T	p.Gln1613Ter	1		?		
c.5156+1G>C		1		?		
c.9040C>T	p.Arg3000Ter	1		?		
c.11625C>A	p.Ser3861Arg	1		?		
c.5647C>T	p. Arg1883Ter	1	ODA defects[1]	?	Sanger sequencing	[10]
c.670C>T	p. Arg224Ter	1		?		
c.3876_4053+1 58del	p.Glu1279-Lys1351del	1		?		
c.1089+1G>A		1		?		
c.1108A>T	p.Ile370Phe	2		?		
c.1645A>G	p.Asn549Asp	2		?		
c.1667A>G	p.Asp556Gly	2		?		
c.10365G>C	p.Gln3455His	3		?		
12397G>T	p.Glu4133Ter	2		?		
c.5557A>T	p.Lys1853Ter	1		?		
c.5710-2A>G	p.Cys1904-Lys1909del	1		?		
c.1619T>C	p.Phe540Leu	1		?		
c.10615C>T	p.Arg3539Cys	1		?		
c.8485G>T	p.Val2829Phe	2		?		
c.8497C>G	p.Arg2833Gly	2		?		
c.1645A>G	p.Asn549Asp	1		?		
c.2253C>A	p.Asn751Lys	4		?		
c.5146C>T	p.Arg1716Trp	1		?		
c.7888A>T	p.Arg2630Trp	1		?		
c.5281C>T	p.Arg1761Ter	2	ODA defects[1]	?	Sanger sequencing	[10,11]
c.8029C>T	p.Arg2677Ter	2		?		
c.7502G>C	p.Arg2501Pro	2		?		
c.13486C>T	p.Arg4496Ter	3		?		
c.6791G>A; c.13194_13197 del	p.Ser2264Asn; p.Asp4398Glufs*16	2		?		
c.4348C>T	p.Gln1450Ter	3	ODA defects	?	Sanger sequencing	[10,12]
c.5545G>A	p.Ala1849Thr	1		?	Sanger sequencing	[12]
c.6988+2T>C	p.Met754Ilefs*5	1		?		

Axoneme =Ax; Dynein arms = DA; Outer dynein arms = ODA; Next-generation sequencing = NGS; WES = Whole exome sequence; 1 = ODA defects in all patients, which in some cases are also associated with IDA defects but not individually discriminated; ? = information not provided in the original paper.

Table A3-2 (continued)

Coding change	Protein change	Families (number)	Main Ultrastructural anomalies	*Situs inversus*	Method	Ref.
c.10815delT	p.Pro3606Hisfs*23	9	ODA defects	?	Sanger sequencing	[9,11,12]
c.6037C>T	p.Arg2013Ter	1	?	?	Sanger sequencing	[11]
c.13458_13459insT	p.Asn4487fs*1	4	ODA defects	?		
c.232C>T	p.Arg78Ter	1		?		
c.832delG	p.Ala278ArgfsTer27	1		?		
c.1627C>T	p.Gln543Ter	1		?		
c.ivs17+2T>C		1		?		
c.7914_7915insA	p.Arg2639Thrfs*19	1		?		
c.3905delT	p.Leu1302Argfs*19	1		?		
c.ivs27+1G>A		1		?		
c.5147G>T	p.Arg1716Leu	1		?		
c.5482C>T	p.Gln1828Ter	1		?		
c.1226C>T	p.Trp3409Ser	1		?		
c.5599_5600insC	p.Leu1867Phefs*35	1		?		
c.ivs76+5G>A		2		?		
c.7039G>A	p.Glu2347Lys	1		?		
c.8167C>T	p.Gln2723Ter	1		?		
c.8404C>T	p.Gln2802Ter	1		?		
c.13426C>T	p.Arg4476Ter	1		?		
c.8528T>C	p.Phe2843Ser	1		?		
c.ivs75–2A>T		1		?		
c.9101delG	p.Gly3034Valfs*22	2	ODA shorter	?	WES / Sanger sequencing	[7,13]
c.2261_2262insT	Met754Ilefs*5	1	ODA defects	?	WES / Sanger sequencing	[14]
c.1121T>C	p.Ile374Thr	1		?		
c.3139G>A	p.Gly1047Arg	1		?		
4361G>A; c.8910_8911delATinsG	p.Arg1454Gln; p.Phe2971Serfs*12	2	ODA defect	?	Sanger sequencing	[15,16]
c.11140A>G	p.Ile3714Val	1	ODA defects	?	Target NGS panel	[5]
c.638C>A	p.Pro213Gln	1		?		
c.6710A>G	p.Asn2237Ser	1	ODA defects	?		

Axoneme = Ax; Dynein arms = DA; Outer dynein arms = ODA; Next-generation sequencing = NGS; WES = Whole exome sequence; ? = information not provided in the original paper.

Table A3-2 (continued)

Coding change	Protein change	Families (number)	Main Ultrastructural anomalies	*Situs inversus*	Method	Ref.
c.5563insA	p.1855Asnfs*5	1	ODA defects	?	Sanger sequencing	[16]
c.8440delGAAccAAA	p.2814fs*1	1		?		
c.10555G>C	p.Gly3519Arg	1		?		
c.ivs74-1G>C		1		?		
c.1828C>T	p.Gln610Ter	1		?		
c.5130insA	p.Arg1711Tfs*36	1		?		
c.5367delT	p.Asn1790IlefsX14	1	ODA defects	?	Target NGS panel	[7]
c.9018C>T	p(=) (splicing disturbed)	1		Yes		
c.7550_7556delAGCTGCC	p.Glu2517Glyfs*52	1		Yes		
c.894C>G	p.Asn298Lys	1		No		
c.5563_5564insA	p.Ile1855Asnfs*6	4	ODA defects	?	Target NGS panel/ Sanger sequencing	[7,11]
c.7778C>T	p.Gly2593Glu	1	?	Yes	WES/ Sanger sequencing	[17]
c.13729G>A	p(=)	1	?	Yes		
c.1090-6A>G		1	ODA defects	No	WES / Sanger sequencing	[4]
c.6230T>C	p.Phe2077Ser	1		No		
c.8498G>A	p.Arg2833His	1		No		
c.10384C>T	p.Gln3462Ter	1		Yes		
c.9427A>T	p.Lys3143Ter	1		No		
c.5983C>T	p.Arg1995Ter	3	ODA defects	Yes	WES / Sanger sequencing	[4,7,13]
c.10615C>T	p.Arg3539Cys	2	ODA defects	No	WES / Sanger sequencing	[4,10]
c.6249G>A	p.Met2083Ile	2	ODA defects	Yes	WES / Sanger sequencing	[4,14]

Axoneme = Ax; Dynein arms = DA; Outer dynein arms = ODA; Next-generation sequencing = NGS; WES = Whole exome sequence; ? = information not provided in the original paper.

Attachment References:

1. Antony, D., Becker-Heck, A., Zariwala, M.A., Schmidts, M., Onoufriadis, A., Forouhan M., et al., 2013. Mutations in CCDC39 and CCDC40 are the Major Cause of Primary Ciliary Dyskinesia with Axonemal Disorganization and Absent Inner Dynein Arms. Hum. Mutat. 34, 462-472. https://doi.org/10.1002/humu.22261.
2. Blanchon, S., Legendre, M., Copin, B., Duquesnoy, P., Montantin, G., Kott, E., et al., 2012. Delineation of CCDC39/CCDC40 mutation spectrum and associated phenotypes in primary ciliary dyskinesia. J. Med. Genet. 49, 410-416. https://doi.org/10.1136/jmedgenet-2012-100867.
3. Becker-Heck, A., Zohn, I.E., Okabe, N., Pollock, A., Lenhart, K.B., Sullivan-Brown, J., et al., 2011. The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis formation. Nat. Genet. 43, 79-84. http://dx.doi.org/10.1038/ng.727.
4. Zariwala, M.A., Gee, H.Y., Kurkowiak, M., Al-Mutairi, D.A., Leigh, M.W., Hurd, T.W., et al., 2013. ZMYND10 Is Mutated in Primary Ciliary Dyskinesia and Interacts with LRRC6. Am J Hum. Genet.93, 336-345. http://doi.org/10.1016/j.ajhg.2013.06.007.
5. Nakhleh, N., Francis, R., Giese, R.A., Tian, X., Li, Y., Zariwala, M.A., et al., 2012. High prevalence of respiratory ciliary dysfunction in congenital heart disease patients with heterotaxy. Circulation. Am. Heart. Assoc. 125, 2232-2242. http://doi.org/10.1161/CIRCULATIONAHA.111.079780.
6. Sui, W., Hou, X., Che, W., Ou, M., Sun, G., Huang, S., et al., 2016. CCDC40 mutation as a cause of primary ciliary dyskinesia: a case report and review of literature. Clin. Respir. J. 10, 614-621. http://doi.org/10.1111/crj.12268.
7. Takeuchi, K., Kitano, M., Kiyotoshi, H., Ikegami, K., Ogawa, S., Ikejiri, M., et al., 2018. A targeted next-generation sequencing panel reveals novel mutations in Japanese patients with primary ciliary dyskinesia. Auris Nasus Larynx. 45, 585-591. http://doi.org/10.1016/j.anl.2017.09.007.
8. Yang, L., Banerjee, S., Cao, J., Bai, X., Peng, Z., Chen, H., et al., 2018. Compound Heterozygous Variants in the Coiled-Coil Domain Containing 40 Gene in a Chinese Family with Primary Ciliary Dyskinesia Cause Extreme Phenotypic Diversity in Cilia Ultrastructure. Front. Genet. 9, 1-10. http://doi.org/10.3389/fgene.2018.00023.
9. Djakow, J., Svobodová, T., Hrach, K., Uhlík, J., Cinek, O., Pohunek, P., 2012. Effectiveness of sequencing selected exons of DNAH5 and DNAI1 in diagnosis of primary ciliary dyskinesia. Pediatr. Pulmonol. 47, 864-875. http://doi.org/10.1002/ppul.22520.
10. Failly, M., Bartoloni, L., Letourneau, A., Munoz, A., Falconnet, E., Rossier, C., et al., 2009. Mutations in DNAH5 account for only 15% of a non-preselected cohort of patients with primary ciliary dyskinesia. J. Med. Genet. 46, 281-286. http://doi.org/10.1136/jmg.2008.061176.
11. Hornef, N., Olbrich, H., Horvath, J., Zariwala, M.A., Fliegauf, M., Loges, N.T., et al., 2006. DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arm defects. Am. J. Respir. Crit. Care Med. 174, 120-126. http://doi.org/10.1164/rccm.200601-084OC.
12. Ferkol, T.W., Puffenberger, E.G., Lie, H., Helms, C., Strauss, K.A., Bowcock, A., et al., 2013. Primary ciliary dyskinesia-causing mutations in Amish and Mennonite communities. J. Pediatr. 163:383-387. http://doi.org/10.1016/j.jpeds.2013.01.061.

13. Kano ,G., Tsujii, H., Takeuchi, K., Nakatani, K., Ikejiri, M., Ogawa, S., et al., 2016. Whole-exome sequencing identification of novel DNAH5 mutations in a young patient with primary ciliary dyskinesia. Mol. Med. Rep. 14, 5077-5083. http://doi.org/10.3892/mmr.2016.5871.
14. Knowles, M.R., Leigh, M.W., Ostrowski, L.E., Huang, L., Carson, J.L., Hazucha, M.J., et al., 2013. Exome Sequencing Identifies Mutations in< i> CCDC114</i> as a Cause of Primary Ciliary Dyskinesia. Am. J. Hum. Genet. 92, 99-106. http://doi.org/10.1016/j.ajhg.2012.11.003.
15. Loges, N.T., Olbrich, H., Fenske, L., Mussaffi, H., Horvath, J., Fliegauf, M., et al., 2008. DNAI2 mutations cause primary ciliary dyskinesia with defects in the outer dynein arm. Am. J. Hum. Genet. 83, 547-558. http://doi.org/10.1016/j.ajhg.2008.10.001.
16. Olbrich, H., Häffner, K., Kispert, A., Völkel, A., Volz, A., Sasmaz, G., et al., 2002. Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. Nat. Genet. 30, 143-144. http://doi.org/10.1038/ng817.
17. Xu, X., Gong, P., Wen, J., 2017. Clinical and genetic analysis of a family with Kartagener syndrome caused by novel DNAH5 mutations. J. Assist. Reprod. Genet. 34, 275-281. http://doi.org/10.1007/s10815-016-0849-3.