Mutation analysis of genes involved in sperm motility:

A study on patients with total sperm immotility

Rute Ribeiro Pereira

Master's degree in Cell and Molecular Biology
Department of Biology
2013/14

Orientation
Mário Sousa, Full Professor Department of Microscopy, Institute of Biomedical Sciences Abel Salazar, University of Porto

Co-orientation

Jorge Oliveira, MSc, Molecular Genetics Unit, Centre of Medical Genetics Dr. Jacinto Magalhães, Hospital Centre of Porto

Todas as correções determinadas
pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

Agradecimentos

Antes de agradecer a todas as pessoas que me apoiaram ao longo deste ano, quero agradecer ao Instituto de Ciências Biomédicas Abel Salazar e ao Centro de Genética Médica Dr. Jacinto Magalhães/Centro Hospitalar do Porto, por me cederem todas as condições físicas e financeiras necessárias à realização e concretização deste projeto ambicioso e também dispendioso.

Passando para as pessoas que me acompanharam e tornaram possível este projeto, em primeiro lugar, preciso de agradecer ao meu Orientador, o Professor Doutor Mário Sousa, por planear este projeto e me dar a oportunidade de trabalhar num tema bastante interessante e num grupo de trabalho fabuloso. Agradeço-lhe por todo apoio e disponibilidade que me deu ao longo do projeto, mesmo não estando diretamente a trabalhar comigo.

Um *Muito Obrigada* ao Jorge Oliveira, o meu coorientador, por ter sido um elemento fundamental neste projeto e por, apesar das "milhentas" coisas que tem a seu cargo, ter estado sempre disponível para me ensinar e orientar. Agradeço o tempo que "*gastou*" comigo e agradeço o facto de ter sempre dado ideias para ser possível continuar o projeto, nomeadamente a aplicação da NGS que enriqueceu bastante o trabalho.

Quero agradecer também à Dra. Rosário Santos, por ter permitido que fizesse o meu trabalho de dissertação na Unidade de Genética Molecular do Centro de Genética Médica Dr. Jacinto de Magalhães/ Centro Hospitalar do Porto, e por toda simpatia e disponibilidade que sempre me demonstrou.

Não posso deixar de dizer que todo o grupo da Unidade de Genética Molecular é fantástico. Foi um prazer trabalhar com todos, agradeço a todos pela boa disposição, simpatia e por estarem sempre disponíveis a ajudar-me e ensinar-me. Aprendi muito com este excelente grupo não só a nível profissional como também pessoal.

Ao Nuno, agradeço as conversas sobre o *Once Upon a Time,* sobre as danças (aprendi muito!!), etc, etc…

Agradeço à Rita por me ter "*cartolado*" assim como por todo o apoio que me deu durante o meu trabalho.

À Dra. Paula e Dra. Isabel agradeço todo o apoio na pesquisa de futuros trabalhos e programas doutorais. Em particular, agradeço à Dra. Paula pela carta de

recomendação que, com certeza também influenciou a minha entrada no programa doutoral em Bioengenharia na área Terapias Celulares e Medicina Regenerativa, no Instituto Superior Técnico.

À Dra. Márcia e à Emília agradeço toda a disponibilidade em me ensinar.

Aos meus amigos, 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!*

Ao Hélder, agradeço a paciência, foste a pessoa em quem mais descarreguei o *stress*! Obrigada por me ouvires reclamar quando as coisas corriam pior, mesmo não percebendo totalmente o que estava a dizer! Obrigada por todo o apoio que me tens vindo a dar.

Em último, mas mais importante, Agradeço à minha família, foram vocês que me fizeram o que sou hoje.

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

Um Muito Obrigada à minha Mãe! Sei de tudo o que passaste para que eu conseguisse formar-me e ser o que sou hoje. Não há palavras para explicar o que fizeste por mim e como estou agradecida por tudo. Obrigada Mãe!

"Cada novo amigo que ganhamos no decorrer da vida aperfeiçoa-nos e enriquece-nos, não tanto pelo que nos dá, mas pelo que nos revela de nós mesmos."

Miguel de Unamuno y Jugo

Resumo

O espermatozoide caracteriza-se por duas partes fundamentais, a cabeça e a cauda (flagelo). O axonema (Ax) é o motor do movimento dos espermatozoides e estende-se por todo o comprimento do flagelo. Este é composto por 9 pares de microtúbulos periféricos um interno (A) e um externo (B) (dupletos), unidos entre si através de pontes de nexina, e ligados a um par de microtúbulos centrais, por projeções radiais medianas. Os microtúbulos A têm 2 braços de dineína, um exterior e outro interior. Ao conjunto dos microtúbulos centrais, da ponte fibrilar que os une e da bainha fibrilar que os rodeia, denomina-se complexo do par central (CPC). O CPC funciona como distribuidor de sinal para a motilidade, distribuindo o sinal para as projeções radiais, que, com o auxílio de um complexo designado complexo regulador das dineínas (DRC), transmitem sinais regulatórios e ativam seletivamente os braços de dineína. Estes convertem, através de ATPases, a energia química contida no ATP em energia mecânica, que originará o movimento dos espermatozoides. As fibras densas externas e a bainha fibrosa são duas estruturas que rodeiam o Ax. A bainha fibrosa é uma importante estrutura que influência/modela a flexão e o movimento flagelar, e contém enzimas glicolíticas que fornecem energia necessária à motilidade

A infertilidade é uma doença que afeta cerca de 10-17% dos casais, sendo uma patologia no elemento masculino, responsável por cerca de 30-45% dos casos. A imotilidade dos espermatozoides destaca-se como uma das principais causas de infertilidade masculina. A base genética e os mecanismos moleculares que controlam a motilidade do flagelo dos espermatozoides não são completamente compreendidos. A discinesia ciliar primária (PCD) e a displasia da bainha fibrosa (DFS) são consideradas as principais síndromes associadas à imotilidade dos espermatozoides. PCD é uma doença autossómica recessiva, em que a maioria dos pacientes é infértil devido a imotilidade provocada por anomalias no Ax. A síndrome de Kartagener (KS), caracterizada por *situs inversus*, sinusite crónica e bronquiectasias, ocorre em 50% dos casos de PCD. A DFS é outra doença maioritariamente caracterizada por hiperplasia e desorganização da bainha fibrosa. Vários genes foram associados a estas síndromes e à formação e função do flagelo.

Os espermatozoides totalmente imóveis de cinco doentes portugueses com astenozoospermia foram previamente analisados por microscopia eletrónica de transmissão. Nessa análise, foram identificados 4 pacientes compatíveis com a síndrome DFS associados a anomalias na bainha fibrosa, no CPC, projeções radiais, e braços de dineína; e um paciente com *situs inversus* em que foi detetada ausência dos braços de dineína e pontes de nexina (compatível com a KS). Assim, o objetivo

deste trabalho foi identificar uma causa genética para os fenótipos dos pacientes anteriormente analisados. Para tal, começamos por investigar na literatura genes envolvidos na DFS e PCD bem como associados às estruturas flagelares em que foram encontradas anomalias nos nossos pacientes. Foram selecionados sete genes: *CCDC39 e CCDC40 (*braços internos de dineína e o DRC*), DNAH5 e DNAI1 (*braços externos de dineína*), RSPH1 (*projeções radiais e CPC*), AKAP3* e *AKAP4* (bainha fibrosa). Foram desenhados oligonucleotídeos iniciadores para as regiões exónicas (e fronteiras intrónicas) de cada gene. Seguidamente, essas regiões foram amplificadas pela reação em cadeia da polimerase (PCR) e sequenciadas pelo método de Sanger. Adicionalmente, foi realizada a sequenciação de todas as regiões exónicas do genoma (exoma) usando sequenciação de nova geração (NGS) no paciente com *situs inversus*.

Com este trabalho, identificamos **9 novas variantes de sequência** (*CCDC39*: c.2540A>G, *CCDC40*:c.2620-92C>T, *CCDC103*:c.104G>C, *DNAH5*:c. 5882+133A>G e c.11570+124G>A, *DNAH6*:c.3167A>T, *GAS8*:c.828C>G, *INSL6*:c.262_263delCC e *SPAG17*:c.4445G>T); e **10 variantes raras** (*CCDC39*:c.233G>A; *CCDC40*:c.2682G>A; *DNAH5*:c.1537-102T>A, c.1537-100_1537del-99delTT, c.3835-3delT, c.7408-84_7408-83delAT , c.10282-81delT, c.10872+84T>C; *DNAH10*:c.7895C>T e *DNAI1*:c.81+61A>G). As variantes nos genes *CCDC103* e I*NSL6* são as mais promissoras a estarem relacionadas com a imotilidade.

O nosso trabalho demonstrou a heterogeneidade e a complexidade genética da DFS, PCD e da motilidade dos espermatozoides. Destacamos as dificuldades em investigar as causas genéticas da imotilidade usando a sequenciação convencional de Sanger, nomeadamente devido à morosidade e custos associados. Também evidenciamos a importância da aplicação da NGS para investigação de doenças geneticamente complexas e para compreender a base genética dos processos altamente complexos como a formação do flagelo de espermatozoides e a motilidade.

Com este trabalho pretendeu-se iniciar o diagnóstico genético dos pacientes com imotilidade total dos espermatozoides, bem como ajudar a encontrar potenciais marcadores genéticos para os indivíduos com este tipo de patologia.

Para investigação futura, identificar quais são os genes (e variantes) que estão envolvidos neste tipo de patologias, como se relacionam, bem como, entender de que modo as alterações detetadas poderão afetar a função da proteína, são aspetos essenciais para planear e desenvolver futuras terapias, como a terapia genética.

Palavras-chave: Displasia da bainha fibrosa; Sequenciação do exoma; Sequenciação de nova geração; Sequenciação de Sanger; Síndrome de Kartagener; mobilidade dos espermatozoides.

Abstract

The spermatozoon is divided into two fundamental parts, the head and the tail (flagellum). The axoneme (Ax) is the flagellar motor of the sperm cell and extends throughout the flagellum. The Ax is composed by 9 peripheral doublets of microtubules (an internal (A) and an external (B)), linked to each other by nexin bridges, and connected by radial spokes (RS) to a single pair of central microtubules. The microtubule A has two dynein arms (DA), outer (ODA) and inner (IDA). The central microtubules are linked by a fibrilar bridge and are surrounded by a fibrilar central sheath, constituting the central pair complex (CPC). The CPC functions like a distributor of motility signal. The CPC gets the signal and distributes to the RS that, allied with the dynein regulatory complex (DRC), transmit regulatory signals that selectively activate the DA. DA converts the chemical energy contained in ATP, through ATPase, into the mechanical energy which will allow the flagellar movement. . The outer dense fibre and the fibre sheath (FS) surround the Ax. The FS influences/modulates the bending and the flagellar beat and provides energy for motility. In addition, it contains glycolytic enzymes that provides energy for motility.

Infertility is a current significant problem that affects about 10-17% of couples worldwide. A male contribution to infertility is found in 30–45% of the cases. There are several causes for male infertility being the sperm immotility one of the major. The genetic basis and molecular mechanisms that control sperm motility are not fully understood. Primary ciliary dyskinesia (PCD) and dysplasia of the fibrous sheath (DFS) are considered the two main disorders associated with sperm immotility. PCD is an autosomal recessive disease causing male infertility due to sperm immotility associated to abnormalities in the Ax. Kartagener syndrome (KS), characterized by a combination of *situs inversus*, chronic sinusitis, and bronchiectasis, occurs in about 50% of PCD patients. DFS is a disease mostly characterized by hyperplasia and disorganization of the FS. Several genes have been described as associated to these syndromes and to the assembly and function of the normal sperm flagellum.

The total immotile spermatozoa of five Portuguese patients with asthenozoospermia were previously analysed by transmission electron microscopy. This ultrastructural analysis revealed 4 cases with DFS associated with disrupted CPC, RS, DA and doublets; and one patient with *situs inversus* who presented absence of DA and nexin bridges (compatible with KS). Therefore, the objective of the present work was to identify genetic alterations that could explain patients' phenotypes. In that way, we searched for genes associated to DFS and PCD, as well

as, with the ultrastructural defects in sperm cells of these patients. Seven genes were initially selected for analysis: *CCDC39* and *CCDC40* (DRC and IDA)*, DNAH5* and *DNAI1* (ODA), *RSPH1* (RS), *AKAP3 and AKAP4* (FS). We designed primers for the exonic regions (plus intronic boundaries) of each gene. Then, these regions were amplified by polymerase chain reaction (PCR) and sequenced by Sanger's method. Additionally, we sequenced all the exonic regions of the genome (exome) in the patient with *situs inversus* using next-generation sequencing (NGS) technology.

As result, we identified **9 new DNA sequence** variants (*CCDC39*:c.2540A>G, *CCDC40*:c.2620-92C>T, *CCDC103*:c.104G>C, *DNAH5*:c. 5882+133A>G e c.11570+124G>A, *DNAH6*:c.3167A>T, *GAS8*:c.828C>G, *INSL6*:c.262_263delCC e *SPAG17*:c.4445G>T); and **10 rare variants** (*CCDC39*:c.233G>A; *CCDC40*:c.2682G>A; *DNAH5*:c.1537-102T>A, c.1537-100_1537-99delTT, c.3835-3delT, c.7408-84_7408-83delAT, c.10282-81delT, c.10872+84T>C; *DNAH10*:c.7895C>T e *DNAI1*:c.81+61A>G). The variants in *CCDC103* and *INSL6* are the most promising to have a role in sperm immotility.

Our work demonstrates the genetic heterogeneity and complexity of DFS, PCD and sperm immotility. We highlight the difficulties for the identification of the genetic causes using conventional (Sanger) sequencing, especially due to the labour and costs involved. We also demonstrate the importance of exome sequencing, as broad mutation screening method, very useful to understand the genetics of highly complex processes as formation/assembly of sperm flagellum and sperm motility. With these work we intended to initiate the genetic diagnosis of this patients, and help to identify potential genetic markers for individuals with this kind of pathology.

Understanding which genes are involved in sperm motility, how they are correlated as well as the exactly protein's function and how these variants may affect its function, is essential to plan and develop, in the future, a gene therapy.

Key words: Dysplasia of the fibre sheath (FSD); next-generation sequencing; Genetic diagnosis; Kartagener syndrome; Sanger sequencing; Sperm immotility

Table of Contents

List of Figures and Tables

Figures

Page

Tables

Page

List of abbreviations

Full name	*Abbreviation*
ATPases associated diverse cellular activities	*AAA*
Anterior Acrosomal Region	*AAR*
A-kinase anchoring protein	*AKAP*
Amplicon Length	*AL*
Annulus	*An*
Acrosomal Region	*AR*
Acrosomal Vesicle	*AV*
Axoneme	*Ax*
Binary Alignment Map	*BAM*
Basal Plate	*BP*
base pairs	*bp*
Central Pair Complex	*CPC*
Dynein arm	*DA*
Dysplasia of the Fibrous Sheath	*DFS*
Dynein Heavy Chain	*DHC*
Deoxyribonucleic Acid	*DNA*
Dynein Regulatory Complex	*DRC*
End Piece	*EP*
Equatorial Region	*ER*
Exome Sequencing	*ES*
Exonic Splicing Enhancer	*ESE*
Fibre Sheath	*FS*
Folicule Stimulating Hormone	*FSH*
Gonodotropin relasing hormone	*GnRH*
Heavy Chain	*HC*
Human Splice Finder	*HSF*
Intermediate Chain	*IC*
Inner Dynein Arms	*IDA*
Kartagener Syndrome	*KS*
Light Chain	*LC*
Luteinizing Hormone	*LH*
Midpiece	*MP*
messenger RNA	*mRNA*
Mutant	*mt*
mitochondrial DNA	*mtDNA*
National Center for Biotechnology Information	*NCBI*
nanograms	*ng*
Next Generation Sequencing	*NGS*
Neck Piece	*NP*
Nuclear Respiratory Factor	*NRF*
Native	*nt*
Outer Dynein Arms	*ODA*
Outer Dynein Arm Docking Complex	*ODA-DC*
Outer Dense Fibres	*ODF*
Post-Acrosomal Region	*PAR*
Proximal Centriole	*PC*
Primary Ciliary Dyskinesia	*PCD*
Primary ciliary dyskinesia protein 1	*Pcdp1*
Polymerase Chain Reaction	*PCR*
AMP-dependent protein kinase	*PKA*
Polymorphism Phenotyping v2	*Polyphen-2*
Principal Piece	*PP*

Posterior Ring	*PR*
Ribonucleic acid	*RNA*
Radial Spokes	*RS*
Segmented (or Striated) Columns	*SC*
Sorting Tolerant From Intolerant	*SIFT*
Single Nucleotide Variants	*SNV*
Spermatozoon (plural spermatozoa)	*Sperm*
Sperm Oocyte Activating Factors	*SOAF*
Serine/Arginine	*SR*
Splice Site Acceptor	*SSA*
Search Tool for the Retrieval of Interacting Genes/Proteins	*STRING*
Transmission electron microscopy	*TEM*
mitochondrial Transcription Factor A	*TFAM*
mitochondrial Transcription Factor B2	*TFB2*
Temperature melting	*Tm*
Transition Proteins	*TP*
microlitres	*µl*
University of California, Santa Cruz	*UCSC*
Variant Call Format	*VCF*
Whole Genome Sequencing	*WGS*
World Health Organization	*WHO*

Amino acid full name	**Abbreviator**	***Amino acid full name***	***Abbreviator***
Alanine	Ala	*Leucine*	Leu
Arginine	Arg	*Lysine*	Lys
Asparagine	Asn	*Methionine*	Met
Aspartic acid	Asp	*Phenylalanine*	Phe
Cysteine	Cys	*Proline*	Pro
Glutamic acid	Glu	*Serine*	Ser
Glutamine	Gln	*Threonine*	Thr
Glycine	Gly	*Tryptophan*	Trp
Histidine	His	*Tyrosine*	Tyr
Isoleucine	Ile	*Valine*	Val

Nucleotide full name	**Abbreviator**	**Nucleotide full name**	**Abbreviator**
Adenine	A	*Thymine*	T
Guanine	G	*Cytosine*	C

I. INTRODUCTION

INTRODUCTION

1. SPERM: A LITTLE OF HISTORY

In 1677, the scientist Antonie van Leeuwenhoek (1632-1723), a great scientist known as the father of microbiology, with hand-made lenses and microscopes, first observed the spermatozoa and called them ‘‘animalacula’’. The word spermatozoa was invented, years later, by Karl von Baër (1792-1876), who established the new science of comparative embryology alongside with comparative anatomy (Magner, 2002). Spermatozoon (sperm; plural spermatozoa) has origin from the Greek word “*sperein*” which means “to sow”. This is because it was assumed that the semen was the crucial factor for sexual reproduction and that females merely provided “*the fertile garden in which the male seed would flourish and grow” (citing* Birkhead and Montgomerie, 2008). However, the really function of the sperm remained uncertain and controversial. Leeuwenhoek speculated that the “animalcules” were able to penetrate the egg and fertilize it. Oppositely, some of his colleagues argue that these “animalcules” were just parasites, and believed that were present in vast number, in excess of what was necessary to fertilize an egg (Birkhead and Montgomerie, 2008). The roles of spermatozoa begun to be clarified from 1780s, when the scientist Lazzaro Spallanzani (1729–1799), using amphibians, showed that the contact between egg and semen is essential for the development of a new animal, ending with the theory of “*aura spermatica”* that stated that the development of the egg would be due to a sort of vapor emanating from the sperm. However, Spallanzani failed to prove the specific role of the sperm. Years later, using dogs, the first artificial insemination was performed, and the roles of the sperm and of the oocyte in sexual reproduction began to be understood (Capanna, 1999). Albert von Kölliker (1817-1905), was one of the first to interpret tissue structure in terms of cellular elements and to conclude that the semen off all animals contains sperm and determine that spermatozoa develop from cells residing in the testes. Franz von Leydig (1821-1908) described the microscopic characteristics of the testicular cells, that nowadays are called Leyding cells and posteriorly, Enrico Sertoli (1842-1910), in 1865 described the Sertoli cells (Christensen, 2007; Staff, 2009).

The history teach us that all advances about structure and function of the sperm walked, walks and will be walking side by side with the development of microscopy and technologies. As we are in the age of high-tech breakthroughs, progressions will be made and consequently many advances on the scientific knowledge about sperm will be certainly obtained.

2. THE SPERM BIOGENESIS: SPERMATOGENESIS

Spermatogenesis is the biological process that forms the spermatozoa and takes place within seminiferous tubule boundaries of the testis. The epithelium of the seminiferous tubule consists of germ cells that form numerous concentric layers penetrated by Sertoli cells, whose one of the functions is to nourish the germ cells and maintain their cellular associations throughout the process of spermatogenesis, which lasts approximately 64 days (Hess and de Franca, 2008).

At birth, in the seminiferous tubules, A-dark spermatogonia (originated during the fetal life from primordial germ cells) remain mitotically inactive until the peri-pubertal period. At the pubertal period the production of gonadotropic hormones induces the mitotic proliferation of A-dark spermatogonia, originating A-pale spermatogonia. After the pubertal period and through a massive mitosis they will allow the continuous production of spermatozoa all over the reproductive male lifespan. This phase is known as proliferative phase. After puberty, A-pale spermatogonia differentiate into B-type spermatogonia, which then enter meiosis and give rise to preleptotene, leptotene, zygotene and pachitene primary spermatocytes. Then, the first meiotic division occurs and originates two secondary spermatocytes. Each experiences the second meiotic division and give rise to four round spermatids. This is designated meiotic phase (De Kretser, 1998; Sutovsky and Manandhar, 2006). Due to the unique architecture of the seminiferous epithelium, the sequence of germ cells types in the seminiferous epithelium is established in a stepwise manner. Consequently, Sertoli cells, A-dark and A-pale spermatogonia are found in newborn and children, while spermatocytes and spermatids appear in the adult testis. Temporarily, the proliferative and meiotic phase of spermatogenesis can be synchronized or can be sequential. Relatively to the localization, the proliferative phase occurs in the nutrient rich basal compartment, while the meiotic phase takes place within the adluminal compartment, where cells are protected by the blood-testis barrier (Hess and de Franca, 2008).

In the last part of spermatogenesis, round spermatids undergo marked morphological changes, but no further cell divisions occur. This stage can be divided into spermiogenesis and spermiation. **Spermiogenesis** consists of three main steps: (i) nuclear condensation and movement of the nucleus to the periphery of the cell; (ii) formation the acrosome, which becomes attached to the surface of the nucleus; and (iii) flagellar formation. These proceeds through 4 steps: Golgi, capping, acrosomal and maturation (De Kretser, 1998; Sutovsky and Manandhar, 2006).

During the **Golgi phase**, the membrane of the acrosomal vesicle (AV), that begins its formation in the pachytene stage, is attached to the nuclear envelope and grows due to the constant influx and fusion of Golgi derived vesicles (Abou-Haila and Tulsiani, 2000). The centrioles, located near to Golgi, are translocated to the pole opposite to the developing acrosome, establishing the polarity of the cell. The distal centriole participates in formation of the axonema (Ax) and the flagellum, while the proximal centriole (PC) will ultimately give rise to the neck piece (NP). The Golgi phase also involves the development of the midpiece (MP), where the mitochondria, which were previously located around the nucleolus, start to didperse close to periphery of the cell and close to plasma membrane (Gupta, 2006). The distal centriole disappears in spermiogenesis, meaning that sperm cells only contain one centriole, the PC.

In the **Cap phase**, the nucleus, the AV and the body of the spermatids become elongated, and chromatin begins to condensate. The Golgi apparatus continues to produce and deliver the proteins and membranes needed for enlargement and differentiation of the AV (Abou-Haila and Tulsiani, 2000; Moreno *et al.*, 2000).

In the **acrosomal phase**, the nucleus becomes further condensed and elongated as well as the AV and the cell. The mitochondria migrate to the surroundings of the developing flagellum and arranged into a helix that terminates at the annulus (An). During this phase, occurs the formation of outer dense fibres (ODF) associated with the Ax and fibrous sheath (FS), thus completing the formation of the tail and the developing spermatozoa orient themselves in a way that their tails point towards the centre of the lumen, away from the epithelium (Gupta, 2006).

In the **maturation phase**, is when occurs the extrusion of the rest of the cytoplasm (the residual body).

During spermiogenesis, histones are remove and replaced by transition proteins and these by protamines (Meistrich *et al.*, 2003). Also, the majority of the mitochondria are removed and just a few are maintained in the MP (Piomboni *et al.*, 2012). After spermiogenesis, it occurs the **spermiation**, which consists in the release of the fully differentiated testicular spermatozoa into the lumen of the seminiferous tubules to travel through to the *rete testis* to the epididymis. Release is attained by disappearance of cell junctions between Sertoli cells and spermatozoa and retraction of the cytoplasm of Sertoli cells. Once released the spermatozoa, Sertoli cells phagocyte the remainder of the residual bodies. This process is complex and can be particularly affected by hormonal modifications, temperature changes and toxins (Donnell *et al.*, 2011).

Spermatogenesis (resumed on the scheme of the Fig. 1) is a remarkable process that involves multiple factors, such as endocrine, autocrine and juxtacrine, paracrine. About 100 genes have been found by genetic engineering to be implicated in spermatogenesis (Escalier, 2001). It is maintained by the secretion of the hypothalamic gonadotropin releasing hormone (GnRH) that stimulates the secretion of follicle stimulating hormone (FSH) and luteinizing hormone (LH) from the pituitary gland. LH is responsible for stimulating the secretion of testosterone by Leydig cells located in the testicular stroma, whereas FSH and testosterone act over Sertoli cells, which are essential, among other important roles, to nourish the developing sperm cells (Holdcraft and Braun, 2004; Chocu *et al.*, 2012).

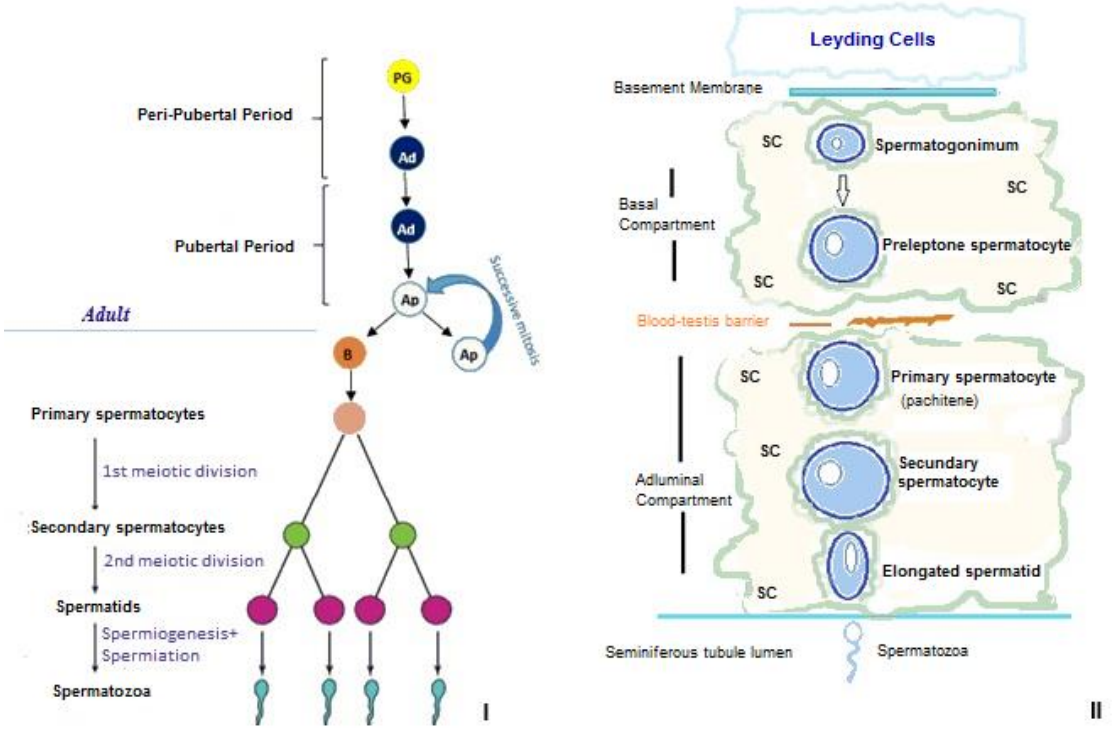
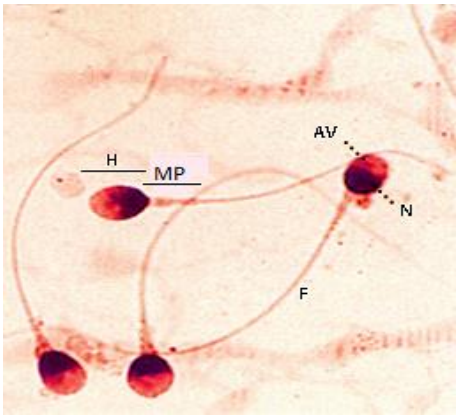


Figure 1. I. Representation of the basic steps that occur in male germ cells during spermatogenesis.
At birth, in the seminiferous tubules, A-dark spermatogonia (Ad), originated from primordial germ cells (PGC), remain mitotically inactive until the peri-pubertal period. In pubertal period the production of gonadotropic hormones induces the mitotic proliferation of A-dark spermatogonia, originating A-pale spermatogonia (Ap), which after the pubertal period, will allow the continuous production of spermatozoa (this is the proliferative phase). A-pale spermatogonia will differentiate into B-type (B), that will enter meiosis and give rise to primary spermatocytes (pink circles). These cells will undergo the first meiotic division of spermatogenesis to originate two secondary spermatocytes (green circles) and each experiences the second meiotic division and give rise to four round spermatids (purple circles). This is designated meiotic phase. The round spermatids undergo marked morphological changes (spermiogenesis) to form the mature spermatozoa, which are released from Sertoli cells into the lumen of the seminiferous tubule (spermiation).
II. Schematic drawing of the seminiferous tubule illustrating the morphological features of germ cells development. The proliferative phase occurs in the nutrient rich basal compartment, while the meiotic phase takes place within the adluminal compartment, where cells are protected by the blood-testis barrier.

3. THE SPERM STRUCTURE

In this section we will make an introduction about the complex and remarkable structure of the sperm cell, with more emphasis to the components involved in sperm motility.

The spermatozoon (Fig. 2 and 3) is divided into two fundamental parts, the sperm head and the sperm tail or flagellum. The head is divided in acrosomal and postacrosomal regions and contains the genetic material in the nucleus and important enzymes for fertilization in the acrosomal vesicle (AV).



The sperm flagellum is responsible for sperm motility, and is divided into four major regions, the neck piece (NP), the midpiece (MP), the principal piece (PP) and the end piece (EP). The spermatozoon is surrounded by a plasma membrane.

Figure 2 Spermatozoa seen in phase contrast microscopy. Head (H), with acrosomal vesicle (AV) and nucleus (N). Flagellum (F), with midpiece (MP) enlarged to contain the mitochondrial sheath.
Image provided by Professor Mário Sousa.

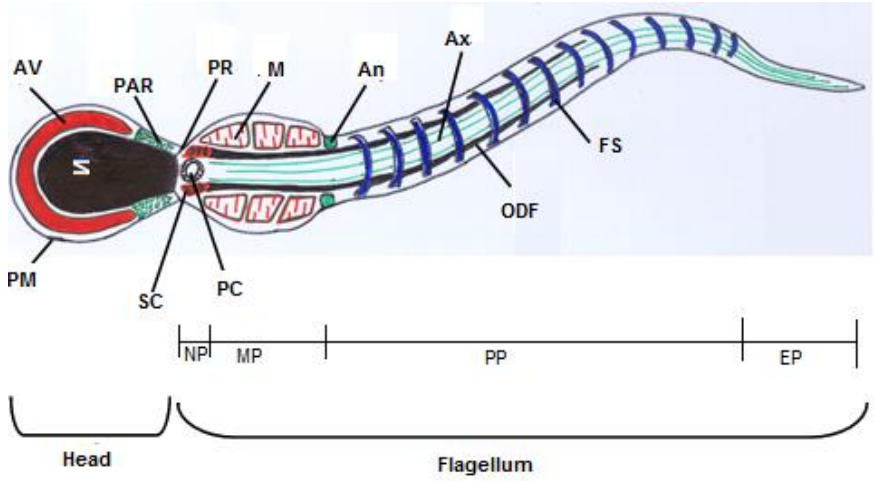


Figure 3. Schematic drawing of a human spermatozoon. The spermatozoon is morphologically divided into two parts: the head and the flagellum. The head is composed by: plasma membrane (PM; that surrounds all sperm cell); acrosomal vesicle (AV); nucleus (N); postacrosomal region (PAR), which connects the head to the neck piece (NP); and by a posterior ring (PR). The sperm flagellum is divided into four major regions, the NP, the midpiece (MP), the principal piece (PP) and the end piece (EP). The MP includes: the proximal centriole (PC), segmented columns (SC), outer dense fibers (ODF), axoneme (Ax, which extends for all flagellum) and mitochondrial sheath (M). The proximal PP starts immediately after the annulus (An) and contains the Ax, the ODF and the rings of fibrous sheath (FS). In the distal PP are only observed the Ax and FS, and the EP only contains the Ax.
Image provided by Professor Mário Sousa.

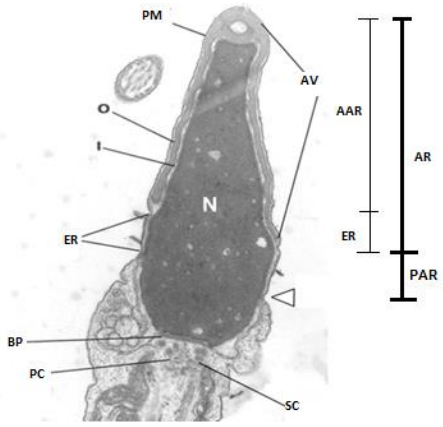
3.1. Sperm plasma membrane

The sperm plasma membrane is a continuous cell boundary and its function consists in maintaining cell integrity and form a dynamic interface between the cell and environment. It has a characteristic subdivision into regional domains that differ in physical, chemical and antigenic properties; in the architectural organization, in the glycoproteins and other components; as well in its function. These domains are dynamic and undergo changes in organization and composition during maturation, capacitation, oocyte-sperm interaction and fusion (Toshimori, 1998; Flesch and Gadella, 2000).

The sperm head plasma membrane is separated from the plasma membrane of the MP by the posterior ring (PR). The membrane of the MP is separated from the plasma membrane of the PP by an annular ring: the annulus (An). Membrane domains not presenting any physical barrier also show limits restricting the mobility of components or restrict certain antigens, which are maintained by a regional specialization of cytoskeletal elements. For instance, the cytoskeletal proteins actin and spectrin are localized in acrosomal region (AR) and PP, whereas vimentin is restricted to the ER (Grudzinskas and Yovich, 1995).

3.2. Sperm Head Structure

The main components of the sperm head are the nucleus and the AV. The head (Fig. 4) is surrounded by the plasma membrane and is morphologically divided into two regions: acrosomal (AR) and postacrosomal (PAR). The AR is further divided into anterior acrosomal region (AAR) and equatorial region (ER). The PAR starts in the ER and ends in the PR and is devoid of the AV coat and is filled with numerous proteins and filaments involved in sperm-egg interaction and activation (Toshimori and Ito, 2003).



According to the World Health Organization (WHO) (WHO, 2010), a “normal” sperm head has about 4.5 µm in length and 3 µm in diameter. It is smooth and generally oval in shape. However, spermatozoa are pleomorphic it is likely that the range of normal sperm both in fertile and infertile men is about 30%.

Figure 4. Ultrastructure of the sperm head. Images from Transmission Electron Microscopy (TEM).
The sperm head is surrounded by the plasma membrane (PM) and is divided into the acrosomal region (AR) and postacrosomal region (PAR). The AR is further divided into anterior acrosomal region (AAR) and equatorial region (ER). The head contains the acrosomal vesicle (AV), with the outer (O) and inner (I) acrosomal membranes; and the nucleus (N). The posterior ring (arrowhead) is the frontier between head and flagellum. Components of the flagellum: proximal centriole (PC), the basal plate (BP) and the segmented columns (SC).
Image provided by Professor Mário Sousa.

3.2.1. The acrosome

The acrosome is an organelle that develops over the anterior half of the sperm head, and comprises 40–70% of the head area. In this region, the nucleus presents a reduced number of small pale areas ("vacuoles") that are uncondensed nuclear regions containing the rRNA genes and other genes that are first transcribed after gamete fusion (Grudzinskas and Yovich, 1995).

The acrosomal membrane has two membranes: the inner and outer. The inner acrosomal membrane covers the nuclear envelope and is highly stable, while the outer acrosomal membrane is unstable and easily disrupted. These differences reflect their respective roles in cumulus penetration and in *zona pellucida* binding and penetration. The acrosomal matrix is composed by a number of different hydrolytic enzymes and other proteins, which are necessary to penetrate the *zona pellucida*. The AV differs from others organelles derived from Golgi apparatus, such lysosomes, since is a secretory vesicle and contains specific enzymes like proacrosin. Proacrosin and hyaluronidase are the most important and the best characterized enzymes of the AV (Grudzinskas and Yovich, 1995; Abou-Haila and Tulsiani, 2000). Acrosin is present only in spermatogenic cells and it is synthesized as a proenzyme (the proacrosin), which after the acrosome reaction is activated to the mature enzyme (Zahn *et al.*, 2002). This enzyme has a vital role in fertilization due to its capability to hydrolyze the *zona pellucida* of the oocyte (Yamagata *et al.*, 1998) and participation in sperm-*zona pellucida* interaction (Howes and Jones, 2002; Mao and Yang, 2013). The absence of proacrosin or reduced activity of acrosin may be related with infertility (De Jonge *et al.*, 1993; Adham *et al.*, 1997).

3.2.2. Perinuclear material

The perinuclear material (also called perinuclear theca) is a cytoskeleton element that covers the apical and lateral part of the mature sperm nucleus. It has a volume practically insignificant but is important to keep the AV attached to the nucleus and is involved in sperm–oocyte interaction and oocyte activation. Moreover, it harbors a complex of signalling proteins collectively referred to as sperm oocyte activating factors that after gamete fusion are dispersed across the ooplasm and trigger the signalling pathways, which lead to oocyte activation, polyspermy block and initiation of zygotic development (Mújica *et al.*, 2003; Sutovsky *et al.*, 2003; Oko and Sutovsky, 2009).

3.2.3. Nucleus

The sperm nucleus contains the genetic material that is completely inactive until fertilization (Auger and Dadoune, 1993). The nuclear envelope has no pores in the apical, lateral and PAR regions. After the PR, the plasma membrane initiates the flagellum. In this region it can be observed redundant nuclear envelope that is rich in nuclear pores. In the base of the nucleus lies a region called implantation fossa that is covered by a layer of electron dense material forming the basal plate (BP), which links the nucleus to the segmented columns (SC) (Pedersen, 1972; Grudzinskas and Yovich, 1995).

The sperm nucleus presents a highly organized and compacted structure, consisting of DNA and heterogeneous nucleoproteins. This condensation enables sperm to compact their DNA content, protecting it during the transit through the male and female genital tracts. The organization and amount of DNA, as well as the arrangement and the composition of the nucleoproteins are completely different from somatic nuclei (Ward and Coffey, 1991; Fuentes-Mascorro *et al*., 2000). Chromosomes demonstrate a different pattern (with a 'looped hairpin' conformation orientating centromeres in the direction to nucleus and distal telomeres to periphery) designed for favouring the unpacking and activation of the male genome after fertilization (Mudrak *et al.*, 2005).

During spermiogenesis the sperm nucleus undergoes several changes. At the round spermatid stage, the round nucleus is transcriptional active and histones are the main nuclear proteins. Afterwards, the transcription ceases and histones are replaced first by transition proteins (TP) and then by protamines. The TP1 and TP2 act on DNA supercoiling and reduce the number of DNA breaks, and errors in these proteins were associated with male infertility (Meistrich *et al.*, 2003; Zhao *et al.*, 2004). The nucleus retains a nucleohistone component (about 15%) (Wykes and Krawetz, 2003), but protamines are the major nuclear sperm proteins. Important roles have been proposed (reviewed by Oliva and Dixon, 1991; Oliva, 2006), as: (1) protection of the genetic material from damage; (2) participation in the imprinting of the paternal genome during spermatogenesis; (3) condensation of sperm genome, generating a more compact and hydrodynamic nucleus. Sperm with most hydrodynamic nucleus will move faster, fertilizing the oocyte first. Abnormal sperm protamines concentrations were correlated with DNA fragmentation (Aoki *et al.*, 2005); defects of late spermiogenesis (Carrell and Liu, 2001) and human male infertility (Aoki *et al*., 2006; Oliva, 2006; Steger *et al*., 2003).

3.3. Sperm Flagellum Structure

The flagellum (Fig. 5) is responsible for sperm motility and contains both the site of energy production and the propulsive apparatus for the cell. To fertilization occur, sperm must reach the oocyte and consequently active motility is a crucial condition for natural fertilization take place. The sperm only becomes fully motile after being exposed to several external and intracellular factors during its storage in the epididymis that gives to these cells the ability to exhibit a vigorous motility and subsequently undergo a further period of maturation during their ascent of the female reproductive tract, which is known as capacitation (Aitken *et al.*, 2007).

The flagellum is almost totality the length of human sperm, with about 55-66 µm long (Holstein and Roosen-Runge, 1981). It consists of four distinct segments (Fig.3): the NP, the MP, the PP and the EP. The NP contains the BP, the proximal centriole (PC) and the SC. The MP contains the Ax, the outer dense fibres (ODF) and the mitochondrial sheath that ends at the An. The PP contains the Ax, the ODF (proximal PP) and the fibrous sheath (FS: proximal and distal PP). The EP contains only the Ax (Grudzinskas and Yovich, 1995).

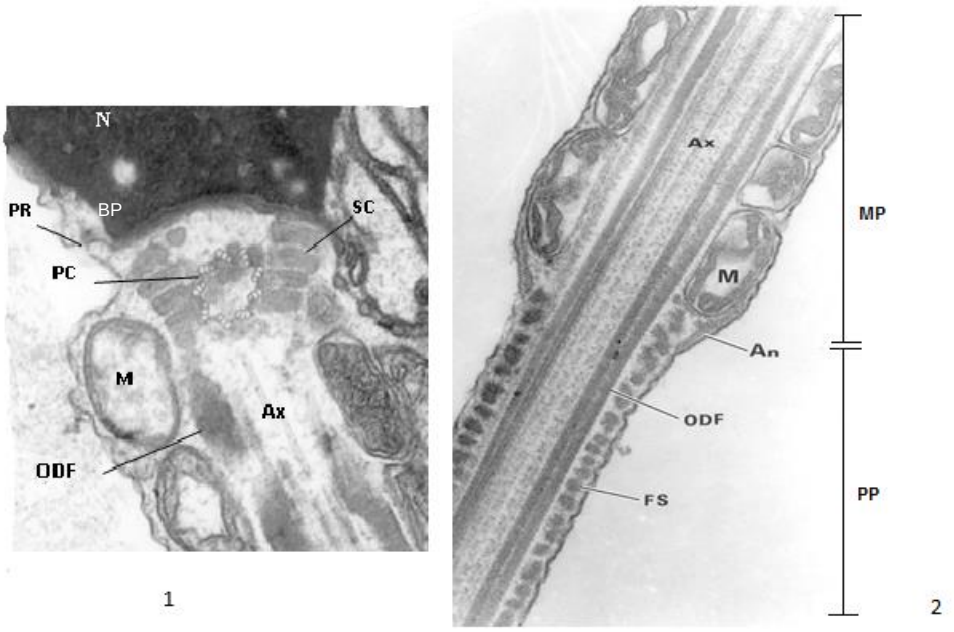


Figure 5. 1-Ultrastructure of sperm neck piece. The posterior ring (PR) makes the connection between the head and the neck piece. The neck piece is composed mainly by the segmented columns (SC), the proximal centriole (PC) and the basal plate (BP).
2- Ultrastructure of sperm midpiece (MP) and principal piece (PP), which are separated by annulus (An). The axoneme (Ax) in the PP is surrounded by the outer dense fibers (ODF). The MP contains the mitochondrial sheath (M) and the PP is distinguished by the presence of a ring of the fibrous sheath (FS).
Image provided by Professor Mário Sousa.

3.3.1. Axoneme (Ax)

The Ax is the flagellar motor of the sperm and therefore knowledge about its structures is essential to understand the mechanisms of sperm motility. The Ax (Fig. 6, 7) is composed by two central microtubules surrounded by nine peripheral microtubule doublets, forming the 9+2 microtubules pattern. The Ax begins to be formed at spermatogenesis Golgi phase by the microtubule polymerizing activity of distal centriole. Subsequently, in the acrosomal phase, the PC attaches to the basal nuclear envelope and originates 9 peripheral microtubule doublets (De Krestser, 1998; Sutovsky and Manandhar, 2006). Each doublet is externally anchored to 9 corresponding asymmetric ODF in MP and proximal PP. The Ax is also surrounded by mitochondria in the MP and by the FS in PP (Afzelius *et al*., 1995; Inaba, 2007).

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. The doublets are linked by nexin links and each consists of an internal complete microtubule (A) onto which is attached a second external and incomplete microtubule (B). Extending from each microtubule A in the direction of the microtubule B of the adjacent doublet exist a pair of projections, named dynein arms (DA) due to the major component of these arms be the protein dynein. DA are designated by their position as either 'inner' or 'outer' (IDA or ODA, respectively) (Grudzinskas and Yovich, 1995). The two central microtubules are linked by a series of regularly spaced linkages (central projections) and are surrounded by a fibrilar central sheath, which constitute the central pair complex (CPC, firstly called central apparatus) of the Ax (Goodenough and Heuser, 1985; Afzelius *et al*., 1995). Microtubule doublets bind to the two central microtubules by radial projections, called radial spokes (RS). Between the RS and both DA, is localized the dynein regulatory complex (DRC), which is the ideally site to mediate signals, either mechanical, enzymatic, or both, between different Ax components (Gardner *et al.*, 1994; Heuser *et al.*, 2009). Recent studies, revealed that DRC forms a continuous connection from the A-tubule to the B-tubule of the microtubule doublet, which aligned to the finding that the DRC is the only structure besides the DA that connects with outer doublets suggest that the DRC is the nexin link (Heuser *et al.*, 2009).

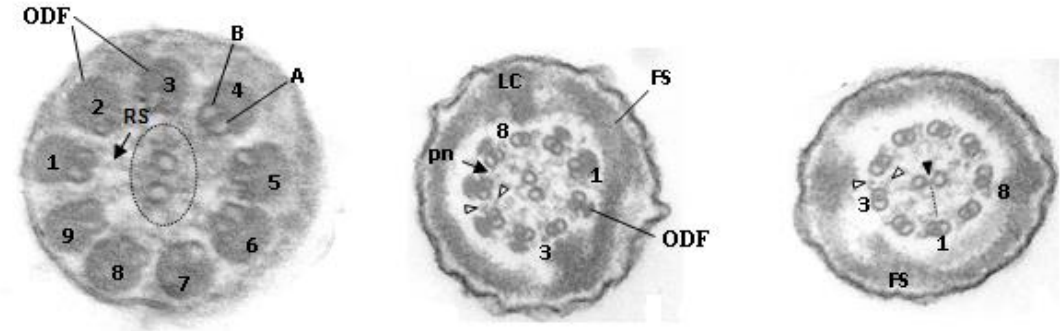


Figure 6. Ultrastructure of the axoneme (Ax). The Ax consists of nine microtubule doublets (1-9) and two central microtubules, forming the 9+2 pattern. Each doublet has an inner peripheral (A) and an outer (B) microtubule and doublets are linked by nexin bridges (pn: between the A and B adjacent pairs). They are connected to the two central microtubules by radial spokes (RS). The microtubule A of each doublet has two dynein arms (white arrowheads). The doublet whose radial projection is perpendicular (dash dotted) to the central microtubules is named No. 1, with the numbering up to 9 following the clockwise direction. The two central microtubules are joined by the central bridge (black arrowhead) and wrapped in a fibrilar sheath (dashed circle). From the neck to the end of the proximal principal piece the outer dense fibers (ODF) are adjacent to the Ax and through all principal piece there are rings of the fibrous sheath (FS).
Image provided by Professor Mário Sousa.

3.3.2 Outer dense fibres (ODF)

The flagellum of the mammalian spermatozoon characteristically possesses 9 ODF positioned around the Ax (Holstein and Roosen-Runge, 1981; Grudzinskas and Yovich, 1995). The ODF are formed during spermiogenesis in acrosomal phase by the SC and are located on the outside of the Ax in the MP and proximal PP of the sperm tail (De Krestser, 1998; Oko, 1998). Each fibre is composed by a medulla (an inner part) that is surrounded by an incomplete cortex (a peripheral area). The medulla is highly stable, composed by several cystein-, serine- and proline-rich intermediate filament-like proteins with a high degree of zinc-dependent disulfide cross-bridging; whereas the cortex is less stable, composed by a single layer of globular subunits (Olson and Sammons, 1980; Grudzinskas and Yovich, 1995).

The ODF have important roles on the flagellum, they provide elasticity (Fawcett, 1970) and structural reinforce, which enhances the tensile strength of the flagellum that is required to overcome the shear forces in the female reproductive tract (Baltz *et al.*, 1990). Anomalies in ODF were reported in spermatozoa of asthenoteratozoospermic men (Haidl *et al*., 1991; Chemes and Rawe, 2003, 2010).

3.3.3 Neck piece (NP)

The NP (also called connecting piece) (Fig. 5) has about 0.5 µm in length and its major components are the *capitulum* (a dense dome-shaped fibrous structure) and the SC (Fawcett, 1970; Grudzinskas and Yovich, 1995).

The base of the nucleus is indented (implantation fossa) and covered by the BP, a layer of electron dense material that contacts with the SC (Pedersen, 1972). The columns are cross-striated with a periodicity of about 6.5 nm between segments, and each segment has nine or ten horizontal striations. The SC are formed

separately, become linked at a late stage in spermiogenesis and the ODF only attach to the distal ends of the SC in the mature spermatozoon. It is believed that the columns are articulated, and may give flexibility to the NP to allow bending without placing stress on the BP link (Grudzinskas and Yovich, 1995).

The centrosome is a cytoplasmic structure that serves as the main microtubule organizing center of the cell as well as a regulator of cell-cycle progression. During formation of cilia, the centrosome replicates and individual centrioles migrate to the cell periphery where they anchorage to the cell membrane to become basal bodies from which Ax originates. In humans, the distal centriole disappears in spermiogenesis, after the sperm Ax being formed, leaving only few remnants in mature spermatozoa. Therefore, the mature spermatozoon has only one functional centriole, the PC, attached to the BP. The main role of the sperm PC is related with the process of fertilization and cell division of the fertilized oocyte. The PC is carried into the oocyte at fertilization, persists throughout sperm nuclear decondensation and organizes the sperm aster. This is a radial array of microtubules that promote pronuclear movement towards the center of the zygote. After that moment it replicates and originates the two asters of the zygote (Sathananthan *et al*., 1996; Nakamura *et al*., 2001; Chemes, 2012).

3.3.4 Midpiece (MP)

MP is about 3.5-5 µm in length and is characterized by the presence of a mitochondrial sheath in form of a helix that surrounds the axonemal complex and the nine ODF (Fig.5) (Holstein and Roosen-Runge, 1981). The mitochondrial helix is kept in form by multiple disulphide bridges composed by a selenium-rich protein that becomes a structural protein in mature spermatozoa (Piomboni *et al.*, 2012).

The human sperm mitochondria are positioned in four helices, and each helix is composed of 11-15 gyres with on average two mitochondria to each gyre. The human sperm mitochondria are more stable than those found in somatic cells, which may help to resist to the stretching and compression of the mitochondria during the flagellar beat (Grudzinskas and Yovich, 1995). Another difference is its ability to use lactate as an oxidative substrate and the presence of specific isoforms of proteins and isoenzymes, such as cytochrome c. These specific functional characteristics adapted sperm mitochondria for sperm motility (Gagnon and de Lamirande, 2006; Piomboni *et al.*, 2012). The mitochondrial inner membrane is the site of energy production for the spermatozoa. However, is an ongoing debate if mitochondria oxidative phosphorylation is enough to generate the energy necessary for sperm flagellar motility. It has been suggested by some reports, that the energy source for sperm motility is mostly given by glycolysis (glycolytic enzymes are linked to the fibrous

sheath, which is near the beating Ax) instead by mitochondria oxidative phosphorylation (whose ATP production is far away from the beating Ax) (reviewed by Ford, 2006; Piomboni *et al*., 2012). Nevertheless, independently of its role in providing energy for sperm motility, mitochondria energy are involved in important cellular process such as cell defence against oxygen reactive species, apotosis and in intracellular and membrane events occurring during spermatogenesis, sperm maturation (in the epididymis), sperm capacitation (during interaction with the endometrium), and fertilization (reviewed by Rajender *et al*., 2010; Amaral *et al.*, 2013).

In the MP it is also found a cytoplasmic droplet that consists of the remains of the residual cytoplasm. The cytoplasmic droplet is formed during the maturation phase of spermiogenesis, accomplished through the actions of Sertoli cells by a process known as “cytoplasmic extrusion”, which occurs before sperm are transported to the epididymis (Rengan *et al.*, 2012). In the tubular lumen of the testis, Sertoli cells extrude and phagocyte most of the germ cell cytoplasm as “residual bodies”, the remnant of which becomes the cytoplasmic droplet. The human spermatozoa retain a small amount of the cytoplasmic droplet around the MP after spermiogenesis, with no negative effects to male fertility. The MP is the major site of water influx and cell volume regulation, thus the vesicles contained by cytoplasmic droplet are important when spermatozoa face hypoosmotic challenges, and therefore is an ideal location for the cytoplasmic droplet (Rengan *et al.*, 2012).

3.3.5 Principal Piece (PP)

The major part of sperm flagellum is the PP with about 44-50 µm in length (Holstein and Roosen-Runge, 1981). The PP is separated from the MP by the An or Jensen’s ring (Fig. 3, 5) that is a traverse ring of dense material found at the end of the mitochondrial sheath. The plasma membrane of the flagellum adhere to the An in the mature spermatozoon, which, in humans, is semicircular in cross section with a curved caudal surface and a plane rostral surface (Fawcett, 1970). The role of An have been speculated for several years, however it was shown that the An is needed to maintain sperm membrane domains (Kwitny *et al.*, 2010).

The PP is distinguished from the other parts of flagellum by the presence of the FS, a cytoskeletal structure surrounding the Ax and the ODF. At the distal part of the PP the FS becomes more compact and smaller until completely disappear. At this time, the EP begins (Holstein and Roosen-Runge, 1981).

The FS (Fig. 3, 5) is a unique characteristic of the spermatozoon and consists of two peripheral longitudinal columns, which are at the plane of the central

microtubules, connected together by a series of ribs. The ribs are composed of closely packed filaments and form a ring around the Ax (Holstein and Roosen-Runge, 1981; Grudzinskas and Yovich, 1995). The longitudinal columns are attached to ODF 3 and 8 in the proximal part of the PP. When the ODF disappears, these columns become associated with microtubule doublets 3 and 8. The FS is formed at spermiogenesis acrosomal phase and its assembly proceeds from distal to proximal along the Ax (Oko, 1998; Eddy *et al.*, 2003).

The proteins that compose the FS have extensive disulphide bonding, which makes the whole structure extremely stable and resistant. These disulphide bonds are also important to stabilize the FS when sperm traverse the epididymis and undergo maturation (Eddy *et al.*, 2003).

Some important roles have been given to the FS (reviewed by Eddy *et a*l., 2003). It is believed that the FS influences and modulates the bending and the flagellar beat and is responsible for docking key components involved in signal transduction pathways, such as cAMP-signalling pathway, which are important to regulate processes that lead to fertilization. In addition, the FS anchors glycolytic enzymes that provide energy for motility of sperm (Eddy *et al.*, 2003; Krisfalusi *et al.*, 2006). FS anomalies were associated to male infertility (Baccetti *et al,* 2005a, b; Escalier and Albert, 2006, Chemes and Rawe, 2010)

3.3.6 End Piece (EP)

In human spermatozoa, the EP has about 6 µm in length and begins when the FS completely disappears. Its only component is the Ax that, in the proximal part of the EP, generally maintains its structure, and is surrounded by the plasma membrane. However in the distal part, only single tubules in Ax are present (Holstein and Roosen-Runge, 1981). Analysing the EP from the beginning to the end, the first structure that disappear are the arms of the doublets. At the same plane, the subunit A becomes hollow. Then, the two central microtubules terminate and two of the doublets move to the middle and are surrounded by seven spaced doublets. At the end, the EP becomes closer, is possible to observe a separation of the doublets and the beginning of the opening of the B subunits. This is followed by a successive termination of the remaining tubules as the end of the flagellum is reached (Pedersen, 1972).

4. MOLECULAR COMPONENTS OF THE SPERM FLAGELLUM

In the previous section the general structure and morphology of human spermatozoa was focused. In this section we will give emphasis to the molecular components of the sperm flagellum, specifically to the main proteins and encoded genes present in the different structures of flagellum associated to sperm motility. The molecular composition of flagellum components has been studied mainly in sperm from marine invertebrates (such as sea urchins and tunicates) and in protists (such as *Chlamydomonas and Tetrahymena*) (Dutcher, 1995; Inaba, 2003, 2011). The molecular components that will be addressed here are merely a small portion of all molecular components and interactions that may exist in the complex sperm flagellum. There is still a long journey to make in order to fully understand all molecular components and mechanisms that compound and are responsible for the assembly of the sperm flagellum.

4.1. Axoneme

The Ax (Fig. 6,7) is the flagellar motor of the sperm cells and the studies performed in marine invertebrates show that it is composed by approximately 250 proteins (Dutcher, 1995). Despite the basic structure of the sperm Ax being quite conserved and similar, some structures, such as ODA and central pair projections, are different between *Chlamydomonas* and metazoan sperm (Inaba, 2007, 2011).

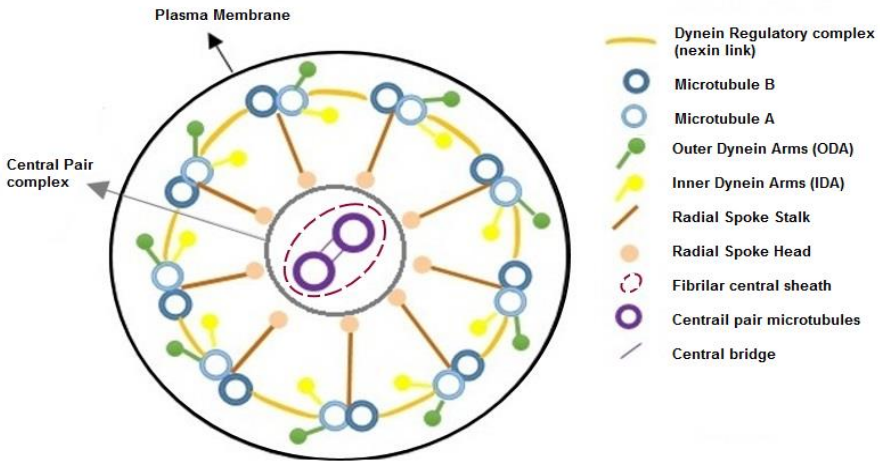


Figure 7. A schematic drawing of the sperm axoneme. The legend of the sperm axoneme components is placed on the right. For more information about its structure see section 3.3.1

The Ax is a sophisticated structure with a cytoskeleton, motors proteins, molecular chaperones, regulatory elements such as Ca^{2+}, binding proteins and protein kinases/phosphatases (reviewed by, Inaba, 2003).

4.1.1 Tubulin

The main structural component of microtubules is tubulin (alpha and beta tubulin), constituting about 70% of the protein mass in an Ax. They belong to a small family of globular proteins and are linked with a wide range of somatic cell functions, such as maintaining structure of the cell, movement of secretory vesicles, organelles, and intracellular substances and in cell division (Dutcher, 1995). The genes of α- and β-tubulin are highly conserved within and among species, with α- and β-tubulins sharing at least 60% of amino acid identity. The α- and β-tubulins are also similar between themselves as α -tubulins share 36–42% identity with β-tubulins. Besides α- and β-tubulins, the tubulin superfamily is composed by γ-, δ-, ζ- and ε- tubulins. Tubulins α-, β- and γ- are ubiquitous in eukaryotes and the genes that encodes this proteins are essential for microtubule polymerization. Tubulins δ- and ε- are widespread but not ubiquitous, whereas ζ-tubulin has been found only in kinetoplastid protozoa. However, knowledge about their role is still scarce (reviewed by Oakley, 2000; Dutcher, 2001).

Tubulin molecules are arranged in rows to form protofilaments, which are then aligned side by side to form the microtubule walls. In doublets of the Ax, the A tubule is composed of 13 protofilaments and the incomplete B tubule is made up of 10 protofilaments (Afzelius *et al.*, 1995). Laterally and throughout most of the microtubule, α-tubulin molecules interact with other α-tubulin molecules, so does, β-tubulin molecules. These laterally self-interaction of α-tubulin and β-tubulin molecules is necessary to the combination of the protofilaments that compose the microtubules (Oakley, 2000). Tubulin is often subjected to post-translational modifications, for example acetylation, palmitoylation, phosphorylation, polyglutamylation and polyglycation (reviewed by Rosenbaum, 2000; Hammond *et al.*, 2008), that are important for proper binding and assembly of axonemal substructures to microtubules and for motility. For instance, it was demonstrated that the polyglutamylation of α-tubulin plays a dynamic role in a dynein-based motility process (Gagnon *et al.*, 1996).

4.1.2 Dyneins

Dyneins are motor proteins that convert the chemical energy contained in ATP into the mechanical energy of movement. Dyneins can be divided into two groups: cytoplasmic dyneins and axonemal dyneins (or flagellar dyneins). The axonemal dyneins are key elements to motility of eukaryotic cilia and flagella.

The organisation of human ODA has not been examined but in other metazoan sperm, the ODA is comprised of two heavy chains (HC) of about 500 kDa; three to five intermediate chains (IC) of about 120–60 kDa; and six light chains (LC) with approximately 30–80 kDa each (Inaba, 2003). In its turn, the IDA is more

complex. In *Chlamydomonas* it was observed that the IDA has eight distinct inner arm HC, which are organized with various IC and LC into seven different molecular complexes, one two-headed isoform and six single-headed isoforms (Porter, 1996).

The **HC** contains the motor machinery that is responsible for transducing chemical energy into directed mechanical force applied to the microtubule surface, possessing the sites of both ATP hydrolysis and ATP-sensitive microtubule binding (Asai and Koonce, 2001; Burgess and Knight, 2004). Each HC comprises three domains: **(1)** the N-terminal domain (also called stem domain), which may be involved in binding of the dynein to microtubules; **(2)** the C-terminal domain (also called a globular head), that includes a highly conserved central section with six AAA modules (AAA for: ATPases associated diverse cellular activities), considered to form a globular subdomain; and **(3)** the microtubule-binding domain (that forms the stalk domain).
The N-terminal domain is connected to the C-terminal globular head. Regarding the AAA motifs, the first four possess P-loop motifs that are well conserved in all dynein sequences. The first P-loop-P1, is the site of ATP hydrolysis that leads to movement; the AAA modules 2 to 4 are non-hydrolytic and possess intact P-loops that possibly are capable of binding nucleotides. The two last AAA domains lack P-loops and are believed that consequently do not bind nucleotides. Between the 4 and 5 AAA modules is located the B-link, which is the functional contact between dynein and the microtubule (King, 2000; Asai and Koonce, 2001; Burgess *et al*., 2003; Burgess and Knight, 2004). 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).

The **IC** and **LC** help to specify the intracellular location of the dynein, regulates its motor activity and are also a source of diversity between species (Asai and Koonce, 2001, Burgess and Knight, 2004). The IC participates in structural attachment of the DA to flagellar microtubules (King, 2000). Relatively to the LC, several functions have been recognised, such as redox-sensitive vicinal dithiols, Ca^{2+} binding (King, 2000) and intraflagellar transport (Pazour *et al.*, 1998). It is known that phosphorylation/dephosphorylation of IC and LC leads to an alteration in motor activity (King, 2000). The variety of structure and function of these chains indicates that many regulatory mechanisms are present and needed for the proper sperm motility.

Multiple dynein genes are found in the genomes of organisms with motile cilia and flagella. Phylogenetic analyses classify these into several groups, each of which may be associated with a specific function. Five human genes are known that encode

ODA dynein heavy chain (DHC) genes: *DNAH5* (Gene ID: 1767, 5p15), *DNAH8* (Gene ID: 1769, 6p21), *DNAH9* (Gene ID: 1770, 17p12), *DNAH11* (Gene ID: 8701, 7p21) and *DNAH17* (Gene ID: 8632, 17q25.3) (Pazour *et al.*, 2006; Yagi, 2009). Relatively to the IDA, there are eight human genes: *DNAH1* (Gene ID: 25981, 3p21.1), *DNAH2* (Gene ID: 146754, 17p13.1), *DNAH3* (Gene ID: 55567, 16p12.3), *DNAH6* (Gene ID: 1768, 2p11.2), *DNAH7* (Gene ID: 56171, 2q32.3), *DNAH10* (Gene ID: 196385, 12q24.31), *DNAH12* (Gene ID: 201625, 3p14.3), *DNAH14* (Gene ID: 127602, 1q42.12) (Pazour *et al.*, 2006; Yagi, 2009). The IC and LC chains are thought to contain at least 5 genes, including *DNAI1* (Gene ID: 27019, 9p13.3), *DNAI2* (Gene ID: 64446, 17q25), *DNAL1* (Gene ID: 83544, 14q24.3), *DNAL4* (Gene ID: 10126, 22q13.1), *DNALI1* (Gene ID: 7802, 1p35.1) and *TXNDC3* (Gene ID: 51314, 7p14.1) (National Center for Biotechnology Information (NCBI) database-accessed in December 2013). Phylogenetic analysis supports the assignment of DNAI1 and DNAI2 as outer dynein arm IC (Pazour *et al.*, 2006).

These chains regulate dynein activity, consequently mutations may result in abnormal ciliary ultrastructure and function. Mutations in some of these genes were already associated to syndromes such as primary ciliary dyskinesia (PCD, see Table I of the following section 5.1).

4.1.3 Dynein Docking Complex and Dynein Regulatory Complex (DRC)

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 (24nm). The ODA-DC contains three polypeptides (DC1-DC3) (Takada *et al.*, 2002). The DC1 and DC2 polypeptides are quite similar. Both have three regions, totalling 236 and 178 amino acids, correspondently, that are >99% likely to form a coiled-coil structure. These two proteins potentially determine the spacing of the ODA (Takada *et al.*, 2002). The DC3 polypeptide has four motifs that possibly bind Ca^{2+}, has important roles in the regulation of the ODA and also plays a role in calcium-regulated ODA activity (Casey *et al.*, 2003).

The DRC functions of dynein regulation and limitation of doublet sliding (Heuser *et al.*, 2009), and is composed of six axonemal proteins (Piperno *et al.*, 1992). Studies using DRC mutants showed that some components of the DRC serve primarily to regulate activity, while others play a role in mediating structural interactions between the DA, the A-tubule of the outer doublet, and the RS (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 (Gardner *et al.*, 1994).

As far as we known, only three DRC components have been cloned and characterized at the molecular level: the *DRC4* in *Chlamydomonas* (Rupp and Porter, 2003); *CMF70* in *Trypanosoma brucei* (Kabututu *et al.*, 2010) and the *DRC1* also in *Chlamydomonas* (Wirschell *et al.*, 2013).

4.1.4 Radial spokes (RS) and Central pair complex (CPC)

The CPC is composed by two structurally asymmetric and biochemically distinct central microtubules, linked by central projections and surrounded by a fibrilar sheath. The central pair interacts with the RS, which by its turn are attached to each microtubules doublets (Fig. 7). The RS and CPC are essential for the regulation of DA. For instance, it was 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 (Smith and Lefebvre, 1997;Smith and Yang, 2004). In addition, central pair may function like a distributor to provide a local signal to the RS that selectively activates subsets of DA (Omoto *et al.*, 1999).

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). Some have been already cloned and important functions have been described. For example, RSP3 anchors the RS to the outer microtubule doublet, contains an AKAP (for A-kinase anchoring protein) domain and binds in vitro the cyclic AMP-dependent protein kinase (PKA) regulatory subunit (Gaillard *et al.*, 2001). This could explain the fact that the regulation of DA by RS/CPC involves events of phosphorylation/dephosphorylation (Porter and Sale, 2000).

In humans it has been already described at least seven radial spoke proteins and its encoding genes are: *RSPH1* (Gene ID: 89765, 21q22.3), *RSPH3* (Gene ID: 83861, 6q25.3), *RSPH4A* (Gene ID: 345895, 6q22.1), *RSPH6A* (Gene ID: 81492, 19q13.3) *RSPH9* (Gene ID: 221421, 6p21.1) *RSPH10B* (Gene ID: 222967, 7p22.1) and *RSPH10B2* (Gene ID: 728194, 7p22.1) (NCBI database http://www.ncbi.nlm.nih.gov/gene/ and Uniprot database, http://www.uniprot.org/, accessed July 2014). RSPH1 gene encodes a RS-head protein and is mainly expressed in respiratory and testis cells. It is important for the proper building of CPC and RS, since mutations in the *RSPH1* leads to an abnormal axonemal configuration with CPC and RS defects (Kott *et al*., 2013; Onoufriadis *et al*., 2014). Castleman and collaborators found mutations in the RS genes *RSPH4A* and *RSPH9* associated to anormalies in CPC (Castleman *et al.,* 2009).

The **CPC** in *Chlamydomonas* flagella also contains approximately 23 polypeptides (1-350 kDa) (Smith and Lefebvre, 1997) and possesses similar domains for molecular assembly and signalling, such as AKAP240 (Gaillard *et al.*, 2001) and FAP221, whose mammalian orthologue is the CaM-binding protein primary ciliary dyskinesia protein 1 (PCDp1). The PCDp1 is known to coordinate the activity of specific dynein isoforms in the control of ciliary motility (DiPetrillo and Smith, 2010, 2011).

4.2. Outer dense fibres (ODF)

The ODF in rat spermatozoa contains at least 14 polypeptides (Oko, 1988), whereas the human ODF consist of about 10 major proteins and of at least 15 minor proteins (Petersen *et al.*, 1999).

Some of its protein encode genes were already identified, and at least four (*ODF1* to 4 genes) have been cloned and characterized in humans. For instance, the human gene that encodes one of the main ODF protein is the *ODF1* (OMIM: 182878, 8q22) and it is only expressed in testis. The protein codified by its gene (ODF1 protein) contains leucine zipper motifs which appear to be responsible for self-interaction and interaction with other ODF components (Shao *et al.*, 1997). The N-terminal leucine zipper motif of the ODF1 protein, allowed the isolation of the interacting proteins ODF2 (firstly called Odf84) (Shao *et al.*, 1997), SPAG4 (Shao *et al.*, 1999a) and SPAG5 (Shao *et al.*, 2001). The human *ODF2* gene (Gene ID: 4957, 9q34) is ubiquitously expressed in many cell types as various splicing variants (Shao *et al.*, 1999b, Schweizer and Hoyer-Fender, 2009). ODF2 protein has been reported to play several important roles such as: formation of primary cilia (Soung *et al.*, 2009) and sperm flagella (Tarnasky *et al.*, 2010). It is also involved in the initial docking of centrioles to membrane (Hoyer-Fender, 2010). Moreover, is important to the basal foot formation and for polarized alignment of basal bodies and coordinated ciliary beating (Kunimoto *et al.*, 2012) and for pre-implantation development (Salmon *et al.*, 2006). The *ODF3* gene (Gene ID: 113746, 11p15.) encodes the ODF protein 3, localized in the flagella of elongated spermatids and along the entire length of the tail in mature sperm (Carvalho *et al.*, 2002). The *ODF4* gene (Gene ID: 146852, 17p13.1) encodes the ODF protein 4, expressed only in the testis (Kitamura *et al.*, 2003).

4.3. Centriole

Centrioles are a small cylindrical structure that together with the pericentriolar-material forms the centrosome, the major microtubule organising centre in animal cells. Among their important cellular roles, are the formation of Ax and flagella (Bettencourt-Dias and Glover, 2007).

PLK4 and SAS6 proteins are master regulators of centriole duplication. Its absence leads to a lack of centriole duplication, whereas an overexpression leads to an increase in the number of microtubule organising centre (Bettencourt-Dias *et al*., 2005; Habedanck *et al*., 2005; Kitagawa *et al*., 2009). The proteins CEP135, SAS5 (or STIL), CP110, Centrin and Ana1/2/3 are also important for procentriole and centriole biogenesis and the CPAP protein plays an important role in centriole elongation (reviewed by Bettencourt-Dias and Glover, 2007; Brownlee and Rogers, 2013). Mutations in some of these regulators or in other components involved in centriole formation/function/duplication may lead to centriolar and centrosomal abnormalities, which by its turn, are involved in male infertility or in abnormal development of the embryos. Defects in centrosomes, have been reported to cause the anomalies in cilia in patients suffering from ‘‘ciliopathies’’, a group of disorders of ciliated cells, such as PCD. These disorders are also associated to male sperm immotility (reviewed by Chemes, 2012).

4.4. Mitochondria

Mitochondria contain about a thousand of distinct proteins involved in various metabolic pathways (Sickmann *et al*., 2003; Pagliarini *et al*., 2008). Despite the established role of mitochondria in energy production (Cooper, 2002), only 14% of its proteins act directly in energy metabolism. The majority are involved in maintaining and expressing the mitochondrial genome, in the transport of metabolites and in metabolism of amino acids, lipids, and iron (Sickmann *et al.*, 2003). The crucial core machinery of mitochondrial gene expression consists of the mitochondrial transcription factor A and B2 (TFAM and TFB2, respectively) and the RNA polymerase. The TFAM has also a role in maintenance and packaging of mitochondrial DNA (mtDNA) (Kang *et al.*, 2007; Kaufman *et al.*, 2007). Mitochondrial biogenesis, respiratory function and control of the expression of genes encoding for cytochromes-c and cytochromes-c-oxidase subunits (vital for ATP production) are governed by nuclear factors, such as the nuclear respiratory factors 1 and 2 (NRF1 and NRF2). NRF1/NRF2 are regulated by the PGC-1α, thus it is the master regulator of mitochondrial biogenesis (Hock and Kralli, 2009; Scarpulla, 2011; Wenz, 2013).

Mitochondrial genetic diseases can result from defects in mtDNA in the form of deletions, point mutations or depletion, which, ultimately, causes loss of oxidative phosphorylation. They may be spontaneous, maternally inherited or a result of inherited nuclear defects in genes that maintain the mtDNA (Copeland, 2008). Mutations of the mtDNA in spermatogenesis cause male infertility because of its important role in energy production for mitosis, meiosis, sperm maturation,

capacitation and motility (Rajender *et al.*, 2010). It was observed that high levels of mutant mtDNA are correlated to sperm immotility (Spiropoulos *et al*, 2002). An explanation could be that, the high levels of mutant mtDNA might increases the likelihood of reactive oxygen species that damage the mtDNA, and thus causing potentially deleterious effects on sperm function (Tremellen, 2008).

4.5. Annulus (An)

The An, in humans, is semicircular in cross section with a curved caudal surface and a plane rostral surface (Fawcett, 1970). The An is needed to maintain sperm membrane domains and its absence produces an interruption in the cytoskeleton at the MP-PP junction (Kwitny *et al.*, 2010).

Septins are essential structural components of the human and mouse An (Mostowy and Cossart, 2012), forming a ring that forms a diffusion barrier required for the mechanical and structural integrity of the sperm (Ihara *et al.*, 2005; Kwitny *et al.*, 2010).

Until now there have been identified 14 human septins, with SEPT1, SEPT4, SETP6, SEPT7 and SEPT12 being located at the An (Ihara *et al.*, 2005; Steels *et al.*, 2007; Lin *et al.*, 2011). Using mouse knockouts it was showed that SEPT4 is essential for the structural and mechanical integrity of the spermatozoon, including for proper mitochondrial architecture and establishment of the An (Ihara *et al.*, 2005; Kissel *et al.*, 2005). Protein SEPT7 is involved in the regulation of sperm maturation and sperm morphology (Chao *et al.*, 2010). Consequently SEPT4/7 were suggested to be used as biomarkers for monitoring the status of spermiogenesis (Sugino *et al.*, 2008). In humans, SEPT12 is expressed only in spermatozoa, namely in the head, neck and An (Steels *et al.*, 2007; Lin *et al.*, 2011, 2012). Important roles in spermiogenesis, including sperm nuclear integrity and tail development were attributed to SEPT12 (Lin *et al.*, 2009, 2012; Miyamoto *et al.*, 2012; Kuo *et al.*, 2013).

Tat1 is an anion transporter of the SLC26 family and is another critical component of the sperm An. It is specifically expressed in male germ cells and co-expressed with Sept4. This anion transporter is thought to be essential for proper sperm tail differentiation and motility (Touré *et al.*, 2007; Lhuillier *et al.*, 2009).

4.6. Fibrous Sheath (FS)

The FS is a cytoskeletal structure exclusive of the PP of sperm and surrounds the Ax and ODF. The FS is believed to influence the degree of flexibility, plane of flagellar motion and the shape of the flagellar beat.

Although the majority of studies on FS proteins were performed in rodents, several proteins were already identified in the human FS too (Beecher *et al.*, 1993;

Kim *et al.*, 1995, 1997; Turner *et al.*, 1998; Vijayaraghavan *et al.*, 1999; Naaby-Hansen *et al*., 2002).

Three of the main FS proteins are of the AKAP family (Eddy *et al.*, 2003). AKAP are scaffolding molecules that organize molecular complexes whose function is determine the precise location and timing of signal transduction events, and combine signal transduction and signal termination molecules in order to modulate signalling pathways. AKAP proteins sequester enzymes such as protein kinases and phosphatases with appropriate substrates to the coordination of phosphorylation and dephosphorylation events (Langeberg and Scott, 2005). The *AKAP4* gene is expressed in the post-meiotic phase of spermatogenesis and encodes an AKAP4 protein, which is the most abundant protein and a major structural component of FS. It is restricted to the PP of the flagellum, has two regions of homology with PKA-anchoring domains (a characteristic of AKAP proteins) and plays a major role in completing FS assembly. It is also important for sperm motility (Turner *et al*., 1998; Brown *et al*., 2003; Eddy *et al*., 2003). The human *AKAP3* gene encodes the second major protein of the FS (Vijayaraghavan *et al.*, 1999), the AKAP3 protein. It has a single PKA binding site and was detected only in the circumferential ribs. It was proposed that *AKAP3* gene is involved in organizing the basic structure of the FS (Mandal *et al.*, 1999; Brown *et al.*, 2003; Eddy *et al*., 2003).

The FS is composed by many more proteins, such Ropporin; Rhophilin and “Calcium-binding tyrosine phosphorylation regulated protein” (Eddy *et al.,* 2003), which are extremely important for FS assembly and function, and thus for sperm motility. Other FS proteins, as a mu-class glutathione-s-transferase, have raised the possibility that the FS could also have a role in protect sperm from oxidative stress (Fulcher *et al.*, 1995).

Another important class of FS proteins are the glycolytic enzymes. Some studies have suggested that the mitochondrial energy production may not be enough to sustain the sperm motility and propose glycolyis as a main method of energy production for flagellar motility (Ford, 2006; Piomboni *et al*., 2012). Most of the glycolytic enzymes are localised in PP and glycolysis seems to be restricted to the PP. Some of the glycolytic enzymes, including two spermatogenic cell-specific forms (glyceraldehyde 3-phosphate dehydrogenase (GAPD) and hexokinase 1), are tightly associated with the FS (Eddy *et al.*, 2003; Turner, 2003; Ford, 2006). Another study showed that male mice lacked progressive motility when the gene for sperm specific *GAPD* has been ‘knocked out’, demonstrating the importance of the glycolysis and the glycolytic enzymes of FS to sperm motility (Miki *et al.*, 2004).

5. FLAGELLAR ABNORMALITIES AND GENETIC BASES OF SPERM IMMOTILITY IN HUMANS

Sperm motility is crucial to natural fertilization occur. Sperm only becomes fully motile after being exposed to several of external and intracellular factors during its journey through epididymis and female reproductive tract (Luconi *et al.*, 2006; Aitken *et al.*, 2007). Flagellum motility is a consequence from ondulatory waves propagating backwards that create forward propulsive thrust along the axis of the flagellum. It is created by the motor activities of the axonemal DA working against the stable microtubule doublets. Signals are transmitted from the CPC through the RS to the DRC whose regulate the dynein activity. The flagellar beat starts with the phosphorylation of the dynein, which activates the ATPase and starts the conversion of chemical energy from ATP hydrolysis into mechanical energy for motility. In the presence of ATP, dyneins produce a unidirectional force towards the minus end of microtubules to generate sliding forces. Consequently, one DA (first dynein) binds to the B tubule of the adjacent outer doublet and generates a downward stroke, resulting in the adjacent microtubules sliding past between pairs of outer doublets. This sequence will be repeated along the length of the microtubules resulting in the propagation of the flagellar bend for all flagellum (Turner, 2003; Gagnon and de Lamirande, 2006; Luconi *et al.*, 2006). Therefore, DA activity must be precisely coordinated both along the length and around the circumference of the flagellum (Heuser *et al.*, 2009).

The sperm motility is thus, a highly complex process with several structural and molecular elements and metabolic pathways involved. Due to this highly complexity, any alteration in external and/or internal factors regulating sperm motion as well as in cellular structure and metabolism involved in generating flagellar beat may result in defects in sperm motility, which consequently results in male infertility.

Asthenozoospermia is the medical term for reduced sperm motility and is one of the main male pathologies underlying infertility (Curi *et al.*, 2003; Ortega *et al.*, 2011). The aetiology of asthenozoospermia is not simple to unravel and often remains unexplained. Ultrastructural defects in the sperm flagellum caused by genetically inherited and congenital defects (Chemes *et al.*, 1998; Chemes, 2000) and necrozoospermia (low percentage of live, and high percentage of immotile, spermatozoa in the ejaculate) are the main causes of asthenozoospermia (Curi *et al.*, 2003; Ortega *et al.*, 2011).

Mitochondria provide energy for spermatogenesis, sperm maturation and capacitation, and are as well involved in cell differentiation, ROS generation,

apoptosis and calcium signalling (Amaral *et al.*, 2013). Therefore, dysfunctions of the human mitochondrial sheath and mutations mtDNA are also a significant cause of asthenozoospermia (Spiropoulos *et al*., 2002; Rajender *et al*., 2010; Piomboni *et al*., 2012). In addition, the lack of the An and disorganization of the MP-PP junction are associated with human asthenozoospermia (Ihara *et al.*, 2005; Lhuillier *et al.*, 2009).

In mouse, many mutations in genes coding for varied products as flagellar transport proteins, structural, motor and signalling proteins are known to cause anomalies in the sperm flagellum leading to motility disorders (Escalier, 2006; Yatsenko *et al.*, 2010; Inaba, 2011). Unfortunately, in humans a strict association between mutations in some genes and alteration in sperm motility are still very scarce. However, due to the high degree of conservation of many of these genes among mice and humans, some genes, isolated or associated to syndromes, already have been proved to be responsible for some cases of human infertility associated with poor sperm motility (Inaba, 2011). Two main disorders are associated with poor sperm motility: Primary Ciliary Dyskinesia (PCD) and Dysplasia of the Fibrous Sheath (DFS).

5.1. Primary Ciliary Dyskinesia (PCD)

PCD (also called immotile cilia syndrome), was first described more than a century ago by Afzelius (Afzelius *et al.*, 1975; Afzelius, 1976). PCD (OMIM: 244400) is a genetically heterogeneous, autosomal recessive disease that is characterized by a generalized paralysis of ciliated cells, including sperm and respiratory cilia, resulting in recurrent infections of the respiratory tract.

Kartagener syndrome (KS), is characterized by the combination of *situs inversus* (reversal of the internal organs), chronic sinusitis, and bronchiectasis, and occurs in about 50% of PCD. Most men with PCD have nearly 100% immotile spermatozoa and thus are infertile. The estimated incidence of PCD is approximately 1 per 15,000 births, but the prevalence of PCD is difficult to determine, mainly due to difficulties related to the diagnosis, ranging from 1/2,000 to 1/40,000 (reviewed in Boon *et al*., 2013; Lie and Ferkol, 2007). In the majority of cases, the results of electron microscopic analysis of sperm cells reveal that microtubule doublets of the Ax lack DA. In addition, it was also found in PCD patient's absence or dislocation of the central microtubules, defects of RS and doublets abnormalities (Afzelius and Srurgess, 1985; Afzelius *et al*., 1995; Chemes and Rawe, 2003).

Given that the typical diagnostic of PCD is the absence of DA, the investigations into the genetic basis of PCD have been focused on DA proteins and several genes are known to be associated with PCD (Table I). The first gene in which mutations were found to be associated with PCD was *DNAI1*, which is an axonemal

dynein IC gene, found in the ODA, and localized on 9p13-p21 (Pennarun *et al.*, 1999). Mutations of *DNAI1* have been identified in a patient with PCD/KS, whose ODA are often, but not always, absent or shortened (Pennarun *et al.*, 1999; Guichard *et al.*, 2001; Zariwala *et al.*, 2006). This gene has been one of the most studied, although a lower occurrence was described (about 10% of PCD patients) (Failly *et al.*, 2008; Djakow *et al.*, 2012). The gene *DNAH5*, localized at 5p15.2, comprises 79 exons and encodes a HC of the ODA. Defects of the ODA were found associated with *DNAH5* mutations in patients with PCD (Olbrich *et al.*, 2002; Hornef *et al.*, 2006). Patients without *DNAH5* mutations did not show ODA defects neither isolated IDA defects (Hornef *et al.*, 2006). These data suggest that *DNAI1* and *DNAH5* genes are important to the function of the ODA complex. Mutations in the *DNAH5* gene are comparatively more frequent in PCD cases than mutations in *DNAI1*, although mutations in both genes are present in up to 38% of all patients with PCD (Geremek and Witt, 2004; Failly *et al.*, 2008, 2009; Djakow *et al.*, 2012; Boon *et al.*, 2013).

The human *CCDC39* gene is localized at 3q26.33 and encodes a protein that was shown to be essential for the assembly of IDA and of the DRC. Thus, mutations in *CCDC39* results in failure to correctly assemble of IDA complexes, the DRC, the CPC and the RS, thus causing axonemal disorganization and dyskinetic beating (Merveille *et al.*, 2011). The *CCDC40* gene, located at 17q25.3, encodes a CCDC40 protein. Mutations in *CCDC40* gene were found in subjects with PCD whose microscopic analyses showed defects in several axonemal structures, 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 the ODA appeared normal (Becker-Heck *et al.*, 2011). Moreover, *CCDC40* appears to be required for axonemal recruitment of *CCDC39* gene. It also physically interacts with the other axonemal components and may serve as a part of the Ax structural scaffold, possibly as a new DRC component (Becker-Heck *et al.*, 2011). CCDC39 and CCDC40 proteins are considered integral components of the DRC and play a role in IDA attachment (Becker-Heck *et al.*, 2011; Merveille *et al.*, 2011). A recent study, using next generation sequencing (NGS) had identified mutations in both genes among 69% of individuals with PCD with IDA and CPC defects that are indistinguishable by microscopy (Antony *et al.*, 2013).

Table I. List of genes known to be associated to PCD and main axonemal ultrastructural defects showed in PCD patients

Gene	Gene Location	Molecular Function	Main ultrastructural defect	Key reference
ARMC4	10p12.1-p11.23	Axonemal docking and targeting of ODA components	Marked reduction of ODA	(Hjeij *et al.*, 2013)
DNAAF3	19q13.4	Assembly of axonemal IDA and ODA and dynein complexes	Absence of ODA and IDA	(Mitchison *et al.*, 2012)
C21orf59	21q22.1	DA assembly	Absence of both ODA and IDA components	(Austin-Tse *et al.*, 2013)
CCDC103	17q21.31	Fundamental factor for DA binding to cilia microtubules	Partial loss of ODA complexes	(Panizzi *et al.*, 2012)
CCDC114	19q13.33	Component of the ODA docking complex	Absence of ODAs	(Onoufriadis *et al.*, 2013)
DRC1	2p23.3	Regulation of the dynein motors	Severe defects in assembly of the N-DRC structure	(Wirschell *et al.*, 2013)
CCDC39	3q26.33	Assembly of DRC and IDA complexes	Incorrect assembly od IDA, DRC, CPC and RS	(Merveille *et al.*, 2011) (Antony et al., 2013)
CCDC40	17q25.3		Misplacement of the doublets and CPC; abnormal RS and defective assembly of IDA and DRC	
DNAAF1 (LRRC50)	16q24.1	Pre-assembly and/or targeting of DA complexes	Marked reduction of both ODA and IDA	(Loges *et al.*, 2009)
DNAAF2 (KTU)	14q21.3	Pre-assembly of DA complexes	Absence or defects of ODA and IDA	(Omran *et al.*, 2008)
DNAH11	7p21	Encodes a ciliary ODA protein	Normal axonemal ultrastructure	(Knowles *et al.*, 2012)
DNAH5	5p15.2	Important for function of the ODA complex	Absence of ODA	(Olbrich *et al.*, 2002)
DNAI1	9p13.3			(Pennarun *et al.*, 1999)
DNAI2	17q25	Assembly of ODA complexes	ODA defects	(Loges *et al.*, 2008)
DNAL1	14q24.3	Involved in the interaction of the axonemal DLC1 with DHC and tubulin	Absence or markedly shortened ODA.	(Mazor *et al.*, 2011)
DYX1C1	15q21.3	Important for axonemal dynein assembly	Disruptions of ODA and IDA	(Tarkar *et al.*, 2013)
HEATR2	7p22.3	Assembly or transport of DA	Absence of DA	(Horani *et al.*, 2012)
HYDIN	16q22.2	N.D.	Defects in CPC	(Olbrich *et al.*, 2012)
LRRC6	8q24.22	Assembly/ transport and transcriptional regulation of DA and dynein proteins	Absence or defects of ODA and IDA	(Horani *et al.*, 2013)
RSPH1	21q22.3	A radial-spoke-head protein	CPC and RS defects	(Kott *et al.*, 2013)
RSPH9	6p21.1	Components of the RS head	Abnormalities in CPC	(Castleman *et al.*, 2009)
RSPH4A	6q22.1.			
SPAG1	8q22.2	Assembly and/or trafficking of the axonemal DA	Defects in ODA and IDA	(Knowles *et al.*, 2013)
NME8 (TXNDC3)	7p14.1	Critical role in the ODA due to its ability to bind to the MTS.	Partial lack and reduction of ODA	(Duriez *et al.*, 2007)
ZMYND10	3p21.3	Required for IDA and ODA assembly	Absence of ODA and IDA	(Moore *et al.*, 2013)

CPC- Central Pair Complex; DA- Dynein Arms; DLC- Dynein Light Chain; DHC- Dynein Heavy Chain; ODA- Outer Dynein Arms; IDA- Inner Dynein Arms; DRC- Dynein Regulatory Complex; N-DRC- Nexin-Dynein Regulatory Complex; N.D. Non-defined; RS-radial spokes

5.2. Dysplasia of the Fibrous Sheath (DFS)

Dysplasia of the Fibrous Sheath (DFS), also called stump tail syndrome, is one of the most severe abnormalities of the sperm flagellum and causes extreme asthenozoospermia (Chemes *et al.*, 1998; Chemes and Rawe, 2003, 2010). Marked hyperplasia and disorganization of the FS is the typical diagnostic finding in these cases. In addition, under light microscopy, the majority of sperm from affected individuals have short, thick, irregular flagella with no clear distinctions among the MP, PP and EP. From electron microscopy analysis it is also observed partial or total lack of DA, absence of the CPC (in about half of the cases), absence of a normal An and disassemble of the mitochondrial sheath. As a consequence, the hypertrophic FS extends up to the sperm neck (Chemes *et al.*, 1998; Chemes, 2000; Baccetti *et al.*, 2005a, 2005b; Chemes and Rawe, 2010, 2003; Moretti *et al.*, 2011). Although only occasionally associated with lack of IDA/ODA, DFS is considered a variant of PCD.

In humans, the AKAP3 and AKAP4 proteins are the most abundant structural proteins of the FS (see previous section 4.6).

To date, there is no therapeutic procedures to DFS phenotype, which aligned to the familial incidence and association with dynein deficiency in respiratory cilia, suggests that the DFS may have an autossomic recessive inheritance (Chemes and Rawe, 2010, 2003), but it is not clearly yet. Until now only mutations in *AKAP3* and *AKAP4* genes have been associated to the DFS phenotype (Chemes *et al.*, 1998; Baccetti *et al.*, 2005a, 2005b; Moretti *et al.*, 2007). Miki and collaborators, demonstrated that male mice lacking AKAP4 protein were infertile, had reduced sperm motility, the FS was incompletely developed, and that other proteins usually found in the principal piece region of the flagellum were either absent or reduced in amount (Miki *et al*., 2002). Baccetti *et al*, detected deletions in *AKAP3/4* binding regions and a moderate diffuse signal after immunostaining for human AKAP4 protein in DFS patients (Baccetti *et al*., 2005a, 2005b). However, there is still controversy since strong evidences for the involvement of specific genes in the pathogenesis of DFS are not yet available (Turner *et al.*, 2001; Chemes and Rawe, 2010).

6. OBJECTIVES

A group of five asthenozoospermic patients from two Assisted Reproductive Medicine centres (Hospital Centre of Vila Nova de Gaia (CHVNG) (n = 4) and Centre of Reproductive Genetics Prof. Alberto Barros (CGR) (n=1)) were previously study by transmission electron microscopy and their sperm ultrastructural analysis was reported in detailed by Sousa and co-workers (Sousa *et al.,* 2014) and is also described in attachment 1.

This ultrastructural analysis revealed 4 cases with DFS associated with disrupted CPC, RS, DA and doublets; and one patient with *situs inversus* (compatible with KS), who presented absence of DA and nexin bridges. Therefore, the objective of the present work was to identify genetic alterations that could explain patients' phenotypes.

The genes *AKAP3, AKAP4, CCDC39, CCDC40, DNAI1, DNAH5* and *RSPH1* are genes involved in the assembly/composition of the FS (*AKAP3*, *4*) and Ax (*CCDC39, CCDC40, DNAI1, DNAH5* and *RSPH1*), namely DA, RS and CPC. Mutations in these genes were associated to defects in structures that are abnormal in the patients selected and also associated to sperm immotility.

To start this study, we proposed to test if mutations in these genes could be related with the observed defects in flagellar structures of infertile men with total absence of sperm motility. In that way, we will amplify by polymerase chain reaction (PCR) and sequence by the Sanger sequencing method the exonic regions of the selected genes and analyse the data with the help of bioinformatics resources.

With this work we expect to be able to offer a genetic diagnosis to the patients, find potential genetic markers for individuals with this medical condition and open the way for further investigations and the development of future genetic treatments.

II. METHODS

Methods

1. PATIENT, ETHICAL CONSIDERATIONS AND BIOLOGICAL MATERIAL

We performed a genetic diagnostic of five individuals with total absence of sperm motility from two Portuguese Assisted Reproductive Medicine centres: the Hospital Centre of Vila Nova de Gaia (CHVNG) and the Centre of Reproductive Genetics Prof. Alberto Barros (CGR). In compliance with the National Law on Medically Assisted Procreation (Law 32/2006) and the guidelines of the National Council on Medically Assisted Procreation (CNPMA, 2008), semen and blood samples were used with the written informed consent of the patients.

Patients' sperm ultrastructural analysis was previously reported in detailed (Sousa *et al.*, 2014). In Attachment 1 is escribed the methodology applied by Sousa *et al.* (2014) to perform the transmission electronic analyses of these patients and its main results. Briefly, ***Patient 1*** and ***4*** with DFS associated with absence of the annulus, CPC and RS (Patient 1: also with allergies and chronic sinusitis). ***Patient 3***, with chronic bronchitis, also presented DFS associated with absence of the annulus, CPC, RS, DA and nexin bridges. ***Patient 2*** had a different presentation of DFS, which was associated with an intact annulus and disorganization of doublets. Finally, ***patient 5*** presented *situs inversus*, severe respiratory symptoms (nasal polyps, chronic sinusitis, rhinitis and bronchitis) and the sperm flagellum lacked DA and nexin bridges. The clinical features of patient 5 are compatible with the KS, however the saccharin test and the measurement of nasal nitric oxide, which are the most popular screening tests for PCD, were not performed. Thus, despite the clinical features being quite compatible, we cannot affirm that this patient present a KS.

2. STUDIED GENES

We evaluated seven different genes that encode essential proteins of the sperm flagellum, *AKAP3* (NM_001278309.1), *AKAP4* (NM_003886.2) *CCDC39* (NM_181426.1), *CCDC40* (NM 017950.3), *DNAH5* (NM_001369.2), *DNAI1* (NM_001281428.1) and *RSPH1* (NM_080860.3) (Table II). Although the majority of the patients had a defined diagnosis, as all presented anomalies in common flagellar structures, we analysed all genes in all patients.

Table II. List of selected genes and a summary of the main phenotype changes.

Gene	Accession number	Gene Location	Molecular Function	Main ultrastructural defect[1]	Key reference
AKAP3	NM_001278309.1	12p13.3	Major structural proteins of FS	Marked hyperplasia and disorganization of the FS	Baccetti *et al.*, 2005a, 2005b; Chemes and Rawe, 2010
AKAP4	NM_003886.2	Xp11.22			
CCDC39	NM_181426.1	3q26.33	Assembly of DRC and IDA complexes	Incorrect assembly od IDA, DRC, CPC and RS	Merveille *et al.*, 2013; Antony *et al.*, 2013
CCDC40	NM 017950.3	17q25.3		Misplacement of the doublets and CPC; abnormal RS and defective assembly of IDA and DRC	
DNAI1	NM_001281428.1	9p13.3	Important for function of the ODA complex	Absence of ODA	Pennarun *et al.*, 1999
DNAH5	NM_001369.2	5p15.2			Olbrich *et al.*, 2002
RSPH1	NM_080860.3	21q22.3	A radial-spoke-head protein	CPC and RS defects	(Kott *et al.*, 2013)

[1] Associated to mutations in referred genes and detectable by TEM.
Abbreviations used: CPC-Central Pair Complex; DRC- Dynein regulatory complex; IDA: Inner Dynein Arms; ODA: Outer Dynein; FS: Fibrous Sheath; RS-Radial Spoke.

3. BLOOD COLLECTION

Blood samples were collected, using EDTA as anticoagulant, at the Centre of Medical Genetics Jacinto Magalhães, Hospital Centre of Porto (CGM-CHP).

4. EXTRACTION OF DNA

DNA from patients peripheral blood lymphocytes were extracted 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. Aliquots with 100ng/µl of DNA were made and stored, also at 4 ° C.

5. PRIMER DESIGN

The sequences of the interest genes were retrieved using the NCBI (http://www.ncbi.nlm.nih.gov/) and the UCSC (University of California, Santa Cruz) Genome Browsers (http://genome.ucsc.edu/index.html). These sequences were inserted in the GenMol program (CGM-CHP institutional use, unpublished), which gives a user friendly interface to visualize the sequences and allows an easier identification of the exonic regions. Then, the Primer Express software (LifeTechnologies) was used to design primers that covered all exonic regions previously identified (with exception of the gene *DNAH5* in which were only designed

primers for forty of a total seventy-nine, known to harbour the major number of mutations (Djakow *et al.*, 2012)).

Primer design parameters were: primer length 18-25 bp; percentage of GC 40%-60%; amplicon length (AL) 400-650 base pairs (bp); and temperature melting (Tm) 59Cº-62Cº. Some primers had punctual modifications in the parameters. The total list of primers used, with reference to the modifications in design parameters are listed in Attachment 2. Each designed primer pair was tested against: i) the presence of dimers using the FastPCR software (version 3.7.7; Institute of Biotechnology, University of Helsinki, Finland); and ii) their specificity towards the regions of interest identified using PrimerBlast tool (Ye *et al.*, 2012). Finally, primers were aligned with the total sequence of the gene to verify their annealing positions, using the software PrimerShow, at http://www.bioinformatics.org/sms/primer_show.html (Perry III, 2002).

6. POLYMERASE CHAIN REACTION

The gene regions of interest identified were amplified, using the primers previously designed, by polymerase chain reaction (PCR), in a thermocycler of models 9700, 9800 or VERITI (LifeTechnologies).

PCR conditions were optimized to each primer pair (see Attachment 2). For that purpose DNA from a healthy man was used as a control. Four PCR programs were applied: **(1) Stand-38c,** with an initial denaturation at 95°C for 2 min, followed by thirty-eight cycles of 95°C for 45 s, specific annealing temperature (Ta, listed in tables of attachment 2) for 30 s, and 72°C for 1 min, with a final extension at 72°C for 10 min; **(2) Stand-35c,** with an initial denaturation at 95°C for 10 min, followed by thirty-five cycles of 95°C for 1 min, specific Ta (listed in tables of attachment 2) for 30 s, and 72°C for 2 min, with a final extension at 72°C for 10 min; **(3) Multi-Dmsc**, a PCR with an initial denaturation at 95°C for 10 min, followed by thirty cycles of 95°C for 1 min, with a slower ramp reduction to 58/63°C for 30 s, and 72°C for 1 min, with a final extension at 72°C for 10 min; and **(4) High-GC**, a PCR with an initial denaturation at 95°C for 5 min, followed by forty cycles of 95°C for 45 s, 58°C for 30 s, and 72°C for 1 min, with a final extension at 72°C for 10 min.

The PCR reaction mixture for the three first conditions (30 µl) 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) and 1 µl of DNA at 100ng/µl. For the High-GC PCR condition, the reaction mixture (30 µl) contained: 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 (Thermo Fisher Scientific)

and 2,5 µl of DNA at 100ng/µl. To amplify more complex genomic regions, the PCR Master Mix was replace by either: i) EmeraldAmp Max HS PCR Master Mix (Takara Bio, Shiga, Japan), which includes a high yield, Hot Start (HS) PCR enzyme, gel loading dye (green) and a density reagent in a 2X *Taq* PCR Master Mix format; ii) ImmoMix™ Red (Bioline), which includes a heat-activated thermostable DNA polymerase and gel loading dye (red); and iii) DreamTaq Green PCR Master Mix 2X (Thermo Fisher Scientific).The list of all the PCR conditions and reaction mixture used for amplify each region are also detailed described in Attachment 2.

PCR products were analyzed by 2% agarose gel electrophoresis: mix of TAE 1X SeaKem LE Agarose (Lonza, Rockland, USA), and 5µL/100ml of GelRed Nucleic Acid Gel Stain, 10,000X in water (Biotium, California, USA) and were used phiX174 DNA BsuRI/HaeIII Marker, ready-to-use (New England BioLabs, County Road Ipswich, USA) as a marker for sizing and approximate quantification of small double-stranded DNA fragments. Finally, the PCR products were visualized and photographed by LAS-3000 (version 2.2, Fuji Film, Tokyo, Japan).

7. SANGER DNA SEQUENCING

Following successful PCR amplification, reactions were enzymatically purified using Illustra ExoStar kit (GE Healthcare- Buckinghamshire, UK). A new asymmetric PCR was prepared based on Sanger sequencing method, where random incorporation of chain-terminating dideoxynucleotides is promoted by a modified DNA polymerase, generating several DNA chains with different sizes. For this asymmetric PCR, BigDye Terminator v1.1 Cycle Sequencing Kit (LifeTechnologies) was used. Next these reactions were purified using a size selection column method Performa DTR (for Dye Terminator Removal) (EdgeBio, Maryland, USA). In the final step, the obtained products were resolved and analysed by high resolution electrophoresis in a 3130xl genetic analyzer (LifeTechnologies)

8. EXOME SEQUENCING BY NEXT-GENERATION SEQUENCING

The exome of the patient 5, compatible with KS, was sequenced using the AmpliSeq strategy in the Ion Proton NGS platform. A total of 75 ng of high quality DNA from the patient were amplified in twelve primer pools amplicons (200pb of average size) 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).

After the Ion Proton adapter sequences and low quality bases trimming using Torrent Suite software (Life Technologies), reads were mapped against the human reference genome hg19 using Torrent Mapping *Alignment* Program (version 4.0.6, Life Technologies). Variant calling was performed by running Torrent Variant Caller plugin (version 4.0), using the optimized parameters for exome sequencing recommended for Ampliseq sequencing (Life Technologies). The obtained Variant Call Format (VCF) file was analysed in GEMINI software (Paila *et al.*, 2013) for annotation and prioritization. Variant filtering was performed using a list of 67 candidate genes previously selected based on the literature (Pereira, in press; Attachment 3) and variants matching the autosomal recessive disease model (either homozygous or two heterozygous variants in the same gene). The Ion Reporter™ Software (Life technologies) was also used to filter rare variants by Gene Ontology (Ashburner *et al.,* 2000), using the key words: sperm, flagellar, motility, axoneme and dynein. Candidate variants were manually checked on the Binary Alignment Map (BAM) file through GenomeBrowse version 2.0.2 (Golden Helix, Bozeman, United States). Seven manually checked variants were selected and submitted to PCR amplification and Sanger sequencing for confirmation as previously described. The detailed list of primers and PCR conditions used to perform these confirmations is available at Attachment 4.

9. DATA ANALYSIS AND INTERPRETATION

Sequencing analyses were performed using SeqScape (V2.5 software; Life Technologies), which is a software designed for mutation detection and analysis and Single Nucleotide Polymorphisms (SNP) discovery and validation. Variants were validated and interpreted using in-house software: GenMol (unpublished) and Variobox (University of Aveiro, Aveiro, Portugal) (Gaspar *et al.*, 2013); and external on-line resources: Exome variant server (http://evs.gs.washington.edu/EVS/), which is a database that compiles information about the variants already described of the Human genome; NCBI dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/), that is a database of SNP and multiple small-scale variations that include insertions/deletions, microsatellites, and non-polymorphic variants. We also used the Mutalyzer 2.0.beta-

29 (https://mutalyzer.nl/) that allows checking if the variant description on sequence cause changes in translation.

In order to perform further bioinformatics analysis of the new and rare DNA variants found in the analysed genes, we used the free available online tools Polyphen-2 (Polymorphism Phenotyping v2; available in http://genetics.bwh.harvard.edu/pph2/) that predicts the possible impact of an amino acid substitution on the structure and function of a human protein (Adzhubei *et al.*, 2010); SIFT ('Sorting Tolerant From Intolerant' available in http://sift.jcvi.org) that predicts, based on the degree of conservation of amino acid residues, whether an amino acid substitution affects protein function (Ng and Henikoff, 2003); MutationTaster (available in http://doro.charite.de/MutationTaster/), that integrates information from different biomedical databases and established analysis tools to predict the disease potential of DNA sequence alterations (Schwarz *et al.*, 2010); Human Splice Finder (HSF) (available in http://www.umd.be/HSF/), which is a tool that predicts the effects of mutations on splicing signals or to identify splicing motifs in any human sequence (Desmet *et al.*, 2009). Some of these tools are integrated in Alamut Visual V2.4 software (Interactive Biosoftware, Rouen, France).

III. RESULTS

Results

1. GENETICS ANALYSIS BY SANGER SEQUENCING

Our main goal was to perform a genetic characterization five patients with total sperm immotility, in order to find a genetic alteration that may be causing the morphological phenotype diagnostic in our patients. Four patients had clinical feature compatible with DFS associated with disrupted CPC, RS, DA and doublets; and one patient has *situs inversus* and its axoneme show absence of DA and nexin bridges (compatible with KS). Consequently, we initiated our research by studying 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.*

Briefly, the genes *AKAP3* and *AKAP4* code a member of A-kinase anchoring proteins (AKAPs), which are involved in the organization of the basic structure and assembly of the FS. Mutations are associated with DFS (Baccetti *et al.*, 2005a, 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 disorganization of the Ax, including disorganization of the peripheral and outer microtubule doublets, absence or shifted CPC, absence of IDA, and anomalies in nexin links and RS (Becker-Heck *et al.*, 2011; Merveille *et al.*, 2011; 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 with the absence of ODA (Pennarun *et al.*, 1999; Olbrich *et al.*, 2002; Hornef *et al.*, 2006). Finally, *RSPH1* gene codes a RS-head protein that is important for the proper building of CPC and RS, since mutations in the *RSPH1* leads to an abnormal axonemal configuration with CPC and RS defects (Kott *et al.*, 2013; Onoufriadis *et al.*, 2014)

We designed specific primers and optimization was achieved by performing several PCR reactions under different conditions, using a DNA sample from a control (fertile male). Once the PCR conditions were optimized (see the PCR conditions for each exon at Attachment 2), we amplified the patients' genomic DNA by PCR (Fig. 8). Next, all the exonic regions and its intronic boundaries (about 100 bp upstream and downstream of each exon) were sequenced by Sanger sequencing method. Due to the elevated number of exons, the gene *DNAH5* was the only exception. In this gene we amplified and sequenced forty exons of the total seventy-nine, that are known to harbour a significant number of mutations (Djakow *et al.*, 2012).

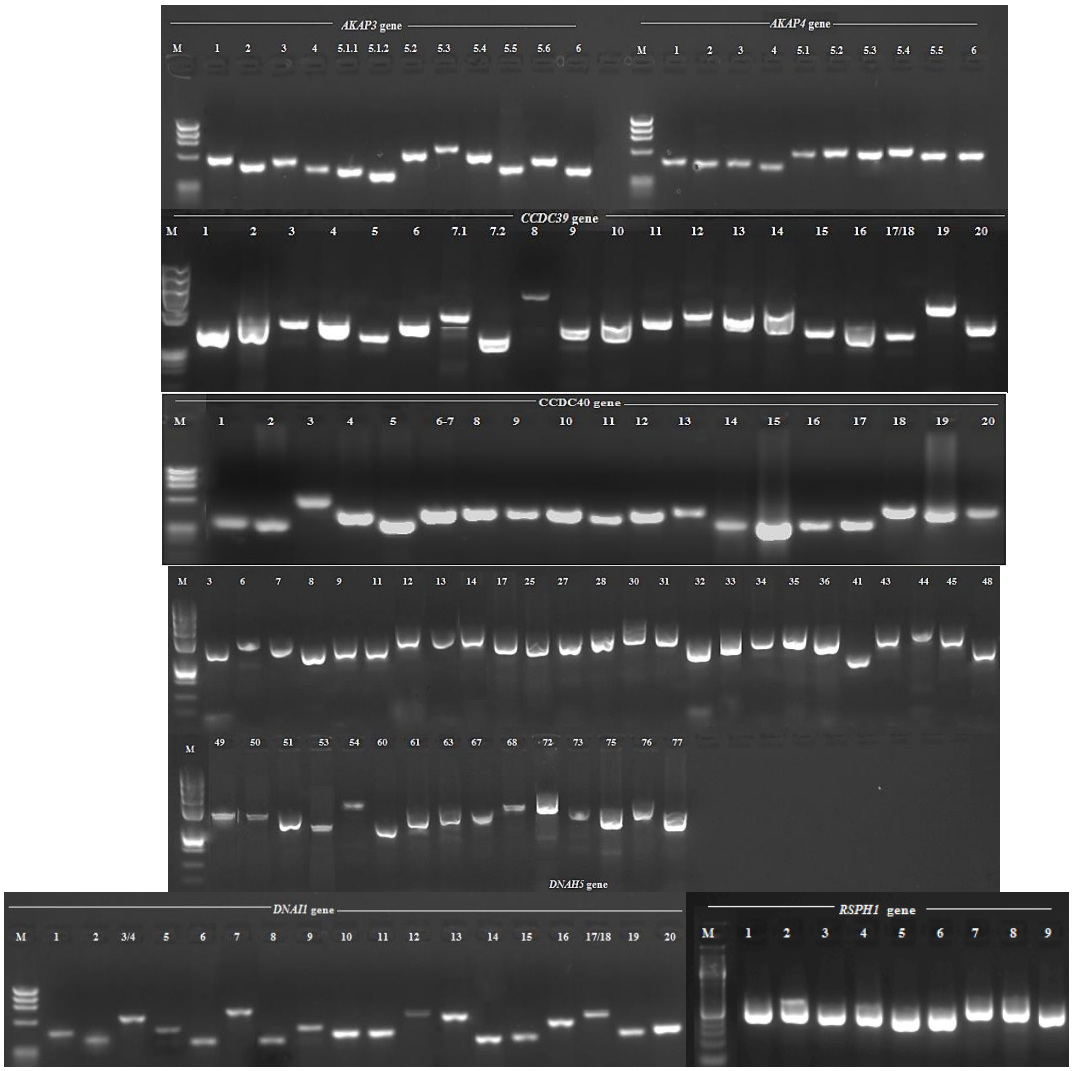


Figure 8. PCR products after analyses by 2% agarose gel electrophoresis: mix of TAE 1x SeaKem LE Agarose (Lonza), and 5µL/100ml of GelRed Nucleic Acid Gel Stain (Biotium). phiX174 DNA BsuRI/HaeIII Marker, ready-to-use (New England BioLabs) were used as a marker (M) for sizing. In each picture the gene's name is listed and in each land are the number of the respective exon (For more information about PCR conditions see Attachment 2).

We used the software SeqScape (Life Technologies) for sequencing analysis, and we found eight DNA sequence variants in *AKAP3* gene, one in *AKAP4* gene, ten in *CCDC39* gene, twenty-six in *CCDC40* gene, four in *DHAI1* gene, forty-five in *DNAH5* and eight in *RSPH1* gene (see Table III).

Variants were validated and interpreted using in-house software: GenMol (unpublished) and Variobox (Gaspar *et al.*, 2013); as well as the external on-line resources: (1) Exome variant server (available in http://evs.gs.washington.edu/EVS/), (2) NCBI dbSNP (available in http://www.ncbi.nlm.nih.gov/projects/SNP), and (3) Mutalyzer 2.0.beta-29 (available in https://mutalyzer.nl/). Variants whose frequency within European population (based on Exome variant server and SNP database from NCBI) was above 1% were considered as polymorphisms, which is the arbitrary cut-off value between a rare variant and polymorphism.

Table III. **DNA sequence variants identified by Sanger sequencing analysis.**

Gene	Patient	DNA change	Protein change [a]	Previously described [b]	Frequency*	Impact
AKAP3	1,2,4 (He) 3,5 (Ho)	c.-522C>G	-	Yes	43.5%	Polymorphism
	1,2,4 (He) 3,5 (Ho)	c.-245+15A>G	-	Yes	53.3%	Polymorphism
	2, 4 (He)	c.-106-96T>G	-	Yes	19.4%	Polymorphism
	1,3,5 (Ho) 2,4 (He)	c.353G>A	p. Gly118Glu	Yes	80.5%	Polymorphism
	1,3,5 (Ho) 2,4 (He)	c.609G>A	p.(=)	Yes	77.0%	Polymorphism
	1 (Ho) 2,4 (He)	c.1026C>T	p.(=)	Yes	51.2%	Polymorphism
	2,4 (He)	c.1499T>C	p. Ile500Thr	Yes	14.6%	Polymorphism
	3,5 (Ho)	c.2331T>C	p.(=)	Yes	29.2%	Polymorphism
AKAP4	5 (Hem)	c.174+71A>G	-	Yes	7.8%	Polymorphism
CCDC39	**3 (He)**	**c.233G>A**	**p.Arg78His**	**Yes**	**0.6%**	**U (Table IV)**
	2 (He)	c.545A>G	p. Thr182Ser	Yes	2.6%	Polymorphism
	2,3,5 (He)	c.738+169G>A	-	Yes	20.8%	Polymorphism
	2,3,5 (He)	c.930+95A>G	-	Yes	50.0%	Polymorphism
	5 (He)	c.1248A>G	p.(=)	Yes	17.1%	Polymorphism
	3,5 (Ho), 4 (He)	c.1359C>T	p.(=)	Yes	53.1%	Polymorphism
	5 (He)	c.1528-43A>G	-	Yes	47.6%	Polymorphism
	5 (He)	c.2301G>A	p.(=)	Yes	10.4%	Polymorphism
	3 (Ho) 4 (He)	c.2397G>A	p.(=)	Yes	22.2%	Polymorphism
	1 (He)	**c.2540A>G**	**p.Glu847Gly**	**No**	**-**	**U (Table IV)**
CCDC40	3 (He)	c.-49C>A	-	Yes	8.4%	Polymorphism
	1-3, 5 (He)	c.29+58C>A	-	Yes	25.0%	Polymorphism
	3 (He)	c.30-191G>T	-	Yes	10.8%	Polymorphism
	3 (He)	c.30-80G>A	-	Yes	10.5%	Polymorphism
	1-3,5 (Ho) 4 (He)	c.207G>C	p.(=)	Yes	99.0%	Polymorphism
	3 (He)	c.676+13T>C	-	Yes	16.0%	Polymorphism
	3 (He)	c.676+59T>C	-	Yes	34.3%	Polymorphism
	1, 3,5 (He)	c.677-4C>G	-	Yes	21.2%	Polymorphism
	3 (He)	c.873C>T	p.(=)	Yes	8.3%	Polymorphism
	1, 3,5 (He)	c.1159+12C>T	p.(=)	Yes	18.9%	Polymorphism
	1,5 (He)	c.1440+136C>T	-	Yes	26.1%	Polymorphism
	5 (He)	c.1441-106A>G		Yes	12.3%	Polymorphism
	3 (Ho)	c.1562+64C>T	-	Yes	7.7%	Polymorphism
	1,3,4 (He) 2 (Ho)	c.1890T>G	p.(=)	Yes	91.9%	Polymorphism

Zygosity: Ho- homozygous; He- heterozygous; Hem- hemizygous (gene locus in cromosome X).

a-*In silico* predition; b-Variants previously described in databases (Exome variant server and/or dbSNP from NCBI). The variants in blue **Bold** are the novel variants (variants not described in databases/literature) and the variants listed on databases but with a non determined frequency or lower than 1% within European population ((*) based on Exome variant server and SNP database from NCBI). These variants were further analyzed and the results are at table IV.

Abbreviations: c, coding sequence; A, Adenine; G, Guanine; T, Thymine; C, Cytosine; p, protein sequence; Gly, Glycine; Glu, Glutamic acid; Ile, Isoleucine; Thr, Threonine; Arg, Arginine; His, Histidine; Ser, Serine; U, variant with unknown effect.

Table III (continuation) DNA sequence variants identified by Sanger sequencing analysis.

Gene	Patient	DNA change	Protein change [a]	Previously described [b]	Frequency*	Impact
CCDC40	5 (He)	c.2255T>C	p.Leu752Pro	Yes	1.6%	Polymorphism
	3 (He)	c.2323G>A	p.Val775Met	Yes	9.9%	Polymorphism
	4 (Ho) 1-3,5 (He)	c.2450-89A>C	-	Yes	70.0%	Polymorphism
	4 (Ho) 1-3,5 (He)	c.2450-53G>A	-	Yes	65.0%	Polymorphism
	3 (He)	**c.2620-92C>T**	**-**	**No**	**-**	**U (Table IV)**
	4 (He)	**c.2682G>A**	**p.(=)**	**Yes**	**0.6%**	**U (Table IV)**
	3 (He)	c.3021+74C>A	-	Yes	17.8%	Polymorphism
	1,2,4,5 (He)	c.3030T>C	p.(=)	Yes	74.8%	Polymorphism
	3 (He)	c.3210A>G	p.(=)	Yes	28.0%	Polymorphism
	3,5 (He)	c.3417A>G	p.(=)	Yes	37.6%	Polymorphism
	3 (He)	c.*15T>C	-	Yes	29.0%	Polymorphism
	3 (He)	c.*149T>C	-	Yes	47.0%	Polymorphism
DNAI1	**1(He)**	**c.81+61A>G**	**-**	**Yes**	**0.3%**	**U (Table IV)**
	2,4 (He)	c.1003G>A	p.Val335Ile	Yes	17.7%	Polymorphism
	1,3, 5 (Ho)	c.1019+42C>T	-	Yes	24.5%	Polymorphism
	2(Ho) 3,4 (He)	c.1489+130G>A	-	Yes	37.2%	Polymorphism
DNAH5	1,2,5 (He) 4 (Ho)	c.661-172A>G	-	Yes	39.7%	Polymorphism
	3 (Ho)	c.975+30A>C	-	Yes	28.2%	Polymorphism
	3 (Ho)	c.1089+141T>C	-	Yes	27.0%	Polymorphism
	4 (Ho) 1,2,5 (He)	c.1197+155G>T	-	Yes	40.0%	Polymorphism
	4 (Ho) 1,2,5 (He)	c.1321-53A>G		Yes	43.0%	Polymorphism
	4 (Ho) 1,2,5 (He)	c.1503T>C	p.(=)	Yes	89.1%	Polymorphism
	3 (Ho)	**c.1537-102T>A**	**-**	**Yes**	**ND**	**U (Table IV)**
	4 (Ho) 1 (He)	**c.1537-100_ 1537-99delTT**	**-**	**Yes**	**ND**	**U (Table IV)**
	3 (Ho)	c.1644+90C>A	-	Yes	30.0%	Polymorphism
	3 (Ho) 4 (He)	c.1645-196A>G	-	Yes	31.4%	Polymorphism
	4(Ho) 1,2,5 (He)	c.1672A>G	p.Thr558Ala	Yes	42.0%	Polymorphism
	1-4(He)	**c.3835-3delT**	**-**	**Yes**	**0.1%**	**U (Table IV)**
	1-4(He)	c.4152A>G	p.(=)	Yes	45.0%	Polymorphism
	1-4 (He)	c.4355+60T>C	-	Yes	50.0%	Polymorphism
	1-4 (He)	c.4374G>T	p.(=)	Yes	44.7%	Polymorphism
	1-5 (Ho)	c.4797-93C>A	-	Yes	97.0%	Polymorphism
	1-4 (He)	c.4797-29C>A	-	Yes	40.3%	Polymorphism
	1-4(He)	c.4951-92G>A	-	Yes	50.0%	Polymorphism
	1-4(He)	c.5114+11T>C	-	Yes	60.8%	Polymorphism

Zygosity : Ho- homozygous; He- heterozygous; Hem- hemizygous (gene locus in cromosome X).
a-In silico predition; b-Variants previously described in databases (Exome variant server and/or dbSNP from NCBI). The variants in blue **Bold** the novel variants (variants not described in databases/literature) and the variants listed on databases but with a non determined frequency or lower than 1% within European population ((*) based on Exome variant server and SNP database from NCBI). These variants were further analyzed and the results are at table IV.

Abbreviations: c, coding sequence; A, Adenine; G, Guanine; T, Thymine; C, Cytosine; p, protein sequence; Leu, Leucine; Pro, Proline; Val, Valine; Met, Methionine; Ile, Isoleucine; Thr, Threonine; Ala, Alanine; U, variant with unknown effect; ND, non-defined.

Table III (continuation) DNA sequence variants identified by Sanger sequencing analysis.

Gene	Patient	DNA change	Protein change [a]	Previously described [b]	F*	Impact
DNAH5	1,2,4,5(He)	c.5114+22A>T	-	Yes	50.3%	Polymorphism
	1-4(He)	c.5114+110A>G	-	Yes	50.0%	Polymorphism
	2 (Ho) 3,4 (He)	c.5115-49G>C	-	Yes	64.6%	Polymorphism
	2(Ho) 3,4(He)	c.5115-5T>C		Yes	36.5%	Polymorphism
	2(Ho) 3,4(He)	c.5172C>T	p.(=)	Yes	35.4%	Polymorphism
	2(Ho) 3,4(He)	c.5710-58G>A	-	Yes	50.0%	Polymorphism
	3(He)	**c.5882+133A>G**	**-**	**No**	**-**	**U (Table IV)**
	1-3,5 (Ho)	c.7408-160G>C	-	Yes	80.9%	Polymorphism
	2 (Ho), 4 (He)	**c.7408-84_7408-83delAT**	**-**	**Yes**	**ND**	**U (Table IV)**
	2 (Ho), 4 (He)	c.7408-72C>T	-	Yes	50.0%	Polymorphism
	2 (Ho), 4(He)	c.7609+19C>T	-	Yes	53.8 %	Polymorphism
	4,2 (Ho), 5(He)	c.8586G>T	p.Leu2862Phe	Yes	32.6%	Polymorphism
	1(Ho)	c.10102-29C>T	-	Yes	26.2%	Polymorphism
	1 (Ho)	c.10140A>G	p.(=)	Yes	29.8%	Polymorphism
	1 (Ho)	**c.10282-81delT**	**-**	**Yes**	**ND**	**U (Table IV)**
	3 (He)	**c.10872+84T>C**	**-**	**Yes**	**ND**	**U (Table IV)**
	1-3 (Ho), 4 (He)	c.11570+35A>G	-	Yes	57.5%	Polymorphism
	3 (He)	**c.11570+124G>C**	**-**	**No**	**-**	**U (Table IV)**
	1-3 (Ho), 4 (He)	c.11761+93G>T	-	Yes	56.6%	Polymorphism
	4 (He), 5 (Ho)	c.12401C>T	p.Ala4134Val	Yes	46.2%	Polymorphism
	4 (He), 5 (Ho)	c.12468A>C	p.(=)	Yes	42.0%	Polymorphism
	5 (Ho), 4 (He)	c.12500-194T>G	-	Yes	40.7%	Polymorphism
	5 (Ho), 1,3,4(He)	c.12910-11C>T	-	Yes	40.1%	Polymorphism
	2,5 (Ho)	c.13348A>G	p.Ile4450Val	Yes	46.6%	Polymorphism
	2,5 (Ho)	c.13359A>G	p.(=)	Yes	48.6%	Polymorphism
	2,5 (Ho)	c.13491+66T>C	-	Yes	57.8%	Polymorphism
RSPH1	1,5 (Ho) 2,4(He)	c.-43C>T	-	Yes	63,97%	Polymorphism
	1,2,5 (Ho) 3,4 He	c.55-232C>T	-	Yes	85,00%	Polymorphism
	1-5 Ho	c.55-189T>C	-	Yes	90,00%	Polymorphism
	1,5 (Ho) 2,4(He)	c.55-181C>T	-	Yes	36,00%	Polymorphism
	1,5 (Ho) 2,4(He)	c.365+130C>G	-	Yes	37,00%	Polymorphism
	1,5 (Ho) 2,4(He)	c.393G>A	-	Yes	64,72%	Polymorphism
	2,3 (He)	c.742G>A	p.Gly248Arg	Yes	23,69%	Polymorphism
	1,4 (He)	c.*53C>T	-	Yes	11,69%	Polymorphism

Zygosity : Ho- homozygous; He- heterozygous; Hem- hemizygous (gene locus in cromosome X).
a-*In silico* predition; b-Variants previously described in databases (Exome variant server and/or dbSNP from NCBI). The variants in blue **Bold** the novel variants (variants not described in databases/literature) and the variants listed on databases but with a non determined frequency or lower than 1% within European population ((*) based on Exome variant server and SNP database from NCBI). These variants were further analyzed and the results are at table IV.

Abbreviations: c, coding sequence; A, Adenine; G, Guanine; T, Thymine; C, Cytosine; p, protein sequence; Leu, Leucine; Phe, Phenylalanine; Ala, Alanine; Val, Valine; Ile, Isoleucine; Gly, Glycine; Arg, Arginine; U, variant with unknown effect; ND, non-defined.

The majority of the variants found in this initial genetic screening using the Sanger sequencing method, were already described in online databases as polymorphisms (frequency >1%). Nevertheless, this screening allowed the identification of **four new DNA variants**: one new variant in *CCDC39* gene (c.2540A>G), one in *CCDC40* gene (c.2620-92C>T) and two in *DNAH5* gene (c.5882+133A>G and c.11570+124G>C) (see Table III); and **nine rare variants** (that is, variants listed on databases but with a non-determined frequency or lower than 1% within European population based on Exome variant server and dbSNP from NCBI): one in *CCDC39* (c.233G>A), one in *CCDC40* (c.2682G>A), six in *DNAH5* (c.1537-102T>A, c.1537-100_1537-99delTT, c.3835-3delT, c.7408-84_7408-83delAT, c.10282-81delT, c.10872+84T>C), and one in *DNAI1* (c.81+61A>G) (see Table III). To predict the impact of the new and rare variants on patients' phenotype, further bioinformatics analysis were performed, using the freely accessible online tools: Polyphen-2, SIFT, HSF and the MutationTaster (see Table IV).

In **patient 1,** we observed forty-three polymorphisms, **four rare variants** (*DNAH5* gene: c.1537-100_1537-99delTT, c.3835-3delT and c.10282-81delT; and *DNAI1* gene: c.81+61A>G) and **one novel variant** in *CCDC39* gene.

The new heterozygous variant **c.2540A>G**, located in the exon 18 of the *CCDC39* gene (Fig. 9A), leads to a change on the protein position 847 of glutamic acid by glycine. Both bioinformatic tools Polyphen-2 and SIFT[1] predict this amino acid change as non-pathogenic with a score of 0.17 and 0.066, respectively. By its turn, the MutationTaster tool predicts that this variant may be disease causing, with a probability of 0.69. According to HSF analysis, there is a marginal increase of the score for a exonic splicing enhancer (ESE) motif, more specifically, the alternative splicing factor ASF/SF2 (wild type (wt) score: 72.9 and mutant (mt, i.e. the score of the altered sequence) score: 82.9). The rare heterozygous variant **c.81+61A>G** located in intron 2 of *DNAI1* gene (Fig. 9B), increases by 53.56% the score of splice site acceptor (SSA). This increase makes the mt score almost equivalent of native score (nt, i.e. normally occurring splice site), being the mt score: 82.97 and the nt score: 83.19. This variant is also predicted to be pathogenic by MutationTaster bioinformatic tool. Both variants are only present in patient 1. The rare variant **c.1537-100_1537-99delTT** is heterozygous in patient one and homozygous in patient 4. The heterozygous variant **c.10282-81delT**, was only detected in patient 1. Both are located at *DNAH5* gene and predicted as being non-pathogenic, since the bioinformatic tools used did not detected any impact (Fig. 9C, D).

[1] In Polyphen-2 scores higher than 0.5 are considered as damaging and in SIFT scores less than 0.05 are considerate deleterious.

The last rare variant found in this patient (1), is a heterozygous deletion of a T nucleotide three bases before the exon 25 (**c.3835-3delT**) in *DNAH5* gene (Fig. 9E). We analyzed this variant by five splice algorithms incorporated in Alamut Visual V2.4 software: SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer and HSF. With exception to NNSPLICE algorithm, which did not detect effect on splice, in the others algorithms this deletion is expected to lead to a slight reduction of the score of nt SSA, from 92.6 to 89.1 in SpliceSiteFinder-like; 12.8 to 10.4 in MaxEntScan; 11.9 to 9.8 in GeneSplicer and 85.85 to 85.29 in HSF. According to MutationTaster, this deletion has a probability to be pathogenic (p=1). This variant was also identified in patients 2, 3 and 4.

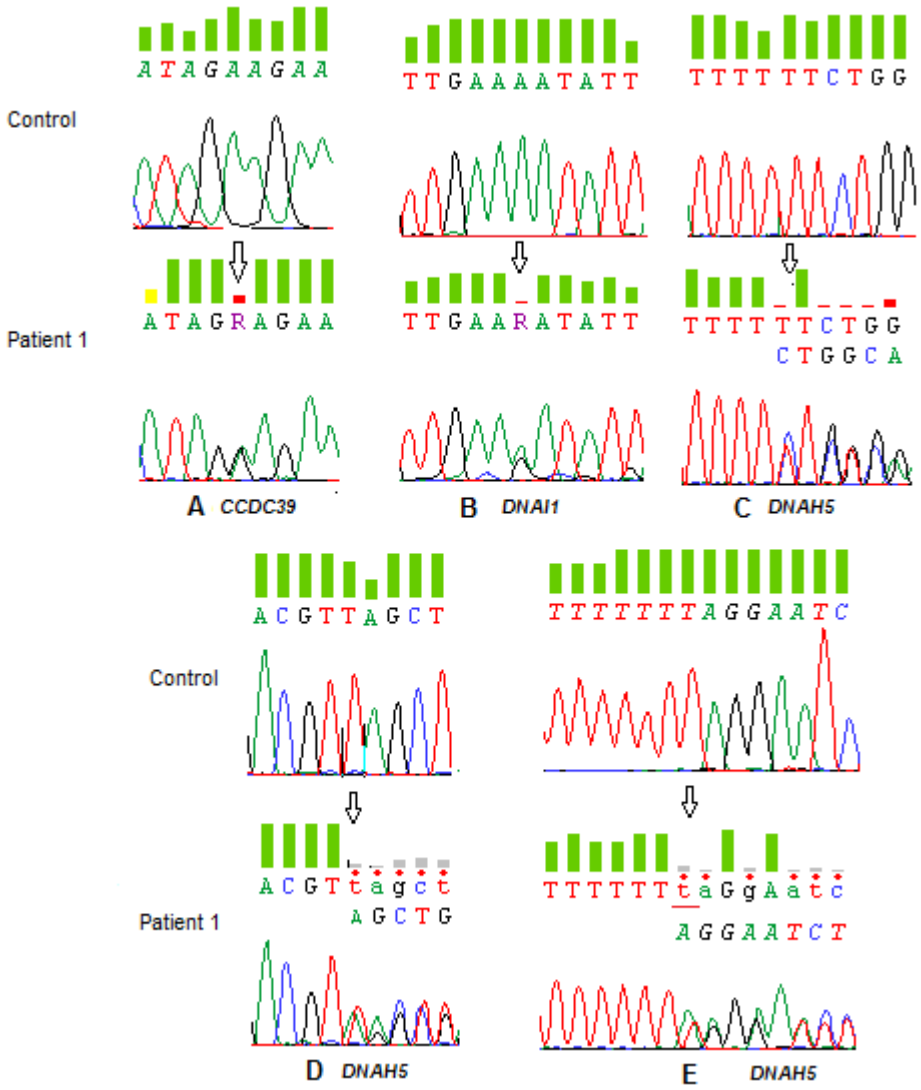


Figure 9. Electropherogram of variants detected in patient 1. A. The new heterozygous variant c.2540A>G located on *CCDC39* gene; **B**. The variant c.81+61A>G in *DNAI1* gene; **C.** The variant c.1537-100_1537-99delTT detected in patient 1 as heterozygous and in patient 4 as homozygous; **D.** The heterozygous variant c.10282-81delT also found in patient 1; E. The heterozygous deletion c.3835-3delT in *DNAH5* gene detected in patient 1 to 4.

The **patient 2** has fifty DNA variants: forty-eight are polymorphisms and two rare variants in *DNAH5* gene: c.3835-3delT and c.7408-84_7408-83delAT. The rare variant **c.7408-84_7408-83delAT** is homozygous in patient 2 and heterozygous in patient 4 was classified as polymorphism after the bioinformatic analysis.

The **patient 3** presents fifty-nine DNA variants, with fifty-two polymorphisms, **four rare variants** (c.233G>A in *CCDC39* gene (Fig. 10A); c.1537-102T>A (Fig. 10B), c.3835-3delT (Fig. 9E), and c.10872+84T>C (Fig. 10C) in *DNAH5)* and **three new variants** (c.2620-92C>T (Fig. 10D) in *CCDC40* gene, c.5882+133A>G (Fig. 10E) and c.11570+124G>C (Fig. 10F) in *DNAH5* gene). The heterozygous variant **c.233G>A** in *CCDC39* gene has lower frequency (0.6%) on control population and leads to a change of the amino acid arginine by a histidine at position 78, but no pathogenic impact is associated to this variant. Likewise, with exception of the variant c.3835-3delT in *DNAH5* gene, already described, the three new variants and the two rare variants are foreseen by bioinformatic tools as polymorphisms.

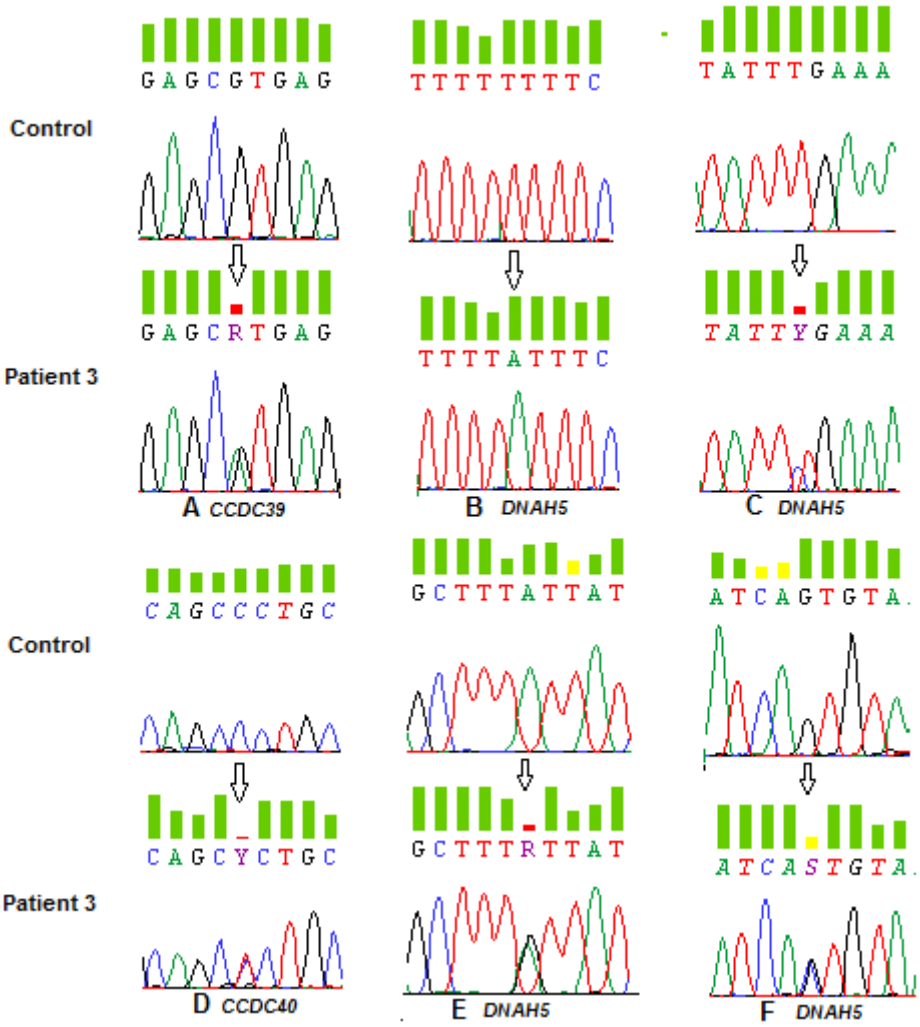
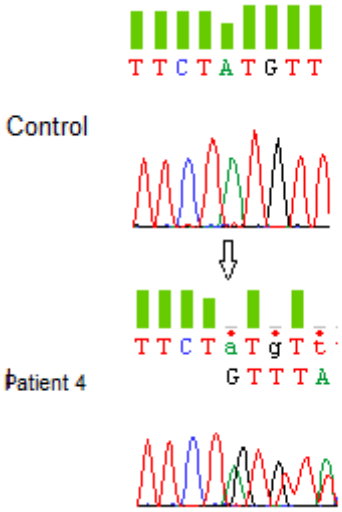


Figure 10. Electropherogram of the rare variants (A-C) and the novel variants (D-F) detected in patient 3. A. The heterozygous variant c.233G>A in *CCDC39* gene; **B**. The homozygous c.1537-102T>A in *DNAH5* gene; **C.** The heterozygous c.10872+84T>C in *DNAH5* gene; **D.** The novel heterozygous variant c.2620-92C>T in *CCDC40* gene; and the two novel heterozygous variants in *DNAH5* gene c.5882+133A>G **(E)** and c.11570+124G>C **(F).**



In **patient 4** were identified fifty-one variants: forty-seven common variants with higher frequency on populations and **four rare variants**. Three of these four rare variants are in *DNAH5* gene: c.1537-100_1537-99delTT (Fig. 9C), c.3835-3delT (Fig. 9E), c.7408-84_7408-83delAT (Fig. 11), and are also present in other patients of this study; the remaining was found in *CCDC40* gene c.2682G>A, and was detected only in this patient. Excluding the variant c.3835-3delT, all variants in this patient, are polymorphisms without predicted pathogenic impact.

Figure 11. Electropherogram of the heterozygous variant c.7408-84_7408-83delAT detected in patient 2 as Homozygous and in patient 4 as heterozygous

Patient 5, with *situs inversus* (compatible with KS), has forty-three DNA variants with any variant with a predicted pathogenic impact.

Table IV. 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	3 (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)
	1 (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	4 (He)	c.2682G>A (exon 16)	0.61%	p. (=)	-	-	No effect on Splice	P (p=0.90)
	3 (He)	c.2620-92C>T (Intron 15)	New variant	-	-	-	No effect on Splice	P (p=0.9)
DNAH5	3 (Ho)	c.1537-102T>A (Intron11)	ND	-	-	-	No effect on Splice	P (p=0.9)
	4 (Ho) 1 (He)	c.1537-100_1537-99delTT (Intron 11)	ND	-	-	-	No effect on Splice	P (p=0.9)
	1-4 (He)	**c.3835-3delT** (intron24)	**0.1%**	**-**	**-**	**-**	**New SSA mt:85.29 nt:85.85**	**Dc (p=1)**
	3 (He)	c.5882+133A>G (Intron 35)	New variant	-	-	-	No effect on Splice	P (p=0.9)
	2 (Ho) 4 (He)	c.7408-84_7408-83delAT (intron 44)	ND				No effect on Splice	P (p=0.9)
	1 (He)	c.10282-81delT (Intron60)	ND	-	-	-	No effect on Splice	P (p=0.9)
	3 (He)	c.10872+84T>C (Intron63)	ND	-	-	-	No effect on Splice	P (p=0.9)
	3 (He)	c.11570+124G>C (Intron67)	New variant	-	-	-	No effect on Splice	P (p=0.9)
DNAI1	**1 (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 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 and SNP database from NCBI); In blue bold are the variants most promising in terms of predicted pathogenic impact.

Abbreviations: 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, wild type (reference non-mutated score); mt, mutant score (i.e. the score of mutated sequence); nt, native splice site (i.e. normally occurring splice site); ESE, Exonic Splice Enhancer.

2. GENETICS ANALYSIS BY EXOME SEQUENCING

2.1. Exome Sequencing metrics

The results presented in the previous section did not supplied sturdy evidences that could explain the phenotype described in these 5 patients, namely in patient 5. Considering the complexity of sperm motility and that those patients present morphological defects in highly complex structures, with an important role in sperm motility, it is conceivable that several genes coding for proteins linked to its assembly, composition, function and regulation might be involved. Therefore, we decided to analyse all exonic regions (exome) by next-generation exome sequencing (ES) of one patient, the patient 5, in order to identify genetic variants that could be underlying this phenotype.

ES of patient 5 (lacking the DA and nexin bridges of its sperm flagellum) was performed using Ion AmpliSeq Workflow and analyzed on an Ion Proton sequencer. The ES run generated 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 addition, 84% of all sequenced bases had a Phred quality scores above Q20, which means that the probability of an incorrect base call is 1 in 100 (Ewing and Green, 1998; Ewing *et al.*, 1998). In this patient, this analysis identified 49,230 single nucleotide variants (SNV) and 3,479 insertions/deletions, which were listed in the Variant Call Format (VCF) file of this patient.

2.1.1. Exome filtering

The number of protein-coding genes within the human genome is estimated to be around 20,000 (Pennisi, 2012), so, as expected, ES generated a huge amount of data. After reads alignment and variant calling more than 50,000 DNA sequence variants were obtained. Consequently, in order to find a DNA variant that may be involved in patient phenotype, we had to filter these variants. As an initial approach we filtered the variants matching the autosomal recessive disease model: either homozygous or two heterozygous candidate variants for the same gene, using the list of 67 candidate genes previously reported in literature as being associated to sperm immotility due to flagellar abnormalities. The recessive disease model was chosen because patient 5 has no family history of infertility and the clinical features found were compatible with KS (PCD), which was consistently described in the literature as having an autosomal recessive inheritance pattern (Boon *et al*., 2013; Lie and Ferkol, 2007).

The GEMINI database framework (Paila *et al.*, 2013) was used for annotation and prioritization of variants. From this filters we narrowed down our analysis to 258 sequence variants in ten different genes. In addition, we also verified if, besides the genes previously selected, other loci could be also involved in the patient's phenotype. For that, we used the Ion Reporter™ Software (Life technologies) to filter the rare variants by Gene Ontology (that covers three domains: cellular component, molecular function and biological process), using the key words: flagellar motility, sperm motility, axoneme and dynein. Using this new filter we selected more 46 DNA variants. In total, with the application of both approaches we reduced from 52709 variants to 304 variants. Our initial filtering steps gave priority to rare variants located in exonic regions and those placed within intronic regions near exons (8bp from exon). For that reason, 167 intronic variants (more than 8bp away from the exon) and 53 variants with a frequency above 1% were excluded. Of the remaining 84 variants, 51 are synonymous, 30 are missense and 3 are frame-shift. The 51 synonymous variants were evaluated using bioinformatic tools and excluded for further analysis since no effects on splicing were predicted. Seventeen of the 30 missense variants belong to the gene *HYDIN* located at chromosome 16q22.2, which has an identical paralogous segment inserted on chromosome 1q21.1 (Doggett *et al.,* 2006). Consequently, we suspected that the high number of novel variants found in this gene may be attributed to this duplication. Our suspicion was also shared by Berg *et al*. (Berg *et al.,* 2011), which add further support to our supposition. As result, these changes were excluded from further analysis. The remaining 16 variants were visually inspected on the BAM file GenomeBrowse version 2.0.2. (Golden Helix). Nine variants were considered sequencing artifacts and seven were selected (Fig. 12) for Sanger sequencing confirmation.

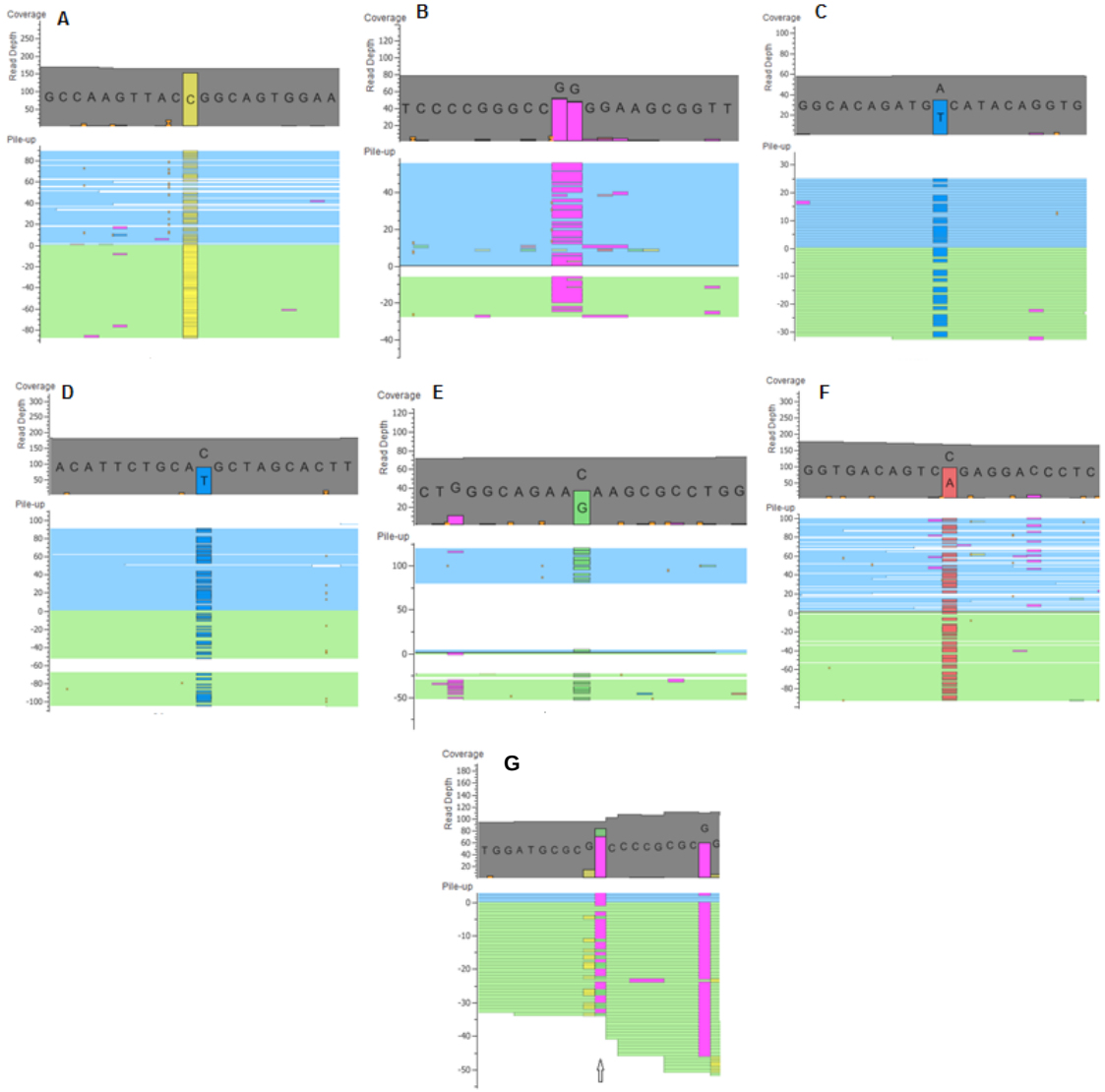


Figure 12. **Alignments of the selected variants after the visual inspection on the BAM file through GenomeBrowse version 2.0.2**. **A.** Variant c.104G>C in *CCDC103*; **B**. c.262_263delCC in *INSL6*; **C.** c.3167A>T in *DNAH6*; **D.** c.7895C>T in *DNAH10*; **E.** c.828C>G in *GAS8*; **F.** c.4445G>T in *SPAG17* ; **G**. c.8delC in *MYCBPAP* gene. The light blue and green lines represent the forward and revers strain, respectively. Citosine (C), Guanine (G), Timine (T) and Adenine nucleotides are the bars yellow, green, blue and red respectively. Deletions are presented as pink bars.

2.2. Exome results

Of the seven variants selected from ES to confirm by Sanger sequencing, six were confirmed as true variants and are present only in patient 5. The variant in *MYCBPAP* gene, c.8delC, proved to be a sequence artifact. These six variants include five missense changes (*CCDC103*:c.104G>C, p.Arg35Pro; *DNAH10*:c.7895C>T, p.Thr2632Met; *DNAH6*:c.3167A>T, p.Asp1056Val; *GAS8*:c.828C>G, p.Asn276Lys; and *SPAG17*:c.4445G>T, p.Arg1482Leu) and one frame-shift deletion (*INSL6*: c.262_263delCC). The variant in *DNAH10* gene is the only variant of these six that were already listed in genetic variant databases available online, but with a reduce frequency of 0.0013. All the others are reported for the first time. Table V resumes the bioinformatic analysis of the six variants selected and figure 13 shows the electropherograms after the Sanguer sequencing.

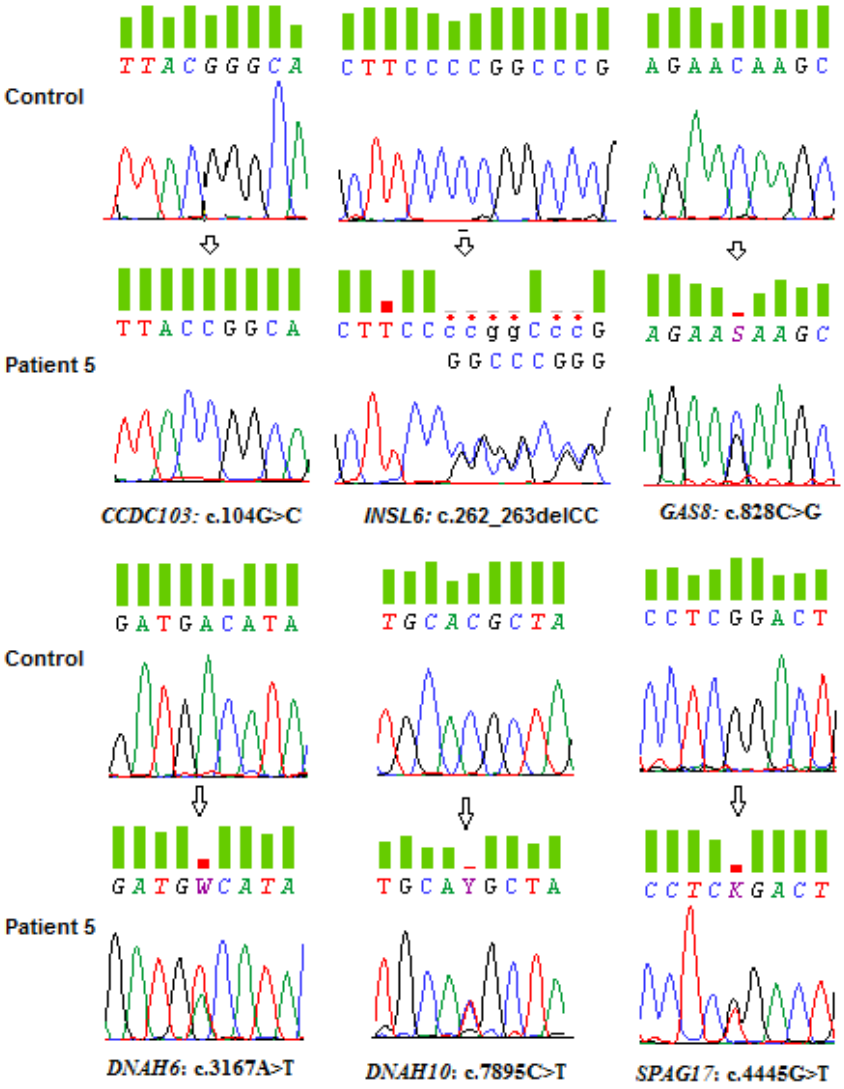


Figure 13. Electropherograms of the variants selected from exome sequencing and after Sanger sequencing confirmation. With the exception of c. 7895C>T variant located in DNAH10, whose frequency on control population is 0.0013, all are reported for the first time.

The novel homozygous variant **c.104G>C** located in exon 2 of *CCDC103* gene (NM_001258398.1; 17q21.31) lead to a change of a conserved amino-acid arginine by a proline at position 35 (Fig. 14). In three different bioinformatic software's this variant is predicted as potentially pathogenic, with the scores of 0 in SIFT, and 0.99 in Polyphen-2 and MutationTaster.

The heterozygous deletion of two cytosine (C) nucleotides at position 262 of exon 1 of the *INSL6* gene (NM_007179.2; 9p24), change the open read frame and may result in a completely different translation and originate a shorter polypeptide (**p.Pro88Glyfs*27**). MutationTaster evaluates this variant as disease causing (p=1). Consequently, is assumed to have a pathogenic impact.

We identified variants in two different genes that code for proteins of dynein HC: *DNAH10* gene (NM_207437.3; 12q24.31) and *DNAH6* gene (NM_001370.1; 2p11.2). According to Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (Franceschini *et al.,* 2013), these two genes may interact with a score of 0.540 (scores closer to 1 mean a highest confidence on result).

The heterozygous missense variant **c.7895C>T** in exon 47 of *DNAH10* gene, is the only of this six variants that is not reported here by the first time, but its frequency is very low (0.13%). This together with the fact that this gene is part of DA HC, led us to consider this variant for further analysis. This variant lead to a change of amino acid threonine by methionine at protein position 2632. This amino acid is much conserved and this alteration is predicted as being damaging by SIFT (s=0.02) and Polyphen-2 (s= 0.98) and as disease causing (p=0.99) by MutationTaster. The other missense variant **c.3167A>T** in exon 20 of *DNAH6* gene is also heterozygous and cause a change in another conserved amino acid at protein position 1056, an acidic polar with a negative charge, the aspartic acid, by a nonpolar and neutral amino acid, the valine. The three bioinformatic tools predicted that may have a pathogenic impact, with scores of 0.00, 0.98 and 0.99 respectively by SIFT, Polyphen-2 and MutationTaster.

The two last missense heterozygous variants **c.828C>G** and **c.4445G>T,** are in exon 7 of *GAS8* (NM_001481.2; 16q24.3) and exon 31 of *SPAG17* (NM_206996.2; 1p12), respectively. The change of the polar and neutral amino acid asparagine by the basic and positive amino acid lysine at position 276, is predicted by MutationTaster and by Polyphen-2 (p=0.99 and s=0.72, respectively) as being possibly pathogenic, however according to SIFT (s=0.38) no pathogenic impact is expected for this variant. For the variant c.4445G>T no pathogenic impact is predicted, despite the change of an arginine to a leucine at position 1482.

Table V. List of the variants detected by exome sequencing and its respective bioinformatic analysis.

Gene	DNA change/ Zigoty	Protein change[a]	Freq.*	SIFT	PolyPhen2	Mutation Taster	Phenotype associated
CCDC103	c.104G>C Homozygous	p.Arg35Pro	New variant	D (s=0)	D (s= 0.99)	Disease Causing (p=0.99)	PCD (Panizzi *et al.*, 2012)
INSL6	c.262_263delCC Heterozygous	p.Pro88Glyfs*27	New variant	na	na	Disease Causing (p=1)	Spermatogenic Failure (Lok *et al.*, 2000; Chen *et al.*, 2011)
DNAH10	c.7895C>T Heterozygous	p.Thr2632Met	0.13%	D (s=0.02)	D (s= 0.98)	Disease Causing (p=0.99)	no phenotype associated
DNAH6	c.3167A>T Heterozygous	p.Asp1056Val	New variant	D (s=0)	D (s= 0.98)	Disease Causing (p=0.99)	no phenotype associated
GAS8	c.828C>G Heterozygous	p.Asn276Lys	New variant	Tolerated (s=0,38)	Possibly Damaging (s= 0.72)	Disease Causing (p=0.99)	Sperm immotility/PCD (Yeh *et al.*, 2002; Horani *et al.*, 2013)
SPAG17	c.4445G>T Heterozygous	p.Arg1482Leu	New variant	Tolerated (s=0,23)	Bening (s=0.005)	P (p=0.9)	PCD (Teves *et al.*, 2013)

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)); In bold blue are the ones most promising in terms of predicted pathogenic impact.
Abbreviators: a, *in silico* prediction; D, Damaging; s, score; p, probability; PCD, primary ciliary dyskinesia; na, non-applicable.

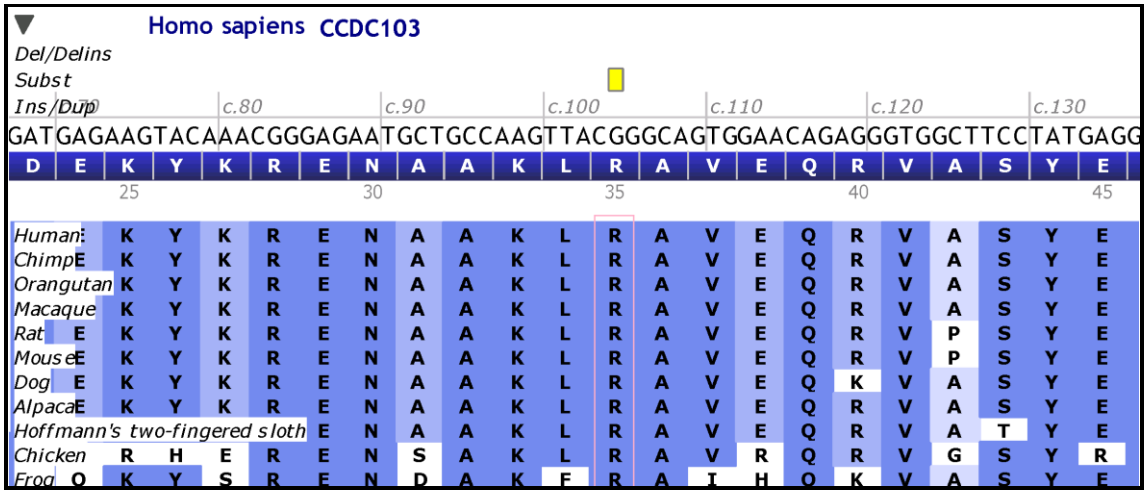


Figure 14. Multiple sequence alignment of the region of the CCDC103 protein sequence altered in the variant p.Arg35Pro, showing the high degree of conservation of this region. Adapted from Alamut Visual V2.4 Software.

IV. DISCUSSION

Discussion

The goal of this work was to genetically characterize five patients from two Portuguese assisted reproductive centres in order find a genetically explanation to the sperm immotility of our patients. In four unrelated subjects were diagnostic with DFS (n=4, patient 1 to 4) and one have clinical features compatible Kartagener Syndrome (KS, a case of PCD) (n=1, patient 5). To achieve this we initially selected seven genes associated to anomalies in sperm flagellum, namely in FS (*AKAP3* and *AKAP4* genes) and Ax (*CCDC39, CCDC40, DNAI, DNAH5* and *RSPH1* genes), and analysed all exonic regions of selected genes in all subjects by Sanger Sequencing. Next, we also performed whole-exome analysis using next-generation sequencing (NGS) technology in the KS patient with *situs inversus* (patient 5).

With this work we identified **nine new** DNA sequence variants (c.2540A>G in *CCDC39*, c.2620-92C>T in *CCDC40*, c.104G>C in *CCDC103*, c. 5882+133A>G and c.11570+124G>A in *DNAH5,* c.3167A>T in *DNAH6*, c.828C>G in *GAS8*, c.262_263delCC in *INSL6* and c.4445G>T in *SPAG17*) and **ten rare** variants (c.233G>A in *CCDC39*; c.2682G>A in *CCDC40;* c.1537-102T>A, c.1537-100_1537-99delTT, c.3835-3delT, c.7408-84_7408-83delAT , c.10282-81delT, c.10872+84T>C in *DNAH5*; c.7895C>T in *DNAH10* and c.81+61A>G in *DNAI1*).

The variants c.2540A>G in *CCDC39*, c.104G>C in *CCDC103*, c.3167A>T in *DNAH6*, c.7895C>T in *DNAH10* and c.262_263delCC in *INSL6* are predicted to be deleterious.

The novel heterozygous variant **c.2540A>G** in *CCDC39* gene was found in patient 1. This patient was characterized has having hyperplastic and hypertrophic FS and sperm Ax without the CPC and the RS, and microtubule doublets were disorganized. Besides changing the amino acid glutamic-acid to glycine at position 847, according to Human Splice Finder (HSF), its variant may slightly increase the score of an alternative splice factor ASF/SF2. This factor is required for spliceosome assembly and are able to influence selection of alternative splice sites (Shepard and Hertel, 2009; Zhou and Fu, 2013). A previous report showed that ASF/SF2 intervenes in the maintenance of genomic stability and preventing DNA fragmentation by protecting cells from the deleterious effects of transcription itself (Li and Manley, 2005). Consequently, it could have an impact on normal splicing and thus modify the protein function. This same patient also presents a heterozygous DNA variant **c.81+61A>G** in *DNAI1* gene that is predicted to make the mt SSA score almost equivalent to the nt (mt: 82.97; nt: 83.19). This increase could lead to a competition for the splice site and eventually have

an impact at the mRNA level. Both genes are essential components of DA and important for sperm motility. *CCDC39* gene is located at chromosome 3q26.33 and is strongly expressed in tissues rich in ciliated cells, namely in nasal brushings, lungs and testis (Merveille *et al.*, 2011). Previous reports state that *CCDC39* is an axonemal protein whose absence results in disorganization of the peripheral microtubule doublets, absent or shifted CPC, partial or complete loss of IDA and the RS, thus causing Ax disorganization and dyskinetic beating (Merveille *et al.*, 2011; Antony *et al.*, 2013). *DNAI1*, located at chromosome 9p13.3, encodes an axonemal dynein intermediate chain (IC) and is highly expressed in trachea and testis. This gene is associated to sperm immotility due to the absence of ODA (Pennarun *et al.*, 1999). Despite their important roles on sperm immotility and the predicted bioinformatic impact of these variants, these are unlike to explain the patient's phenotype considering that they are heterozygous. Further experimentally studies are required to conclude about the real pathogenicity of both variants.

The heterozygous deletion of a T in *DNAH5* gene (**c.3835-3delT**) was present in four patients (patient 1 to 4) whose analyses of transmission electronic microscopy revealed hyperplastic and hypertrophic FS. Despite being present in four of five patients, this variant has a very low frequency (< 0.1%) within a large control sample. Bioinformatic analysis for this variant suggested a possible splicing effect but there was only a marginal reduction of the score of the nt SSA. It is also considered as likely to cause disease by MutationTaster. Here again, based on this data alone it is not possible to classify this variant as a polymorphism, neither to conclude about its pathogenic impact. Consequently, and considering its high frequency within our patients, this variant requires further studies.

There are some inconsistencies between the ultrastructural spermatic changes found in the patients and the feature previously reported linked to *DNAH5* mutations. The gene *DNAH5*, located at chromosome 5p15.2, encodes for an axonemal HC dynein protein that is strongly expressed in lung and kidney and weakly expressed in brain, heart, and testis (Olbrich *et al*., 2002). It is the most mutated gene in PCD patients that exhibit ODA defects (Hornef *et al*., 2006; Djakow *et al*., 2012). However, only one of the four patients lacked DA in its sperm Ax. We might speculate if a heterozygous change in this gene, may not be sufficient to cause the absence of DA but can have an effect, together with other genetic modifiers, on the flagellar movement. Two previous works detected mutations in another gene, *DNAH11* that encodes for an axonemal HC dynein protein in patients with abnormal ciliary beating pattern but with a typical axonemal ultrastructure (Schwabe *et al.*, 2008; Knowles *et al.*, 2012). Nevertheless, it is important to consider that the respiratory tract of our patients

was not clinically evaluated, and that sperm and respiratory cells carrying the same mutations may have distinct phenotypes regarding the DA (Fliegauf *et al.*, 2005).

In this *DNAH5* gene, we were initially expecting to find pathogenic variants, particularly on patients 3 and 5 whose sperm Ax lack DA. However, as previously stated *DNAH5* gene is a large gene with seventy nine exons and only half of these were sequenced in this work. The majority of the reported mutations were described in the selected exons (Djakow *et al.,* 2012). However, we cannot exclude the hypotheses that the remaining thirty-nine exons might carry variants that could be involved in the phenotype.

We also raised the hypothesis that mutations in *AKAP3* and *AKAP4* genes, most abundant structural proteins of the FS and with important roles in sperm motility, could explain the phenotypes of this patients (1 to 4), (Brown *et al.*, 2003; Eddy *et al.*, 2003). This possibility was based on several previous observations. Miki and collaborators, demonstrated that male mice lacking AKAP4 protein were infertile due to reduced sperm motility and by development anomalies in FS (Miki *et al.*, 2002). Baccetti *et a*l detected deletions in *AKAP3/4* binding regions and a moderate diffuse signal after immunostaining for human AKAP4 protein in DFS patients (Baccetti *et al*., 2005a, 2005b). However, our results, in a limited number of patients, did not support the hypotheses that *AKAP3* and *AKAP4* are major genetic causes of DFS. Our results are in agreement with Turner *et al* that did not detected mutations in these genes as well (Turner *et al.*, 2001). Moretti *et al* studied a group of patients with high presence of sperm necrosis and total immotility where AKAP4 protein or tubulin labelling was absent, but similarly, they did not detected gene mutations (Moretti *et al.*, 2007). This demonstrates that the role of *AKAP3/4* is still unclear and leads us to postulate that DFS is a multigenic trait that may be caused by mutations in (several) other genes yet to be identified related to FS function and/or assemble. FS assembly proceeds from distal to proximal along the Ax and occurs during the spermiogenesis, at acrosomal phase. Thus, the FS defects could be caused by mutations in spermiogenic genes related to the FS, for instance, genes related to formation, transport or attachment of the longitudinal columns or in genes that command the development of the flagellum. Further studies about molecular elements involved in human FS development and structure as well as studies about mechanisms of intraflagellar transport of cell factors necessary for the FS and flagellum formation, need to be done to fully understand the genetic cause of the DFS.

Regarding patient 5, who presented *situs inversus*, severe respiratory symptoms (nasal polyps, chronic sinusitis, rhinitis and bronchitis) and its sperm flagellum lacked the DA and nexin bridges, no candidate DNA variants were detected

in selected genes that may justify the patient's phenotype. This was not anticipated by our group given the function of selected genes and the patient's phenotype. On the other hand, considering the complexity of the sperm Ax structures and of the sperm motility process, is likely that several genes are involved in formation/ assembly/ transport of Ax structures and consequently several variants in multiple genes can give rise to this phenotype. Hence, to increase our knowledge about which are these genes, we decided to perform an ES analysis. From the six variants confirmed by Sanger sequencing, only the missense variant p.Arg1482Leu in *SPAG17*, were predicted to be a polymorphism by all bioinformatic software.

The *GAS8* gene (originally designated GAS11), located at chromosome 16q24.3, is expressed predominantly in testis. The heterozygous variant **c.828C>G,** is a missense variant that leads to an amino acid change at protein position 276 (**p.Asn276Lys**) and occurs in a growth-arrest specific microtubule binding domain (pfam accession number (http://pfam.xfam.org/:13851) that is a highly conserved central region among metazoan. Two bioinformatic tools predict as being possibly pathogenic (see Table V). Immunohistochemical analyses showed that *GAS8* is present in spermatids, in the tail of the spermatids and also in flagella of mature spermatozoa (Yeh *et al.*, 2002). *GAS8* may be associated with tail maturation and may be involved in the attachment of the dynein regulatory complex (DRC) to microtubules (Yeh *et al.*, 2002; Bekker *et al.*, 2007). Studies in Zebrafish lines demonstrated that *GAS8* is required for normal motility of cilia (Colantonio *et al.*, 2008). Altogether, these facts suggest that *GAS8* might be important for motility.

The heterozygous missense variants **p.Thr2632Met** in *DNAH10* gene and **p.Asp1056Val** in *DNAH6* gene are both located in genes encoding for dynein HC protein. Each HC comprises[2] an N-terminal domain; a C-terminal segment, that includes a highly conserved central section with six AAA modules (the modules 1 to 4 have P-loop nucleotide binding motifs (P1-4)); and a microtubule-binding domain (Kim, 2000). According to NCBI Conserved Domain Database (http://www.ncbi.nlm.nih.gov/cdd/), the amino acid change in DNAH10 protein at position 2632, occurs in the third AAA module (an AAA_7 super family domain; P3), which is a highly conserved region and its P-loop includes an ATP binding site. Silvanovich *et al* hypothesized that in cytoplasmic dynein, the P3 function is required for the ATP-induced release of the dynein complex from microtubules and showed that mutation of P3 appears to block the ATP binding and hydrolysis at P-loop 1, which, if it is true also for axonemal dyneins, could ultimately affect the motility (Silvanovich *et al.*,

[2] See introduction for more detailes

2003). Relatively to the amino acid change in DNAH6 protein at 1056 position, it is localized in the N-terminal region, which is important to specify the intracellular location of the dynein isoform by the binding of accessory proteins to this domain and is required for assembly of the dynein particle (Asai and Koonce, 2001). As far as we know, this is the first time that these two dynein HC genes are associated to sperm motility in Humans. However, RNAi knockdown of *DNAH10* in *Trypanosoma brucei* leads to flagellum immotility and to structural defects in the Ax (Zukas *et al.*, 2012). Despite the role of that these genes may play in DA, considering the autosomal recessive inheritance pattern of the KS, is questionable that these two heterozygous variants are sufficient to cause the totally lack of DA. Unless that the predicted interaction of these two proteins, as suggested by STRING, has some biological meaning.

In *CCDC103* gene, the conversion of an arginine by a proline at protein position 35, caused by the homozygous variant **c.104G>C**, according to Polyphen-2 and SIFT is predicted to generate a protein damage with scores of 0.99 and 0, respectively. MutationTaster also predicted that this protein may cause a disease phenotype (p=0.99). Multiple sequence alignment of the protein sequences shows that the arginine residue is highly conserved (Fig. 14). The *CCDC103* gene is a DA attachment factor, whose mutations cause absence or severe defects in the IDA and/or ODA in PCD patients with *situs inversus* and paralysis of respiratory cell cilia beat (Panizzi *et al.*, 2012). Altogether, the available data suggests that this variant may be responsible for the lack of DA in our patient. Panizzi *et al* did not focus their study in fertility; neither had we found any study that associated sperm immotility to DNA sequence variants in this gene. As result, this is the first report of a DNA sequence variant in *CCDC103* gene that might be associated to sperm motility in a patient with total sperm immotility, absence of DA and *situs inversus.*

The *INSL6* gene, located at chromosome 9p24, belong to the relaxin family of peptide hormones whose members are involved in several reproductive functions (Anand-Ivell *et al.*, 2013; Anand-Ivell and Ivell, 2014) and is predominantly expressed in male germ cells (Lok *et al.*, 2000). Burnicka-Turek *et al* using Insl6-null mice demonstrated that Insl6 deficiency causes an arrest of spermatogenesis at late stages of meiotic prophase, leading to a reduction of the spermatozoa production and causing immotility to the ones that are actually produced. Hence, they proposed that *INSL6* gene is required for normal spermatogenesis (Burnicka-Turek *et al.*, 2009). Chen and collaborators found a heterozygous missense mutation in exon 2 of *INSL6* that changed the arginine residue at position 171 to histidine, in one patient with spermatogenic failure. In their work, they proposed that this mutation might be

responsible for its phenotype, due to the probable disruption of the processing of INSL6 pro-hormone (Chen *et al.*, 2011). We found a heterozygous deletion of two C nucleotides in cDNA position 262_263, that disrupts the reading frame and may affect the protein function. Thus, ultimately, according to previous reports about its function, this deletion is likely to disrupt the spermatogenesis and hence contributes to sperm immotility of our patient.

Overall, the homozygous variant c.104G>C of *CCDC103* gene and the frame-shift variant c.262_263delCC in *INSL6* gene, were both considered as the most plausible causes for the absence of DA and for the total sperm immotility in patient 5.

Here again, to understand the real impact of all predicted pathogenic variants, detected in this study, and as a proposal of future work, further studies, namely expression studies, for instance study the RNA to analyse if the variants indeed affect the splicing and thus altered the normal expression of the mRNA; and also studies using knock-out animal models, need to be carried out.

Unfortunately, patient's RNA was inaccessible to continue our investigation, and thus considered beyond the scope of this work. However, RNA samples obtained from sperm or testicular cells are important to provide further evidences about the effect produced by these variants and clarify about its impact. RNA studies performed in fertile and infertile males have already proven to be useful to identify differentially expressed genes (Carreau *et al,* 2007; Garrido *et al*, 2009; Hamatani, 2012). Nevertheless, this is not a simple analysis, since sperm contains very limited amounts of messenger RNA (mRNA), roughly estimated to contain about 0.015 pg of mRNA. Spermatozoon is now considered as a dynamic cell that, under specific conditions, RNA transcriptional changes occur in response to environmental stimulate and epigenetic alterations (Miller *et al.*, 2005; Miller and Ostermeier, 2006).

The knock-out animal models could be a good alternative to (or combined with) RNA analysis to determine the impact of these variants. Mouse models have long been important tools in medical research and there are a number of reports that used them to study genes involved in male infertility (Escalier, 2001, 2006; Tamowski *et al.*, 2010). Also *Chlamydomonas reinhardtii* has proved to be an excellent model for studying the molecular components of Ax and to the identification of genes involved in axonemal assembly (Inaba, 2007, 2011).

Besides the lack of these experimental studies, we also consider relevant to study the information about the patient's family history, which could help drawing a pedigree chart and also stablish the disease's inheritance pattern. Samples from parents would also be very helpful to understand if a particular genetic variant was inherited from their parents or arose as *de novo*. Depending on the patient's pedigree,

the interpretation about the pathogenicity of the variants would have been more accurate.

In this work we used two different sequencing technologies: the automated Sanger sequencing and NGS. The automated Sanger sequencing method is based on the original method developed by Frederick Sanger during the 1970s. This approach generates DNA fragments labelled with a fluorescent dye, and then DNA sequencers carry out capillary electrophoresis for size separation, detection of dye fluorescence, in order to create electropherograms. This method had dominated Genetics for the last two decades and is still considered the "gold-standard" for genetic testing and also used to confirm NGS results, since it is highly accurate. Nevertheless, Sanger sequencing has some limitations such as its costs, time consumption and the low throughput, when comparing with NGS, in which many genes are analysed at once.

In order to increase our knowledge about genetics of sperm immotility and considering the genetic heterogeneity of PCD and the highly complexity of FSD and sperm motility, that are probably inherited as a complex multigenic traits we decided to use the NGS technologies. It generates an enormous amount of data in a more cost-effective and faster way (Metzker, 2010) than Sanger sequencing. Among NGS technologies available, ES has already demonstrated great potential to identify genes associated to Mendelian disease, as is an efficient strategy to selectively sequence the entire coding regions of the genome (Bamshad *et al.*, 2011, Ku *et al.*, 2012a). Comparing with other NGS technologies, such as Whole Genome Sequencing (WGS), ES is currently the most effective technical approach available since it reduces the amount of data, therefore making the bioinformatic analysis easier, as well as, being less expensive (Bamshad *et al.*, 2011; Parla *et al.*, 2011). ES has already contributed to the identification of new genes involved in PCD (Horani *et al*., 2012; Knowles *et al*., 2013; Kott *et al.,* 2013; Onoufriadis *et al.,* 2014). Nevertheless, ES use still faces some limitations, such the considerable rate of false-positive variants because of mismapped/aligned reads or sequencing errors (Bamshad *et al.,* 2011). In addition, ES only covers the exonic regions, meaning that approximately 99% of the human genome is not included in this assay. Thereby ES is not able to identify the structural and non-coding variants that may be associated with the disease. It is conceivable that, with the reduction of costs and the development of more powerfully and robust bioinformatic tools, WGS may eventually gain insights and become a standard approach for these studies.

As a future work, it was already discussed that would be important to perform further experimental analysis and analyse the patient's familiy. In addition, it will be important to further clinically characterize these patients and perform ES analysis on

the remaining four. Moreover, it would be also interesting to study the transcriptome of sperm or testis cells, for example using RNA-Seq (RNA sequencing method based on NGS). It is known that human genome is composed of regulatory regions that are not part of the exome and mutations in those regions (e.g., promoter) can potentially affect protein production and consequently their function. RNA-seq allows the discovery of new transcripts, the study of alternative splicing patterns, the detection of fusion transcripts and allele-specific expression analysis. However, it should be done only as a complementary analysis since the transcriptome is very dynamic changing with time and tissue/cell type depending on the conditions. Transcriptome sequencing is cheaper and avoids some technical limitations of ES such as enrichment steps, but does not detect variants in non-transcribed genes (Ku *et al.*, 2012b) or mutations causing non-sense mediated decay. Consequently, and contrasting with ES, it is not widely used to detect variants underlying Mendelian disorders.

The knowledge of the genetic basis of sperm immotility is important to alert the patients that need to use Assisted Reproductive techniques of risks of transmitting the mutation to its offspring. Moreover, understanding which genes are involved in sperm motility, how they are correlated as well as how exactly protein works and how these variants may affect its function, is essential to plan and develop, in the future, a gene therapy. Several advances have been done to develop genetic therapies for diseases caused by single-gene defects, such as Cystic Fibrosis (Ikpa *et al.*, 2014), haemophilia (High *et al.*, 2014), muscular dystrophy (Sumner, 2006), and sickle cell anemia (Frenette and Atweh, 2007). However, in more complex genetic disease, we are still in the beginning and much work still yet to be done.

To conclude, our study typifies the difficulties to study clinical and genetically heterogeneous diseases (namely DFS and PCD) or complex phenotypical traits (such as sperm immotility) by conventional (Sanger) sequencing. Sperm immotility may be caused by defects in multiple genes and/or by the interaction of multiple variants located in different genes. In our study, **nine new DNA sequence variants** were identified, with two in particular being considered the most likely genetic defects that explain the phenotype of the patient with *situs inversus* and lack of DA. Finally, this dissertation corroborates that ES is the most efficient approach to study these pathologies, allowing further knowledge about the genetic causes of sperm immotility and infertility.

V. REFERENCES

REFERENCES

Abou-Haila a, Tulsiani DR. Mammalian sperm acrosome: formation, contents, and function. *Arch Biochem Biophys* 2000, **379**:173–182.

Adham IM, Nayernia K, Engel W. Spermatozoa lacking acrosin protein show delayed fertilization. *Mol Reprod Dev* 1997, **46**:370–376.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010, **7**:248–249.

Afzelius BA, Eliasson R, Johnsen O, Lindholmer C. Lack of dynein arms in immotile human spermatozoa. *J Cell Biol* 1975, **66**:225–232.

Afzelius BA. A human syndrome caused by immotile cilia. *Science* 1976, **193**:317–319.

Afzelius BA, Srurgess JM. The immotile-cilia syndrome: a microtubule-associated defect. *Crit Rev Biochem Mol Biol* 1985, **19**:63–87.

Afzelius BA, Dallai R, Lanzavecchia S, Bellon PL. Flagellar structure in normal human spermatozoa and in spermatozoa that lack dynein arms. *Tissue&Cell* 1995, **27**:241–247.

Aitken RJ, Nixon B, Lin M, Koppers AJ, Lee YH, Baker MA. Proteomic changes in mammalian spermatozoa during epididymal maturation. *Asian J Androl* 2007, **9**:554–564.

Anand-Ivell R, Dai Y, Ivell R. Neohormones as biomarkers of reproductive health. *Fertil Steril* 2013, **99**:1153–1160.

Anand-Ivell R, Ivell R. Regulation of the reproductive cycle and early pregnancy by relaxin family peptides. *Mol Cell Endocrinol* 2014, **382**:472–479.

Antony D, Becker-Heck A, Zariwala MA, Schmidts M, Onoufriadis A, Forouhan M, Wilson R, Taylor-Cox T, Dewar A, Jackson C, *et al.* Mutations in CCDC39 and CCDC40 are the Major Cause of Primary Ciliary Dyskinesia with Axonemal Disorganization and Absent Inner Dynein Arms. *Hum Mutat* 2013, **34**:462–472.

Amaral A, Lourenço B, Marques M, Ramalho-Santos J. Mitochondria functionality and sperm quality. Reproduction 2013;146:R163–R174.

Aoki VW, Moskovtsev SI, Willis J, Liu L, Mullen JBM, Carrell DT. DNA Integrity Is Compromised in Protamine-Deficient Human Sperm. *J Androl* 2005, **26**:741–748.

Aoki VW, Emery BR, Liu L, Carrell DT. Protamine Levels Vary Between Individual Sperm Cells of Infertile Human Males and Correlate With Viability and DNA Integrity. *J Androl* 2006, **27**:890–898.

Asai DJ, Koonce MP. The dynein heavy chain: structure, mechanics and evolution. *Trends Cell Biol* 2001, **11**:196–202.

Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, Davis A, Dolinski K, Dwight S, Eppig J, *et al*. (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25, 25–29.

Auger J, Dadoune J-P. Nuclear status of human sperm cells by transmission electron microscopy and image cytometry: changes in nuclear shape and chromatin texture during spermiogenesis and epididymal transit. *Biol Reprod* 1993, **49**:166–175.

Austin-Tse C, Halbritter J, Zariwala MA, Gilberti RM, Gee HY, Hellman N, Pathak N, Liu Y, Panizzi JR, Patel-King RS. Zebrafish Ciliopathy Screen Plus Human Mutational Analysis Identifies C21orf59 and CCDC65 Defects as Causing Primary Ciliary Dyskinesia. *Am J Hum Genet* 2013, **93**:672–686.

Baccetti B, Collodel G, Estenoz M, Manca D, Moretti E, Piomboni P. Gene deletions in an infertile man with sperm fibrous sheath dysplasia. *Hum Reprod* 2005a, **20**:2790–2794.

Baccetti B, Collodel G, Gambera L, Moretti E, Serafini F, Piomboni P. Fluorescence in situ hybridization and molecular studies in infertile men with dysplasia of the fibrous sheath. *Fertil Steril* 2005b, **84**:123–129.

Baltz JM, Williams PO, Cone RA. Dense fibers protect mammalian sperm against damage. *Biol Reprod* 1990, **43**:485–491.

Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011, **12**:745–755.

Becker-Heck A, Zohn IE, Okabe N, Pollock A, Lenhart KB, Sullivan-Brown J, McSheene J, Loges NT, Olbrich H, Haeffner K, *et al.* The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis formation. *Nat Genet* 2011, **43**:79–84.

Beecher KL, Homyk M, Lee CG, Herr JC. Evidence that 68-kilodalton and 54-51-kilodalton polypeptides are components of the human sperm fibrous sheath. *Biol Reprod* 1993, **48**:154–164.

Bekker JM, Colantonio JR, Stephens AD, Clarke WT, King SJ, Hill KL, Crosbie RH. Direct interaction of Gas11 with microtubules: implications for the dynein regulatory complex. *Cell Motil Cytoskeleton* 2007, **64**:461–473.

Berg JS, Evans JP, Leigh MW, Omran H, Bizon C, Mane K, Knowles MR, Weck KE, Zariwala MA. Next generation massively parallel sequencing of targeted exomes to identify genetic mutations in primary ciliary dyskinesia: implications for application to clinical testing. *Genet Med* 2011, **13**:218–229.

Bettencourt-Dias M, Rodrigues-Martins A, Carpenter L, Riparbelli M, Lehmann L, Gatt MK, Carmo N, Balloux F, Callaini G, Glover DM. SAK/PLK4 is required for centriole duplication and flagella development. *Curr Biol* 2005, **15**:2199–2207.

Bettencourt-Dias M, Glover DM. Centrosome biogenesis and function: centrosomics brings new understanding. *Nat Rev Mol Cell Biol* 2007, **8**:451–463.

Birkhead TR, Montgomerie R. Three centuries of sperm research. In Birkhead TR, Hosken DJ, Pitnick SS, editors. *Sperm Biol an Evol Perspect* 2008, p. 1–43

Boon M, Jorissen M, Proesmans M, Boeck K. Primary ciliary dyskinesia, an orphan disease. *Eur J Pediatr* 2013, **172**:151–162.

Brown PR, Miki K, Harper DB, Eddy EM. A-kinase anchoring protein 4 binding proteins in the fibrous sheath of the sperm flagellum. *Biol Reprod* 2003, **68**:2241–2248.

Brownlee CW, Rogers GC. Show me your license, please: deregulation of centriole duplication mechanisms that promote amplification. *Cell Mol Life Sci* 2013, **70**:1021–1034.

Burgess SA, Walker ML, Sakakibara H, Knight PJ, Oiwa K. Dynein structure and power stroke. *Nature* 2003, **421**:715–718.

Burgess SA, Knight PJ. Is the dynein motor a winch? *Curr Opin Struct Biol* 2004, **14**:138–146.

Burnicka-Turek O, Shirneshan K, Paprotta I, Grzmil P, Meinhardt A, Engel W, Adham IM. Inactivation of insulin-like factor 6 disrupts the progression of spermatogenesis at late meiotic prophase. *Endocrinology* 2009, **150**:4348–4357.

Capanna E. Lazzaro Spallanzani: At the roots of modern biology. *J Exp Zool* 1999, **285**:178–196.

Carreau S, Lambard S, Said L, Saad A, Galeraud-Denis I. RNA dynamics of fertile and infertile spermatozoa. *Biochemical Society Transactions* 2007, **35:**634-636.

Carrell DT, Liu L. Altered Protamine 2 Expression Is Uncommon In Donors of Known Fertility, but Common Among Men With Poor Fertilizing Capacity, and May Reflect Other Abnormalities of Spermiogenesis. *J Androl* 2001, **22**:604–610.

Carvalho CE de, Tanaka H, Iguchi N, Ventelä S, Nojima H, Nishimune Y. Molecular cloning and characterization of a complementary DNA encoding sperm tail protein SHIPPO 1. *Biol Reprod* 2002, **66**:785–795.

Casey DM, Inaba K, Pazour GJ, Takada S, Wakabayashi K, Wilkerson CG, Kamiya R, Witman GB. 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. *Mol Biol Cell* 2003, **14**:3650–3663.

Castleman VH, Romio L, Chodhari R, Hirst RA, Castro SCP De, Parker KA, Ybot-gonzalez P, Emes RD, Wilson SW, Wallis C, *et al.* Mutations in Radial Spoke Head Protein Genes RSPH9 and RSPH4A Cause Primary Ciliary Dyskinesia with Central-Microtubular-Pair Abnormalities. *Am J Hum Genet* 2009, **84**:197–209.

Chao H-CA, Lin Y-H, Kuo Y-C, Shen C-J, Pan H-A, Kuo P-L. The expression pattern of SEPT7 correlates with sperm morphology. *J Assist Reprod Genet* 2010, **27**:299–307.

Chemes HE, Olmedo SB, Carrere C, Oses R, Carizza C, Leisner M, Blaquier J. Ultrastructural pathology of the sperm flagellum : association between flagellar pathology and fertility prognosis in severely asthenozoospermic men. *Hum Reprod* 1998, **13**:2521–2526.

Chemes HE. Phenotypes of Sperm Pathology : Genetic and Acquired Forms in Infertile Men. *J Androl* 2000, **21**:799–808.

Chemes HE, Rawe VY. Sperm pathology: a step beyond descriptive morphology. Origin, characterization and fertility potential of abnormal sperm phenotypes in infertile men. *Hum Reprod Update* 2003, **9**:405–428.

Chemes HE, Rawe VY. The making of abnormal spermatozoa: cellular and molecular mechanisms underlying pathological spermiogenesis. *Cell Tissue Res* 2010, **341**:349–357.

Chemes, HE. Sperm centrioles and their dual role in flagellogenesis and cell cycle of the zygote. In Heide S, editor. *The Centrosome: Cell and Molecular Mechanisms of Functions and Dysfunctions in Disease* 2012, p. 33-48.

Chen G-W, Luo X, Liu Y-L, Jiang Q, Qian X-M, Guo Z-Y. R171H missense mutation of INSL6 in a patient with spermatogenic failure. *Eur J Med Genet* 2011, **54**:e455–e457.

Chocu S, Calvel P, Rolland AD, Pineau C. Spermatogenesis in mammals : proteomic insights. *Syst Biol Reprod Med* 2012, **58**:179–190.

Christensen AK. A history of Leydig cell research. *Leydig Cell Heal Dis* 2007, 3–30.

Colantonio JR, Vermot J, Wu D, Langenbacher AD, Fraser S, Chen J-N, Hill KL. The dynein regulatory complex is required for ciliary motility and otolith biogenesis in the inner ear. *Nature* 2008, **457**:205–209.

Cooper G. Mitochondria. In *Cell A Mol Approach,* 2002, Available from: http://www.ncbi.nlm.nih.gov/books/NBK9896/.

Copeland WC. Inherited mitochondrial diseases of DNA replication. *Annu Rev Med* 2008, **59**:131–146.

Curi SM, Ariagno JI, Chenlo PH, Mendeluk GR, Pugliese MN, Sardi Segovia LM, Repetto HEH, Blanco AM. Asthenozoospermia: analysis of a large population. *Syst Biol Reprod Med* 2003, **49**:343–349.

De Jonge CJ, Tarchala SM, Rawlins RG, Binor Z, Radwanska E. Acrosin activity in human spermatozoa in relation to semen quality and invitro fertilization. *Hum Reprod* 1993, **8**:253–257.

De Kretser DM, Loveland KL, Meinhardt A, Simorangkir D, Wreford N. Spermatogenesis. *Human Reproduction* 1998, **13**:1-8.

Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009, **37**:e67–e67.

DiPetrillo CG, Smith EF. Pcdp1 is a central apparatus protein that binds Ca2+-calmodulin and regulates ciliary motility. *J Cell Biol* 2010, **189**:601–612.

DiPetrillo CG, Smith EF. The Pcdp1 complex coordinates the activity of dynein isoforms to produce wild-type ciliary motility. *Mol Biol Cell* 2011, **22**:4527–4538.

Djakow J, Svobodová T, Hrach K, Uhlík J, Cinek O, Pohunek P. Effectiveness of sequencing selected exons of DNAH5 and DNAI1 in diagnosis of primary ciliary dyskinesia. *Pediatr Pulmonol* 2012, **47**:864–875.

Doggett NA, Xie G, Meincke LJ, Sutherland RD, Mundt MO, Berbari NS, Davy BE, Robinson ML, Rudd MK, Weber JL. A 360-kb interchromosomal duplication of the human HYDIN locus. *Genomics* 2006, **88**:762–771.

Donnell LO, Nicholls PK, Bryan MKO, Mclachlan RI, Stanton PG. Spermiation: The process of sperm release. *Spermatogenesis* 2011, **1**:14–35.

Duriez B, Duquesnoy P, Escudier E, Bridoux A-M, Escalier D, Rayet I, Marcos E, Vojtek A-M, Bercher J-F, Amselem S. A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia. *Proc Natl Acad Sci* 2007, **104**:3336–3341.

Dutcher SK. Flagellar assembly in two hundred and fifty easy-to-follow steps. *Trends Genet* 1995, **11**:398–404.

Dutcher SK. The tubulin fraternity: alpha to eta. *Curr Opin Cell Biol* 2001, **13**:49–54.

Eddy EM, Toshimori K, O'Brien D a. Fibrous sheath of mammalian spermatozoa. *Microsc Res Tech* 2003, **61**:103–115.

Escalier D, Albert M. New fibrous sheath anomaly in spermatozoa of men with consanguinity. *Fertil Steril* 2006, **86**:219–e1.

Escalier D. Impact of genetic engineering on the understanding of spermatogenesis. *Hum Reprod Update* 2001, **7**:191–210.

Escalier D. Knockout mouse models of sperm flagellum anomalies. *Hum Reprod Update* 2006, **12**:449–461.

Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998, **8**:186–194.

Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment. *Genome Res* 1998, **8**:175–185.

Failly M, Saitta A, Munoz A, Falconnet E, Rossier C, Santamaria F, Santi MM de, Lazor R, DeLozier-Blanchet CD, Bartoloni L. DNAI1 mutations explain only 2% of primary ciliary dykinesia. *Respiration* 2008, **76**:198–204.

Failly M, Bartoloni L, Letourneau A, Munoz A, Falconnet E, Rossier C, Santi MM De, Santamaria F, Sacco O, DeLozier-Blanchet CD. Mutations in DNAH5 account for only 15% of a non-preselected cohort of patients with primary ciliary dyskinesia. *J Med Genet* 2009, **46**:281–286.

Fawcett DW. A comparative view of sperm ultrastructure. *Biol Reprod* 1970, **2**:90–127.

Flesch FM, Gadella BM. Dynamics of the mammalian sperm plasma membrane in the process of fertilization. *Biochim Biophys Acta* 2000, **1469**:197–235.

Fliegauf M, Olbrich H, Horvath J, Wildhaber JH, Zariwala MA, Kennedy M, Knowles MR, Omran H. Mislocalization of DNAH5 and DNAH9 in respiratory cells from patients with primary ciliary dyskinesia. *Am J Respir Crit Care Med* 2005, **171**:1343–1349.

Ford WCL. Glycolysis and sperm motility: does a spoonful of sugar help the flagellum go round? *Hum Reprod Update* 2006, **12**:269–274.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Jensen L. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research* 2013, **41**:D808-D815.

Frenette PS, Atweh GF. Sickle cell disease: old discoveries, new concepts, and future promise. *J Clin Invest* 2007, **117**:850–858.

Fuentes-mascorro G, Serrano H, Rosado A. Sperm chromatin. *Archives of andrology* 2000, 45:215–225

Fulcher KD, Welch JE, Klapper DG, O'Brien DA, Eddy EM. Identification of a unique μ-class glutathione S-transferase in mouse spermatogenic cells. *Mol Reprod Dev* 1995, **42**:415–424.

Gagnon C, White D, Cosson J, Huitorel P, Eddé B, Desbruyères E, Paturle-Lafanechère L, Multigner L, Job D, Cibert C. The polyglutamylated lateral chain of alpha-tubulin plays a key role in flagellar motility. *J Cell Sci* 1996, **109**:1545–1553.

Gagnon C, Lamirande E de. Controls of sperm motility. In Jonge CJ De, Barratt CLR, editors. *Sperm Cell- Prod Matur Fertil Regen* 2006, p. 108–134.

Gaillard AR, Diener DR, Rosenbaum JL, Sale WS. Flagellar Radial Spoke Protein 3 Is an a-Kinase Anchoring Protein (Akap). *J Cell Biol* 2001, **153** :443–448.

Gardner LC, O'Toole E, Perrone CA, Giddings T, Porter ME. Components of a" dynein regulatory complex" are located at the junction between the radial spokes and the dynein arms in Chlamydomonas flagella. *J Cell Biol* 1994, **127**:1311–1325.

Garrido N, Martinez-Conejero J, Jauregui J, Horcajadas JA, Simon C, Remohi J, Meseguer M. Microarray analysis in sperm from fertile and infertile men without basic sperm analysis abnormalities reveals a significantly different transcriptome. *Fertility and sterility* 2009, **91**, 1307-1310.

Gaspar P, Lopes P, Oliveira J, Santos R, Dalgleish R, Oliveira JL. Variobox: Automatic Detection and Annotation of Human Genetic Variants. *Human mutation* 2013, **35**(2), 202-207.

Geremek M, Witt M. Primary ciliary dyskinesia: genes, candidate genes and chromosomal regions. *J Appl Genet* 2004, **45**:347–361.

Goodenough U, Heuser J. Substructure of Inner Dynein Arms, Radial Spokes, and the Central Pair/Projection Complex of Cilia and Flagella. *The journal of cell biology* 1985, **100**:2008-2018

Grudzinskas JG, Yovich JL. Sperm structure and function. In Grudzinskas JG, Yovich JL, editors. *Gametes - Spermatozoon* 1995, p.45-70.

Guichard C, Harricane M-C, Lafitte J-J, Godard P, Zaegel M, Tack V, Lalau G, Bouvagnet P. Axonemal Dynein Intermediate-Chain Gene (DNAI1) Mutations Result in Situs Inversus and Primary Ciliary Dyskinesia (Kartagener Syndrome). *Am J Hum Genet* 2001, **68**:1030–1035.

Gupta GS. Spermatogenesis. *Proteomics Spermatogenes* 2006, , p. 1–20. Available from: http://books.google.pt/books?id=U1ENgT2_FPsC&dq=Spermatogenesis.&hl=pt-PT&source=gbs_navlinks_s

Habedanck R, Stierhof Y-D, Wilkinson CJ, Nigg EA. The Polo kinase Plk4 functions in centriole duplication. *Nat Cell Biol* 2005, **7**:1140–1146.

Haidl G, Becker A, Henkel R. Poor development of outer dense fibres as a major cause of tail abnormalities in the spermatozoa of asthenoteratozoospermic men. *Hum Reprod* 1991, **6**:1431–1438.

Hamatani T. Human spermatozoal RNAs. *Fertility and sterility*, 2012, **97**: 275-281.

Hammond J, Cai D, Verhey KJ. Tubulin modifications and their cellular functions. *Curr Opin Cell Biol* 2008, **20**:71–76.

Hess RA, Franca LR de. Spermatogenesis and cycle of the seminiferous epithelium. In Cheng CY editor. *Mol Mech Spermatogenes Spermatogenes* 2008, p1-16.

Heuser T, Raytchev M, Krell J, Porter ME, Nicastro D. The dynein regulatory complex is the nexin link and a major regulatory node in cilia and flagella. *J Cell Biol* 2009, **187**:921–933.

High KH, Nathwani A, Spencer T, Lillicrap D. Current status of haemophilia gene therapy. *Haemophilia* 2014, **20**:43–49.

Hjeij R, Lindstrand A, Francis R, Zariwala MA, Liu X, Li Y, Damerla R, Dougherty GW, Abouhamed M, Olbrich H. ARMC4 Mutations Cause Primary Ciliary Dyskinesia with Randomization of Left/Right Body Asymmetry. *Am J Hum Genet* 2013, **93**:357–367.

Hock MB, Kralli A. Transcriptional control of mitochondrial biogenesis and function. *Annu Rev Physiol* 2009, **71**:177–203. Annual Reviews.

Holdcraft RW, Braun RE. Hormonal regulation of spermatogenesis. *Int J Androl* 2004, **27**:335–342.

Holstein AF, Roosen-Runge E. Spermatozoa. *Atlas Hum Spermatogenes* 1981, Grosse Verlag Berlin.

Horani A, Druley TE, Zariwala MA, Patel AC, Levinson BT, Arendonk LG Van, Thornton KC, Giacalone JC, Albee AJ, Wilson KS. Whole-Exome Capture and Sequencing Identifies HEATR2 Mutation as a Cause of Primary Ciliary Dyskinesia. *Am J Hum Genet* 2012, **91**:685–693.

Horani A, Ferkol TW, Shoseyov D, Wasserman MG, Oren YS, Kerem B, Amirav I, Cohen-Cymberknoh M, Dutcher SK, Brody SL. LRRC6 mutation causes primary ciliary dyskinesia with dynein arm defects. *PLoS One* 2013, **8**:e59436.

Hornef N, Olbrich H, Horvath J, Zariwala M a, Fliegauf M, Loges NT, Wildhaber J, Noone PG, Kennedy M, Antonarakis SE, *et al.* DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arm defects. *Am J Respir Crit Care Med* 2006, **174**:120–126.

Howes L, Jones R. Interactions between zona pellucida glycoproteins and sperm proacrosin/acrosin during fertilization. *J Reprod Immunol* 2002, **53**:181–192.

Hoyer-Fender S. Centriole maturation and transformation to basal body. *Semin Cell Dev Biol* 2010, **21**:142–147.

Ihara M, Kinoshita A, Yamada S, Tanaka H, Tanigaki A, Kitano A, Goto M, Okubo K, Nishiyama H, Ogawa O. Cortical organization by the septin cytoskeleton is essential for structural and mechanical integrity of mammalian spermatozoa. *Dev Cell* 2005, **8**:343–352.

Ikpa PT, Bijvelds MJC, Jonge HR de. Cystic fibrosis: Toward personalized therapies. *Int J Biochem Cell Biol* 2014, **52**:192–200.

Inaba K. Molecular Architecture of the Sperm Flagella : Molecules for Motility and Molecular Architecture of the Sperm Flagella : Molecules for Motility and Signaling. *Zoolog Sci* 2003, **20**:1043–1056.

Inaba K. Molecular basis of sperm flagellar axonemes. *Ann N Y Acad Sci* 2007, **1101**:506–526.

Inaba K. Sperm flagella : comparative and phylogenetic perspectives of protein components Unicellular algae Chlamydomonas : *Mol Hum Reprod* 2011, **17**:524–538.

Kabututu ZP, Thayer M, Melehani JH, Hill KL. CMF70 is a subunit of the dynein regulatory complex. *J Cell Sci* 2010, **123** :3587–3595.

Kang D, Kim SH, Hamasaki N. Mitochondrial transcription factor A (TFAM): roles in maintenance of mtDNA and cellular functions. *Mitochondrion* 2007, **7**:39–44.

Kaufman BA, Durisic N, Mativetsky JM, Costantino S, Hancock MA, Grutter P, Shoubridge EA. The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. *Mol Biol Cell* 2007, **18**:3225–3236.

Kim YH, McFarlane JR, Almahbobi G, Stanton PG, Temple-Smith PD, De Kretser DM. Isolation and partial characterization of rat sperm tail fibrous sheath proteins and comparison with rabbit and human spermatozoa using a polyclonal antiserum. *J Reprod Fertil* 1995, **104**:107–114.

Kim YH, De Kretser DM, Temple-Smith PD, Hearn MT, McFarlane JR. Isolation and characterization of human and rabbit sperm tail fibrous sheath. *Molecular human reproduction* 1997, **3**:307-313.

King SM. The dynein microtubule motor. *Biochim Biophys Acta - Mol Cell Res* 2000, **1496**:60–75.

Kissel H, Georgescu M-M, Larisch S, Manova K, Hunnicutt GR, Steller H. The Sept4 Septin Locus Is Required for Sperm Terminal Differentiation in Mice. *Dev Cell* 2005, **8**:353–364.

Kitagawa D, Busso C, Flückiger I, Gönczy P. Phosphorylation of SAS-6 by ZYG-1 Is Critical for Centriole Formation in C. elegans Embryos. *Dev Cell* 2009, **17**:900–907.

Kitamura K, Miyagawa Y, Iguchi N, Nishimura H, Tanaka H, Nishimune Y. Molecular cloning and characterization of the human orthologue of the oppo 1 gene encoding a sperm tail protein. *Mol Hum Reprod* 2003, **9**:237–243.

Knowles MR, Leigh MW, Carson JL, Davis SD, Dell SD, Ferkol TW, Olivier KN, Sagel SD, Rosenfeld M, Burns KA. Mutations of DNAH11 in patients with primary ciliary dyskinesia with normal ciliary ultrastructure. *Thorax* 2012, **67**:433–441.

Knowles MR, Ostrowski LE, Loges NT, Hurd T, Leigh MW, Huang L, Wolf WE, Carson JL, Hazucha MJ, Yin W. Mutations in SPAG1Cause Primary Ciliary Dyskinesia Associated with Defective Outer and Inner Dynein Arms. *Am J Hum Genet* 2013, **93**:711–720.

Kott E, Legendre M, Copin B, Papon J-F, Dastot-Le Moal F, Montantin G, Duquesnoy P, Piterboth W, Amram D, Bassinet L. Loss-of-Function Mutations in RSPH1 Cause Primary Ciliary Dyskinesia with Central-Complex and Radial-Spoke Defects. *Am J Hum Genet* 2013, **93**:561–570.

Krisfalusi M, Miki K, Magyar PL, O'Brien DA. Multiple Glycolytic Enzymes Are Tightly Bound to the Fibrous Sheath of Mouse Spermatozoa. *Biol Reprod* 2006, **75** :270–278.

Ku C-S, Cooper DN, Polychronakos C, Naidoo N, Wu M, Soong R. Exome sequencing: dual role as a discovery and diagnostic tool. *Ann Neurol* 2012a, **71**:5–14.

Ku C-S, Wu M, Cooper DN, Naidoo N, Pawitan Y, Pang B, Iacopetta B, Soong R. Exome versus transcriptome sequencing in identifying coding region variants. *Expert Rev Mol Diagn* 2012b, **12**:241–251. Expert Reviews.

Kunimoto K, Yamazaki Y, Nishida T, Shinohara K, Ishikawa H, Hasegawa T, Okanoue T, Hamada H, Noda T, Tamura A, *et al.* Coordinated Ciliary Beating Requires Odf2-Mediated Polarization of Basal Bodies via Basal Feet. *Cell* 2012, **148**:189–200.

Kuo P-L, Chiang H-S, Wang Y-Y, Kuo Y-C, Chen M-F, Yu I, Teng Y-N, Lin S-W, Lin Y-H. SEPT12-Microtubule Complexes Are Required for Sperm Head and Tail Formation. *Int J Mol Sci* 2013, **14**:22102–22116.

Kwitny S, Klaus A V, Hunnicutt GR. The annulus of the mouse sperm tail is required to establish a membrane diffusion barrier that is engaged during the late steps of spermiogenesis. *Biol Reprod* 2010, **82**:669–678.

Langeberg LK, Scott JD. A-kinase-anchoring proteins. *J Cell Sci* 2005, **118**:3217–3220.

Lhuillier P, Rode B, Escalier D, Lorès P, Dirami T, Bienvenu T, Gacon G, Dulioust E, Touré a. Absence of annulus in human asthenozoospermia: case report. *Hum Reprod* 2009, **24**:1296–1303.

Li X, Manley JL. Inactivation of the SR Protein Splicing Factor ASF / SF2 Results in Genomic Instability. *Cell* 2005, **122**:365–378.

Lie H, Ferkol T. Primary Ciliary Dyskinesia. *Drugs* 2007, **67**:1883–1892.

Lin Y-H, Lin Y-M, Wang Y-Y, Yu I, Lin Y-W, Wang Y-H, Wu C-M, Pan H-A, Chao S-C, Yen PH. The expression level of septin12 is critical for spermiogenesis. *Am J Pathol* 2009, **174**:1857–1868.

Lin Y-H, Kuo Y-C, Chiang H-S, Kuo P-L. The role of the septin family in spermiogenesis. *Spermatogenesis* 2011, **1**:298–302.

Lin Y-H, Wang Y-Y, Chen H-I, Kuo Y-C, Chiou Y-W, Lin H-H, Wu C-M, Hsu C-C, Chiang H-S, Kuo P-L. SEPTIN12 Genetic Variants Confer Susceptibility to Teratozoospermia. *PLoS One* 2012, **7**:e34011.

Loges NT, Olbrich H, Fenske L, Mussaffi H, Horvath J, Fliegauf M, Kuhl H, Baktai G, Peterffy E, Chodhari R, *et al.* DNAI2 mutations cause primary ciliary dyskinesia with defects in the outer dynein arm. *Am J Hum Genet* 2008, **83**:547–558.

Loges NT, Olbrich H, Becker-Heck A, Häffner K, Heer A, Reinhard C, Schmidts M, Kispert A, Zariwala M a, Leigh MW, *et al.* Deletions and point mutations of LRRC50 cause primary ciliary dyskinesia due to dynein arm defects. *Am J Hum Genet* 2009, **85**:883–889.

Lok S, Johnston DS, Conklin D, Lofton-Day CE, Adams RL, Jelmberg AC, Whitmore TE, Schrader S, Griswold MD, Jaspers SR. Identification of INSL6, a new member of the insulin family that is expressed in the testis of the human and rat. *Biol Reprod* 2000, **62**:1593–1599.

Luconi M, Forti G, Baldi E. Pathophysiology of sperm motility. *Front Biosci* 2006, **11**:1433–1447.

Magner, LN. A history of the life sciences, revised and expanded. CRC Press, 2002. Available from: http://books.google.pt/books?id=EnG0gshuCBAC .

Mandal A, Naaby-Hansen S, Wolkowicz MJ, Klotz K, Shetty J, Retief JD, Coonrod SA, Kinter M, Sherman N, Cesar F, *et al.* FSP95, A Testis-Specific 95-Kilodalton Fibrous Sheath Antigen That Undergoes Tyrosine Phosphorylation in Capacitated Human Spermatozoa. *Biol Reprod* 1999, **61**:1184–1197.

Mao H-T, Yang W-X. Modes of acrosin functioning during fertilization. *Gene* 2013, **526**:75–79.

Mazor M, Alkrinawi S, Chalifa-Caspi V, Manor E, Sheffield VC, Aviram M, Parvari R. Primary Ciliary Dyskinesia Caused by Homozygous Mutation in DNAL1</i>, Encoding Dynein Light Chain 1. *Am J Hum Genet* 2011, **88**:599–607.

Meistrich M, Mohapatra B, Shirley C, Zhao M. Roles of transition nuclear proteins in spermiogenesis. *Chromosoma* 2003, **111**:483–488.

Merveille A-C, Davis EE, Becker-Heck A, Legendre M, Amirav I, Bataille G, Belmont J, Beydon N, Billen F, Clément A, *et al.* CCDC39 is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs. *Nat Genet* 2011, **43**:72–78.

Metzker ML. Sequencing technologies—the next generation. *Nat Rev Genet* 2010, **11**:31–46.

Miki K, Willis WD, Brown PR, Goulding EH, Fulcher KD, Eddy EM. Targeted Disruption of the Akap4Gene Causes Defects in Sperm Flagellum and Motility. *Dev Biol* 2002, **248**:331–342.

Miki K, Qu W, Goulding EH, Willis WD, Bunch DO, Strader LF, Perreault SD, Eddy EM, O'Brien DA. Glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme, is required for sperm motility and male fertility. *Proc Natl Acad Sci USA* 2004, **101**:16501–16506. National Acad Sciences.

Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988, **16**:1215.

Miller D, Ostermeier GC, Krawetz SA. The controversy, potential and roles of spermatozoal RNA. *Trends Mol Med* 2005, **11**:156–163.

Miller D, Ostermeier GC. Spermatozoal RNA: why is it there and what does it do? *Gynécologie Obs Fertil* 2006, **34**:840–846.

Mitchison HM, Schmidts M, Loges NT, Freshour J, Dritsoula A, Hirst RA, O'Callaghan C, Blau H, Dabbagh M Al, Olbrich H. Mutations in axonemal dynein assembly factor DNAAF3 cause primary ciliary dyskinesia. *Nat Genet* 2012, **44**:381–389.

Miyamoto T, Tsujimura A, Miyagawa Y, Koh E, Namiki M, Horikawa M, Saijo Y, Sengoku K. Single nucleotide polymorphisms in the SEPTIN12 gene may be associated with azoospermia by meiotic arrest in Japanese men. *J Assist Reprod Genet* 2012, **29**:47–51.

Moore DJ, Onoufriadis A, Shoemark A, Simpson MA, Lage PI Zur, Castro SC de, Bartoloni L, Gallone G, Petridi S, Woollard WJ. Mutations in ZMYND10</i>, a Gene Essential for Proper Axonemal Assembly of Inner and Outer Dynein Arms in Humans and Flies, Cause Primary Ciliary Dyskinesia. *Am J Hum Genet* 2013, **93**:346–356.

Moreno RD, Ramalho-Santos J, Sutovsky P, Chan EKL, Schatten G. Vesicular Traffic and Golgi Apparatus Dynamics During Mammalian Spermatogenesis : Implications for Acrosome Architecture. *Biol Reprod* 2000, **63**:89–98.

Moretti E, Scapigliati G, Pascarelli NA, Baccetti B, Collodel G. Localization of AKAP4 and tubulin proteins in sperm with reduced motility. *Asian J Androl* 2007, **9**:641–649.

Moretti E, Geminiani M, Terzuoli G, Renieri T, Pascarelli N, Collodel G. Two cases of sperm immotility: a mosaic of flagellar alterations related to dysplasia of the fibrous sheath and abnormalities of head-neck attachment. *Fertil Steril* 2011, **95**:e19–23.

Mostowy S, Cossart P. Septins: the fourth component of the cytoskeleton. *Nat Rev Mol Cell Biol* 2012, **13**:183–194.

Mudrak O, Tomilin N, Zalensky A. Chromosome architecture in the decondensing human sperm nucleus. *J Cell Sci* 2005, **118** :4541–4550.

Mújica A, Navarro-García F, Hernández-González EO, Lourdes Juárez-Mosqueda M De. Perinuclear theca during spermatozoa maturation leading to fertilization. *Microsc Res Tech* 2003, **61**:76–87.

Naaby-Hansen S, Mandal A, Wolkowicz MJ, Sen B, Westbrook VA, Shetty J, Coonrod SA, Klotz KL, Kim Y-H, Bush LA, *et al*. CABYR, a novel calcium-binding tyrosine phosphorylation-regulated fibrous sheath protein involved in capacitation. *Developmental biology* 2002, **242**: 236-254.

Nakamura S, Terada Y, Horiuchi T, Emuta C, Murakami T, Yaegashi N, Okamura K. Human Sperm Aster Formation and Pronuclear Decondensation in Bovine Eggs Following Intracytoplasmic Sperm Injection Using a Piezo-Driven Pipette: A Novel Assay for Human Sperm Centrosomal Function. *Biol Reprod* 2001, **65** :1359–1363.

Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003, **31**:3812–3814.

Oakley BR. An abundance of tubulins. *Trends Cell Biol* 2000, **10**:537–542.

Oko R. Comparative analysis of proteins from the fibrous sheath and outer dense fibers of rat spermatozoa. *Biol Reprod* 1988, **39**:169–182.

Oko, R. Occurrence and formation of cytoskeletal proteins in mammalian spermatozoa. *Andrologia*, 1998, **30**: 193-206.

Oko R, Sutovsky P. Biogenesis of sperm perinuclear theca and its role in sperm functional competence and fertilization. *J Reprod Immunol* 2009, **83**:2–7.

Olbrich H, Häffner K, Kispert A, Völkel A, Volz A, Sasmaz G, Reinhardt R, Hennig S, Lehrach H, Konietzko N, *et al.* Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. *Nat Genet* 2002, **30**:143–144.

Olbrich H, Schmidts M, Werner C, Onoufriadis A, Loges NT, Raidt J, Banki NF, Shoemark A, Burgoyne T, Turki S Al. Recessive HYDINMutations Cause Primary Ciliary Dyskinesia without Randomization of Left-Right Body Asymmetry. *Am J Hum Genet* 2012, **91**:672–684.

Oliva R, Dixon GH. Vertebrate protamine genes and the histone-to-protamine replacement reaction. *Prog Nucleic Acid Res Mol Biol* 1991, **40**:25–94.

Oliva R. Protamines and male infertility. *Hum Reprod Update* 2006, **12**:417–435.

Olson GE, Sammons DW. Structural chemistry of outer dense fibers of rat sperm. *Biol Reprod* 1980, **22**:319–332.

Omoto CK, Gibbons IR, Kamiya R, Shingyoji C, Takahashi K, Witman GB. Rotation of the central pair microtubules in eukaryotic flagella. *Mol Biol Cell* 1999, **10**:1–4.

Omran H, Kobayashi D, Olbrich H, Tsukahara T, Loges NT, Hagiwara H, Zhang Q, Leblond G, O'Toole E, Hara C. Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins. *Nature* 2008, **456**:611–616.

Onoufriadis A, Paff T, Antony D, Shoemark A, Micha D, Kuyt B, Schmidts M, Petridi S, Dankert-Roelse JE, Haarman EG, *et al.* Splice-site mutations in the axonemal outer dynein arm docking complex gene CCDC114 cause primary ciliary dyskinesia. *Am J Hum Genet* 2013, **92**:88–98.

Onoufriadis A, Shoemark A, Schmidts M, Patel M, Jimenez G, Liu H, Thomas B, Dixon M, Hirst RA, Rutman A. Targeted NGS gene panel identifies mutations in RSPH1 causing primary ciliary dyskinesia and a common mechanism for ciliary central-pair agenesis due to radial spoke defects. *Hum Mol Genet* 2014, ddu046.

Ortega C, Verheyen G, Raick D, Camus M, Devroey P, Tournaye H. Absolute asthenozoospermia and ICSI: what are the options? *Hum Reprod Update* 2011, **17**:684–692.

Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong S-E, Walford GA, Sugiana C, Boneh A, Chen WK. A mitochondrial protein compendium elucidates complex I disease biology. *Cell* 2008, **134**:112–123.

Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput Biol* 2013, **9**:e1003153.

Panizzi JR, Becker-heck A, Castleman VH, Al-mutairi D, Liu Y, Loges NT, Pathak N, Austin-tse C, Sheridan E, Schmidts M, *et al.* CCDC103 mutations cause primary ciliary dyskinesia by disrupting assembly of ciliary dynein arms. *Nat Genet* 2012, **44**:714–719.

Parla JS, Iossifov I, Grabill I, Spector MS, Kramer M, McCombie WR. A comparative analysis of exome capture. *Genome Biol* 2011, **12**:R97.

Pazour GJ, Wilkerson CG, Witman GB. A dynein light chain is essential for the retrograde particle movement of intraflagellar transport (IFT). *J Cell Biol* 1998, **141**:979–992.

Pazour GJ, Agrin N, Walker BL, Witman GB. Identification of predicted human outer dynein arm genes: candidates for primary ciliary dyskinesia genes. *J Med Genet* 2006, **43**:62–73.

Pedersen H. Further observations on the fine structure of the human spermatozoon. *Zeitschrift für Zellforsch und Mikroskopische Anat* 1972, **123**:305–315.

Pennarun G, Escudier E, Chapelin C, Bridoux a M, Cacheux V, Roger G, Clément A, Goossens M, Amselem S, Duriez B. Loss-of-function mutations in a human gene related to Chlamydomonas reinhardtii dynein IC78 result in primary ciliary dyskinesia. *Am J Hum Genet* 1999, **65**:1508–1519.

Pennisi E. Encode Project Writes Eulogy For Junk DNA. *Science (80-)* 2012, **337**:1159–1160.

Perry III WL. Short Technical Report JavaScript DNA Translator: DNA-Aligned Protein Translations. *Biotechniques* 2002, **33**:1318–1320.

Pereira R, Oliveira J, Sousa M (2014) A molecular approach to sperm immotility in Humans: A review. Medicina Reproductiva y Embriología Clínica, edited by Elsevier Spain (*in press*).

Petersen C, Füzesi L, Hoyer-Fender S. Outer dense fibre proteins from human sperm tail: molecular cloning and expression analyses of two cDNA transcripts encoding proteins of approximately 70 kDa. *Mol Hum Reprod* 1999, **5**:627–635.

Piomboni P, Focarelli R, Stendardi A, Ferramosca A, Zara V. The role of mitochondria in energy production for human sperm motility. *Int J Androl* 2012, **35**:109–124.

Piperno G, Mead K, Shestak W. The inner dynein arms I2 interact with a“ dynein regulatory complex” in Chlamydomonas flagella. *J Cell Biol* 1992, **118**:1455–1463.

Piperno G, Mead K, LeDizet M, Moscatelli A. Mutations in the“ dynein regulatory complex” alter the ATP-insensitive binding sites for inner arm dyneins in Chlamydomonas axonemes. *J Cell Biol* 1994, **125**:1109–1117.

Porter ME, Sale WS. The 9 + 2 axoneme anchors multiple inner arm dyneins and a network of kinases and phosphatases that control motility. *J Cell Biol* 2000, **151**:F37–42.

Porter ME. Axonemal dyneins: assembly, organization, and regulation. *Curr Opin Cell Biol* 1996, **8**:10–17.

Rajender S, Rahul P, Mahdi AA. Mitochondria, spermatogenesis and male infertility. *Mitochondrion* 2010, **10**:419–428.

Rengan AK, Agarwal A, Linde M van der, Plessis SS du. An investigation of excess residual cytoplasm in human spermatozoa and its distinction from the cytoplasmic droplet. *Reprod Biol Endocrinol* 2012, **10**:1–8.

Roberts AJ, Kon T, Knight PJ, Sutoh K, Burgess SA. Functions and mechanics of dynein motor proteins. *Nat Rev Mol Cell Biol* 2013, **14**:713-726

Rosenbaum J. Cytoskeleton: Functions for tubulin modifications at last. *Curr Biol* 2000, **10**:R801–R803.

Rupp G, Porter ME. A subunit of the dynein regulatory complex in Chlamydomonas is a homologue of a growth arrest–specific gene product. *J Cell Biol* 2003, **162**:47–57.

Salmon NA, Pera RAR, Xu EY. A Gene Trap Knockout of the Abundant Sperm Tail Protein, Outer Dense Fiber 2, Results in Preimplantation Lethality. *Genesis* 2006, **44**:515–522.

Sathananthan a H, Ratnam SS, Ng SC, Tarín JJ, Gianaroli L, Trounson a. The sperm centriole: its inheritance, replication and perpetuation in early human embryos. *Hum Reprod* 1996, **11**:345–356.

Scarpulla RC. Metabolic control of mitochondrial biogenesis through the PGC-1 family regulatory network. *Biochim Biophys Acta (BBA)-Molecular Cell Res* 2011, **1813**:1269–1278.

Schwabe GC, Hoffmann K, Loges NT, Birker D, Rossier C, Santi MM De, Olbrich H, Fliegauf M, Failly M, Liebers U. Primary ciliary dyskinesia associated with normal axoneme ultrastructure is caused by DNAH11 mutations. *Hum Mutat* 2008, **29**:289–298.

Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010, **7**:575–576.

Schweizer S, Hoyer-Fender S. Mouse Odf2 localizes to centrosomes and basal bodies in adult tissues and to the photoreceptor primary cilium. *Cell Tissue Res* 2009, **338**:295–301.

Shao X, Tarnasky HA, Schalles U, Oko R, Hoorn FA van der. Interactional Cloning of the 84-kDa Major Outer Dense Fiber Protein Odf84 leucine zippers mediate associations of Odf84 AND Odf27. *J Biol Chem* 1997, **272**:6105–6113.

Shao X, Tarnasky HA, Lee JP, Oko R, Hoorn FA Van Der. Spag4 , a Novel Sperm Protein , Binds Outer Dense- Fiber Protein Odf1 and Localizes to Microtubules of Manchette and Axoneme. *Develomental Biol* 1999a, **211**:109–123.

Shao X, Murthy S, Demetrick DJ, Hoorn FA Van der. Human outer dense fiber gene, ODF2, localizes to chromosome 9q34. *Cytogenet Genome Res* 1999b, **83**:221–223.

Shao X, Xue J, Hoorn FA van der. Testicular protein Spag5 has similarity to mitotic spindle protein Deepest and binds outer dense fiber protein Odf1. *Mol Reprod Dev* 2001, **59**:410–416.

Shepard PJ, Hertel KJ. The SR protein family. *Genome Biol* 2009, **10**:242.

Sickmann A, Reinders J, Wagner Y, Joppich C, Zahedi R, Meyer HE, Schönfisch B, Perschil I, Chacinska A, Guiard B. The proteome of Saccharomyces cerevisiae mitochondria. *Proc Natl Acad Sci* 2003, **100**:13207–13212.

Silvanovich A, Li M, Serr M, Mische S, Hays TS. The third P-loop domain in cytoplasmic dynein heavy chain is essential for dynein motor function and ATP-sensitive microtubule binding. *Mol Biol Cell* 2003, **14**:1355–1365.

Smith EF, Lefebvre PA. The role of central apparatus components in flagellar motility and microtubule assembly. *Cell Motil Cytoskeleton* 1997, **38**:1–8.

Smith EF, Yang P. The radial spokes and central apparatus: Mechano-chemical transducers that regulate flagellar motility. *Cell Motil Cytoskeleton* 2004, **57**:8–17.

Soung NK, Park JE, Yu LR, Lee KH, Lee JM, Bang JK, Veenstra TD, Rhee K, Lee KS. Plk1-dependent and -independent roles of an ODF2 splice variant, hCenexin1, at the centrosome of somatic cells. *Develomental Cell* 2009, **16**:539–550.

Sousa M, Oliveira E, Alves A, Gouveia M, Figueiredo H, Ferraz L, Barros A, Sá R (2014) Ultrastructural analysis of five patients with total sperm immotility. Zygote: in press.

Spiropoulos J, Turnbull DM, Chinnery PF. Can mitochondrial DNA mutations cause sperm dysfunction? *Mol Hum Reprod* 2002, **8**:719–721.

Staff DG. *The Facts on File Biology Handbook* 2009, Available from: http://books.google.pt/books?id=9pkNvJuvxpwC .

Steels JD, Estey MP, Froese CD, Reynaud D, Pace-Asciak C, Trimble WS. Sept12 is a component of the mammalian sperm tail annulus. *Cell Motil Cytoskeleton* 2007, **64**:794–807.

Steger K, Fink L, Failing K, Bohle RM, Kliesch S, Weidner W, Bergmann M. Decreased protamine-1 transcript levels in testes from infertile men. *Mol Hum Reprod* 2003, **9**:331–336.

Sugino Y, Ichioka K, Soda T, Ihara M, Kinoshita M, Ogawa O, Nishiyama H. Septins as Diagnostic Markers for a Subset of Human Asthenozoospermia. *J Urol* 2008, **180**:2706–2709.

Sumner CJ. Therapeutics development for spinal muscular atrophy. *NeuroRx* 2006, **3**:235–245.

Sutovsky P, Manandhar G, Wu A, Oko R. Interactions of sperm perinuclear theca with the oocyte: Implications for oocyte activation, anti-polyspermy defense, and assisted reproduction. *Microsc Res Tech* 2003, **61**:362–378.

Sutovsky P, Manandhar G. Mammalian spermatogenesis and sperm structure: anatomical and compartmental analysis. In *Sperm Cell-Production, Matur Fertil Regen* 2006, p. 1–31. Cambridge University Press,.

Takada S, Wilkerson CG, Wakabayashi K, Kamiya R, Witman GB. The Outer Dynein Arm-Docking Complex : Composition and Characterization of a Subunit (Oda1) Necessary for Outer Arm Assembly. *Mol Biol Cell* 2002, **13**:1015–1029.

Tamowski S, Aston KI, Carrell DT. The use of transgenic mouse models in the study of male infertility. *Syst Biol Reprod Med* 2010, **56**:260–273.

Tarkar A, Loges NT, Slagle CE, Francis R, Dougherty GW, Tamayo J V, Shook B, Cantino M, Schwartz D, Jahnke C. DYX1C1 is required for axonemal dynein assembly and ciliary motility. *Nat Genet* 2013, **45**:995–1003.

Tarnasky H, Cheng M, Ou Y, Thundathil J, Oko R, Hoorn F van der. Gene trap mutation of murine outer dense fiber protein-2 gene can result in sperm tail abnormalities in mice with high percentage chimaerism. *BMC Dev Biol* 2010, **10**:67.

Teves ME, Zhang Z, Costanzo RM, Henderson SC, Corwin FD, Zweit J, Sundaresan G, Subler M, Salloum FN, Rubin BK. Sperm-Associated Antigen–17 Gene Is Essential for Motile Cilia Function and Neonatal Survival. *Am J Respir Cell Mol Biol* 2013, **48**:765–772.

Toshimori K. Maturation of mammalian spermatozoa: modifications of the acrosome and plasma membrane leading to fertilization. *Cell Tissue Res* 1998, **293**:177–187.

Toshimori K, Ito C. Formation and organization mammalian sperm head. *Arch Histol Cytol* 2003, **55**:383–396.

Touré A, Lhuillier P, Gossen JA, Kuil CW, Lhôte D, Jégou B, Escalier D, Gacon G. The testis anion transporter 1 (Slc26a8) is required for sperm terminal differentiation and male fertility in the mouse. *Hum Mol Genet* 2007, **16**:1783–1793.

Tremellen K.Oxidative stress and male infertility—a clinical perspective. *Human reproduction update* 2008, **14**: 243-258.

Turner RMO, Johnson LR, Haig-Ladewig L, Gerton GL, Moss SB. An X-linked Gene Encodes a Major Human Sperm Fibrous Sheath Protein, hakap82: genomic organization, protein kinase a-rii binding, and distribution of the precursor in the sperm tail . *J Biol Chem* 1998, **273** :32135–32141.

Turner RMO, Musse MP, Mandal A, Klotz KEN, Friederike C, Jayes L, Herr JC, Gerton GL, Moss SB, Chemes HE. Molecular Genetic Analysis of Two Human Sperm Fibrous. *J Androl* 2001, **22**:302–315.

Turner RM. Tales from the tail: what do we really know about sperm motility? *J Androl* 2003, **24**:790–803.

Vijayaraghavan S, Liberty GA, Mohan J, Winfrey VP, Olson GE, Carr DW. Isolation and Molecular Characterization of AKAP110, a Novel, Sperm-Specific Protein Kinase A-Anchoring Protein. *Mol Endocrinol* 1999, **13**:705–717.

Ward W, Coffey D. DNA Packaging and Organization in Mammalian Spermatozoa: Comparison with Somatic Cells. *Biol Reprod* 1991, **44**:569–574.

Wenz T. Regulation of mitochondrial biogenesis and PGC-1α under cellular stress. *Mitochondrion* 2013, **13**:134–142.

WHO. *WHO laboratory manual for the examination and processing of human semen*. 2010, Geneva, Switzerland: WHO.

Wirschell M, Olbrich H, Werner C, Tritschler D, Bower R, Sale WS, Loges NT, Pennekamp P, Lindberg S, Stenram U. The nexin-dynein regulatory complex subunit DRC1 is essential for motile cilia function in algae and humans. *Nat Genet* 2013, **45**:262–268.

Wykes SM, Krawetz S a. The structural organization of sperm chromatin. *J Biol Chem* 2003, **278**:29471–29477.

Yagi T. Bioinformatic approaches to dynein heavy chain classification. *Methods Cell Biol* 2009, **92**:1–9.

Yamagata K, Murayama K, Okabe M, Toshimori K, Nakanishi T, Kashiwabara S, Baba T. Acrosin Accelerates the Dispersal of Sperm Acrosomal Proteins during Acrosome Reaction. *J Biol Chem* 1998, **273** :10470–10474.

Yang P, Diener DR, Yang C, Kohno T, Pazour GJ, Dienes JM, Agrin NS, King SM, Sale WS, Kamiya R, *et al.* Radial spoke proteins of Chlamydomonas flagella. *J Cell Sci* 2006, **119**:1165–1174.

Yatsenko AN, Iwamori N, Iwamori T, Matzuk MM. The power of mouse genetics to study spermatogenesis. *J Androl* 2010, **31**:34–44.

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 2012, **13**:134.

Yeh S-D, Chen Y-J, Chang ACY, Ray R, She B-R, Lee W-S, Chiang H-S, Cohen SN, Lin-Chao S. Isolation and properties of Gas8, a growth arrest-specific gene regulated during male gametogenesis to produce a protein associated with the sperm motility apparatus. *J Biol Chem* 2002, **277**:6311–6317.

Zahn A, Furlong LI, Biancotti JC, Ghiringhelli PD, Marín-Briggiler CI, Vazquez-Levin MH. Evaluation of the proacrosin/acrosin system and its mechanism of activation in human sperm extracts. *J Reprod Immunol* 2002, **54**:43–63.

Zariwala MA, Leigh MW, Ceppa F, Kennedy MP, Noone PG, Carson JL, Hazucha MJ, Lori A, Horvath J, Olbrich H, *et al.* Mutations of DNAI1 in primary ciliary dyskinesia - Evidence of founder effect in a common mutation. *Am J Respir Crit Care Med* 2006, **174**:858–866.

Zhao M, Shirley CR, Hayashi S, Marcon L, Mohapatra B, Suganuma R, Behringer RR, Boissonneault G, Yanagimachi R, Meistrich ML. Transition nuclear proteins are required for normal chromatin condensation and functional sperm development. *genesis* 2004, **38**:200–213.

Zhou Z, Fu X-D. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* 2013, **122**:191–207.

Zukas R, Chang AJ, Rice M, Springer AL. Structural analysis of flagellar axonemes from inner arm dynein knockdown strains of Trypanosoma brucei. *Biocell* 2012, **36**:133–142.

VI. ATTACHMENTS

ATTACHMENT 1

In this attachment are described the methodology applied by Sousa *et al.* (2014) to perform the transmission electronic analyses of the immotile sperm of the five patients selected for the present genetic study and its main results.

Methodology

1. *Selection of patients*

Five patients, from Hospital Centre of Vila Nova de Gaia (CHVNG) (n = 4) and Centre of Reproductive Genetics Prof. Alberto Barros (CGR) (n=1) were selected based on total absence of sperm motility by semen analysis and ultrastructural confirmation.

2. *Transmission electron microscopy*

Semen samples were processed for transmission electron microscopy (TEM) at the Dept. of Microscopy, Lab. of Cell Biology, Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto.

Samples were fixed with Karnovsky (2.5% glutaraldehyde, 4% paraphormaldehyde, 0.15 M sodium cacodylate buffer) (Sigma-Aldrish, St. Louis, USA; Merck, Darnstradt, Germany) for 30 min at room temperature and then for 2 h at 4ºC, washed in 0.15 M sodium cacodylate buffer, pH 7.3 (Merck), post-fixed with 2% osmium tetraoxide (Merck) in buffer for 2h at 4°C, washed in buffer (10 min), serially dehydrated in ethanol (Panreac, Barcelone, Spain), equilibrated with propylene oxide (Merck) and embedded in Epon (Sigma). Semithin and ultrathin sections were prepared with a diamond knife (Diatome, Hatfield, Switzerland) in a LKB ultramicrotome (Leika Microssystems, Weltzlar, Germany). Ultrathin sections collected on 200 mesh formvar carbon-coated copper grids (Taab, Berks, England), stained with 3 % aqueous uranyl acetate (20 min) (BDH, Poole, England) and Reynolds lead citrate (10 min) (Merck). Ultrathin sections were observed in a JEOL 100XII transmission electron microscope (JEOL, Tokyo, Japan) operated at 60 Kv.

Main Results

Four of the five patients of this study (patient 1-4) were previously characterized as having clinical features compatible with dysplasia of fibrous sheath (DFS).

The ***patient 1*** (Fig. 15), had a reduced midpiece (MP) with a few dispersed mitochondria that presented a light appearance. This patient lack its annulus (An), thus the MP was continuous with the principal piece (PP) with hypertrophy, hyperplasia and disruption of fibrous sheath (FS) and with FS invading the MP. The axoneme (Ax) contains the nine microtubules doublets and the dynein arms (DA) but lacks the central pair complex (CPC) and the radial spokes (RS), thus showing a 9+0 pattern.

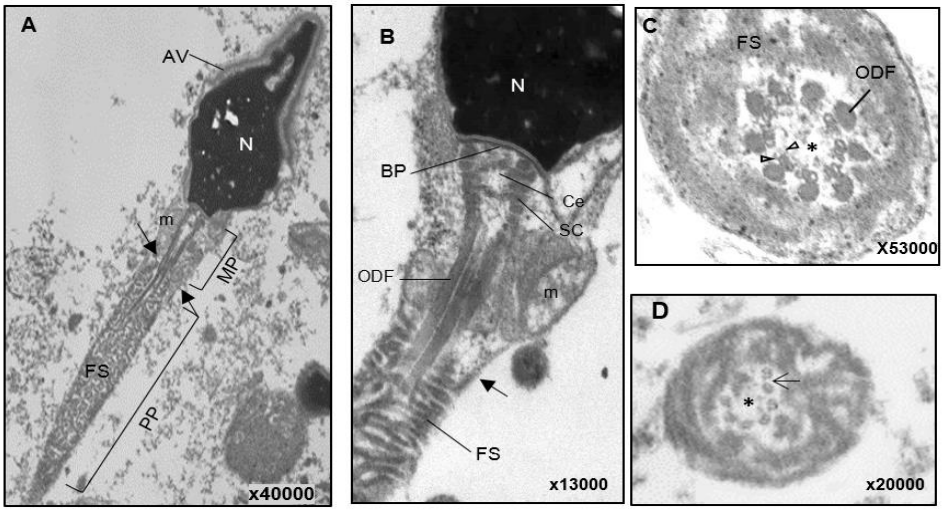


Figure 15. Ultrastructure of sperm from Patient 1. A. Longitudinal section of the sperm showing the head with a regular nucleus (N) and a displaced acrosomal vesicle (AV). At the midpiece (MP) is possible to see that mitochondria (m) are scarce and that the dysplastic fibrous sheath (FS) invades the MP region. No annulus material (arrows) is seen to distinguish the MP to the principal piece (PP); **B.** The basal plate (BP), centriole (Ce), striated columns (SC) and outer dense fibres (ODF) appear normal in the neck region, whereas mitochondria (m) display pale areas; **C.** The axoneme at the proximal PP, in which dynein arms (white arrowheads) are present but the central pair complex and the radial spokes (*) are absent. **D.** The axoneme at the distal PP, also without the central pair complex and radial spokes (*) and with microtubule doublet disorganization (arrow). Images provided by Professor Mário Sousa.

The ***patient 2*** (Fig. 16) show an enlarged MP, due to the presence of extra cytoplasm, and contained aggregated and misaligned mitochondria. The An material was present. At PP the FS material forms a unilateral longitudinal row and was seen a disorganized FS with a mild hypertrophy and hyperplasia. Any normal Ax were find, and most of them presenting a variable reduced number of microtubule doublets, RS and ODF, with presence of the CPC.

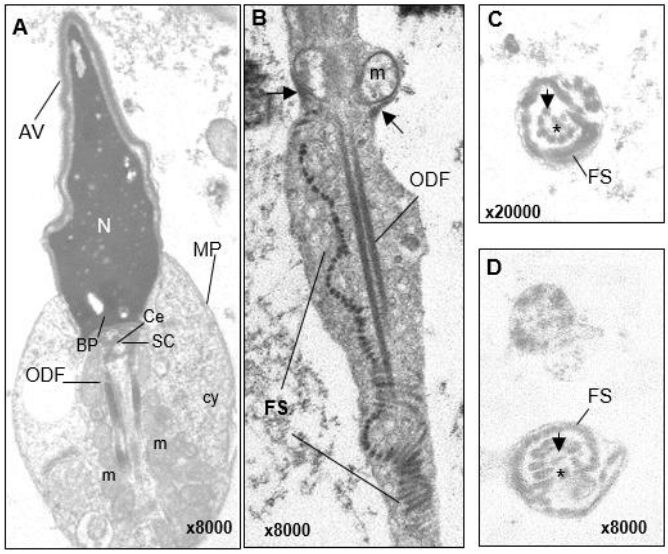


Figure 16. Ultrastructure of sperm from Patient 2 A. Longitudinal section of the sperm showing an irregular nucleus (N) and a displaced acrosomal vesicle (AV). The basal plate (BP), centriole (Ce), striated columns (SC) and outer dense fibers (ODF) show a typical structure. The apical region of the midpiece (MP) is enlarged with aggregates of pale mitochondria (m); **B** The annulus (arrows) is present and the apical region of the principal piece presents an unilateral row of FS followed by a dysplastic FS; **C, D** The axoneme is disorganized presenting a reduced number of microtubule doublets (arrow); the central pair complex is present (*). Images provided by Professor Mário Sousa.

In ***patient 3*** *(*Fig. 17*)*, the TEM analysis show that the MP was small, with a few pale and disorganized mitochondria and that no An material was present. Like in the patient 1, the PP was enlarged by the presence of hyperplastic and hypertrophic FS that invaded the MP. At the Ax level, absence of DA, nexin bridges and CPC, with presence of a variable number of RS were detected.

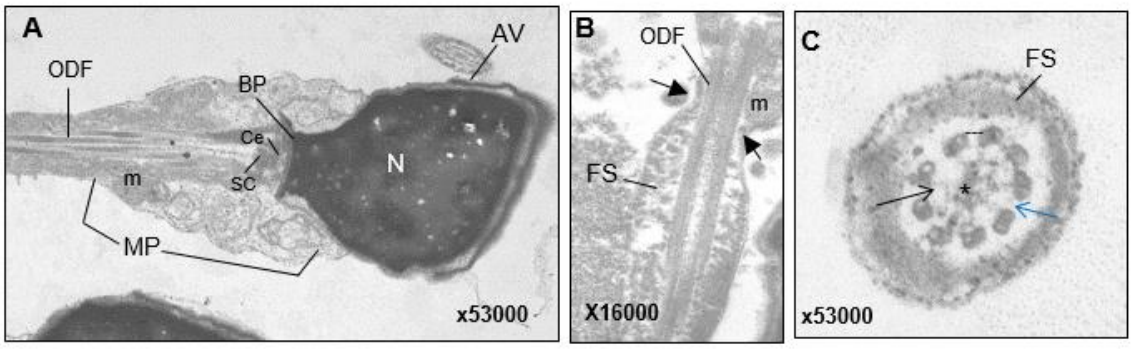


Figure 17. **Ultrastructure of sperm from Patient 3 A**. Longitudinal section of the sperm showing an irregular nucleus (N) and a displaced acrosomal vesicle (AV). At midpiece (MP) the basal plate (BP), centriole (Ce), striated columns (SC) and outer dense fibres (ODF) show a typical structure. **B.** There was no visible annulus material (arrow) and is observed a hyperplastic and hypertrophic FS that invaded the MP; **C**.The axoneme shows absence of the central pair complex (*), a reduced number of radial spokes (black arrow), absence of dynein arms (blue arrows) and nexin bridges (dotted line). Images provided by Professor Mário Sousa.

In ***Patient 4*** (Fig. 18), no An material was observed and the MP was invaded by *a* hyperplastic and hypertrophic *FS*. In the Ax, most of the sperm presented absence of the CPC and RS. Some sperm evidenced also total microtubule disorganization with an increased number of microtubule doublets and ODF.

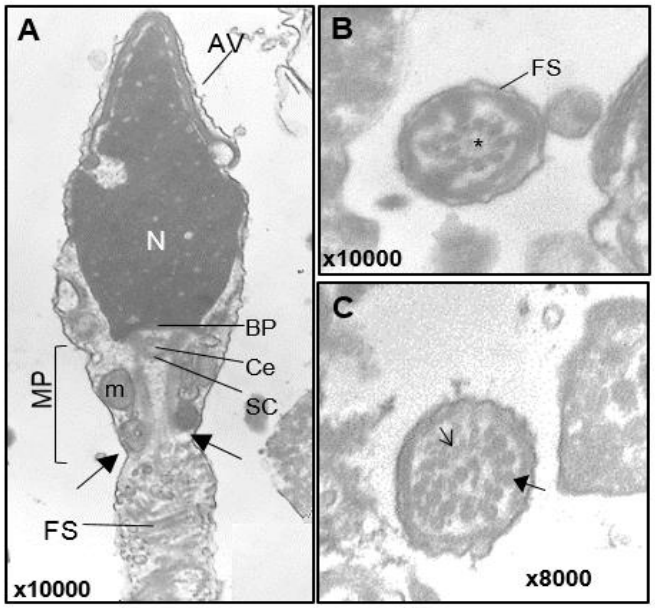


Figure 18. Ultrastructure of sperm from Patient 4 A. Longitudinal section of the sperm showing the head with a displaced acrosomal vesicle (AV). At midpiece (MP) the basal plate (BP), centriole (Ce), striated columns (SC) and outer dense fibres (ODF) show a typical structure. Also there was no visible annulus material (arrow) and is observed a hyperplastic and hypertrophic FS that invaded the MP; **B.** The axoneme showing absence of the central pair complex and radial spoke (*). **C.** A disorganized axoneme, with an increased number of microtubule doublets (small arrow) and ODF (large arrow). Images provided by Professor Mário Sousa.

In ***patient 5*** (Fig. 19), the MP was short and enlarged due to the presence of aggregated and misaligned pale mitochondria. At the Ax, all sections showed absence of DA and nexin links, with presence of RS and CPC. Besides, this patient reported chronic respiratory complains, with nasal polyps, chronic sinusitis, rhinitis and bronchitis. He also presented a *situs inversus totalis*. Therefore this patient present clinical feature compatible with Kartagener syndrome (KS), however, as the respiratory cells were not evaluated, cannot be affirmed that is a KS.

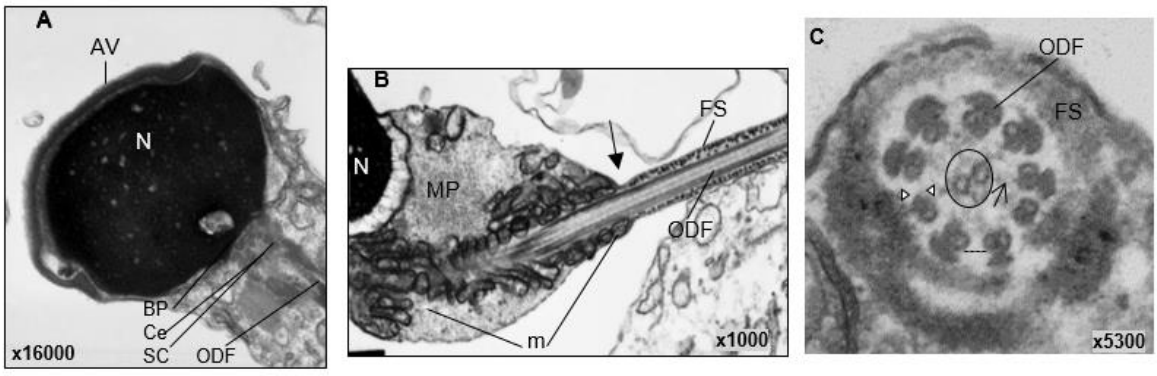


Figure 19 Ultrastructure of sperm from Patient 5. A Shows an abnormal sperm head, with displaced acrosomal vesicle (AV). At midpiece (MP) the basal plate (BP), centriole (Ce), striated columns (SC) and outer dense fibres (ODF) show a typical structure. **B**. The annulus is present (arrow), and the MP is short and enlarged due to the presence of aggregated and misaligned pale mitochondria (m). At the principal piece the ODF and the FS have a normal appearance. **C**. The axoneme shows a conserved central pair complex (circle) and radial spokes (arrow); dynein arms (white arrowheads) and nexin bridges (dotted line) are missing. Images provided by Professor Mário Sousa.

ATTACHMENT 2

This attachment contains the primers and amplification conditions that were used to sequence the following genes: *AKAP3, AKAP4, CCDC39, CCDC40, DNAI1, DNAH5* and *RSPH1*.

The standard primer design parameters were: primer length 18-25 bp; percentage of GC 40%-60%; amplicon length (AL) 400-650 bp; and Tm 59Cº-62Cº. Some primers have punctual modifications in the parameters, which are here referred.
PCR conditions that were applied to each gene regions:

1) **Stand-38c**, with an initial denaturation at 95°C for 2 min, followed by thirty-eight cycles of 95°C for 45 s, specific Ta (listed in tables below) for 30 s, and 72°C for 1 min, with a final extension at 72°C for 10 min;
2) **Stand-35c**, with an initial denaturation at 95°C for 10 min, followed by thirty-five cycles of 95°C for 1 min, specific Ta (listed in tables below) for 30 s, and 72°C for 2 min, with a final extension at 72°C for 10 min;
3) **High-GC**, a PCR with an initial denaturation at 95°C for 5 min, followed by forty cycles of 95°C for 45 s, 58°C for 30 s, and 72°C for 1 min, with a final extension at 72°C for 10 min.
4) **Multi-Dmsc**, a PCR with an initial denaturation at 95°C for 10 min, followed by thirty cycles of 95°C for 1 min, with a slower ramp reduction to 58/63°C for 30 s, and 72°C for 1 min, with a final extension at 72°C for 10 min;

The standard PCR the reaction mixture (30 µl) contained: 15µl of **PCR Master Mix** [(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 at 10 pmol/µl (Thermo Fisher Scientific, Einsteinstrasse, Germany) and 1 µl of DNA at 100ng/µ).

For the High-GC PCR condition, the reaction mixture (30 µl) contained: 12.5 µl of PCR Master Mix (Promega); 1.5 µl of sterile bidistilled water; 5 µl of Betaine 5M (Sigma-Aldrich, Missouri 63103, USA); 1,5 µl of DMSO (Bioline, London, United Kingdom), 1 µl of each primer at10 pmol/µl (Thermo Fisher Scientific) and 2,5 µl of DNA at 100ng/µl.

To amplify some exons, the PCR Master Mix was replaced by either: i) **EmeraldAmp** Max HS PCR Master Mix (Takara Bio, Shiga, Japan), which includes a high yield, Hot Start (HS) PCR enzyme, gel loading dye (green) and a density reagent in a 2X *Taq* PCR Master Mix format; ii) **ImmoMix™** Red (Bioline), which includes a heat-activated thermostable DNA polymerase and gel loading dye (red); and iii) **DreamTaq Green PCR Master Mix 2X** (Thermo Fisher Scientific).

Table1. List of primers used to amplify all the exonic regions of *CCDC39* gene (NM_181426.1), including the respective modifications in primer design and the PCR conditions used.

Exon	Primer Forward	Primer Reverse	Changes in Primer design Parameters	PCR Conditions/ PCR Mixture
1	CTCAACCGGAAGTTTCGCC	TGAAATAGGCAGAGAGGGTAAGGG		Stand-38c;Ta=58ºC PCR MM
2	AGCCTTCAGAACTGTGAGGAATTT	CAGCCTAGGAAACAGAGCGAGA		Stand-38c;Ta=58ºC PCR MM
3	TTTCTTACAACTAGTGAGCAACACC	CCATGTGCCTTTCACTGTCC	Tm min: 55ºC; AL min: 300bp	Stand-38c;Ta=58ºC PCR MM
4	GCATATTATAAACATCTGGAAGGC	GAAACAGTCAGCAATATTTTATGGG	Tm min: 55ºC; GCs min: 35%; AL min:300bp	Stand-38c;Ta=55ºC PCR MM
5	ATTGAGTTTTTATGCAGCTATATGC	CAATTCTAACTGTCAAACAGAGAGC	Tm min: 55ºC; GCs min: 32%;	Stand-38c;Ta=55ºC Emerald
6	CTTACAGAGACTATAAGCGCACAGG	ATCTACTGACTCTTTCTTGTTCCGC	Tm min: 58ºC	Stand-38c;Ta=55ºC PCR MM
7.1	TCTACGACCAGTGAAAGGCC	TGACATACCTCACCCTTCAGC	Tm min: 56ºC; GCs min: 35%	Stand-35c;Ta=63ºC Emerald
7.2	CAAGTTTTTGGAAAGTGAGATTGG	AATGAAAAGGTAACCTACAGAATGG	Tm min: 56ºC; GCs min: 35%	Stand-35c;Ta=63ºC Emerald
8	GGTTAGAACTTAAAGTGGGCAGAA	GATAGAAAACGTGCCTAAAAAGC	no GC clamp; Tm min:56ºC;	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	TGCCTAGTCCAAAGTCATAAGAAAG	TCCTCACAAGACATCTGTAAGGATT		Stand-38c;Ta=61ºC PCR MM
12	TCCCCAGACAATACTGAGGG	AAATTTCTTTACAAGACTGGAGGG	Tm min: 56ºC; GCs min: 36%;	Stand-38c;Ta=58ºC PCR MM
13	ATGGCCATTCATAGGGCTT	CAGCTAGTTCTCCTGACATCATC		Stand-38c;Ta=58ºC PCR MM
14	AGACACAATGAAACTGTGGCAGG	CCTTTGTTAAGTGTAATAGTGGCCC	Tm min: 56 ºC	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	Tm min: 55ºC; AL min: 350bp	Stand-38c;Ta=55ºC PCR MM
18-19	AAATTAGAAAGAGTGACCAAACAGG	GCGGTGATGTAGAAGTGGC	Tm min: 55ºC; GCs min: 35%;	Stand-38c;Ta=55ºC PCR MM
20	TTTGCCAATAATGAGAGGGTC	TGTAGCCTTGTCAAGAATCTGC		Stand-38c;Ta=58ºC PCR MM

Abbreviations: bp, base pair; CG min; minimal CG content; GC max, maximal GC content; Tm min, minimal temperature melting; Tm max, maximal temperature melting; AL min, minimal amplicon length; Ta, annealing temperature; MM, Master Mix; Emerald, EmeraldAmp Max HS PCR Master Mix; ImoMix, ImmoMix™ Red master mix.

Table 1. List of primers used to amplify all the exonic regions of *CCDC40* gene (NM_ 017950.3), including the respective modifications in primer design and the PCR conditions used.

Exon	Primer Forward	Primer Reverse	Changes in Primer design Parameters	PCR Conditions/ PCR Mixture
1	GTGGAGATGGCACAAAGGCC	GAGGACAAAGGGACAGGAGGG	GCs max:64%; AL min: 350 bp	High GC; PCR MM
2	TTCATGGAGCTTTGCTTTTGTG	CACAGTGCCCGTACTCTGATTC	Tm min: 56ºC	Stand38c; Ta=58ºC ImoMix
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	AGGAACGACAAGAGGGAGGG	GCATCTTGAACGGGTGTGG		Stand38c; Ta=58ºC ImoMix
16	GACAGAGAATGAGTTCGTGCGC	TGGTCTTCAACGGACAGAGGC		Stand-38c;Ta=58ºC PCR MM
17	AAACGTTTGCATAAGGAGCCC	CCTGGAAATCTCACCCATTGC		Stand-38c;Ta=58ºC PCR MM
18	CTCCATGTCCCTTGTTCCTCC	TGCAGTCTCCCAGACAGGC	GCs max: 64%;	High GC; PCR MM
19	CTGCTGAACTGTGCCAGTCTCC	TGGTGCTATTACTGGCCTGGG		Stand-38c;Ta=58ºC ImoMix
20	CTTTTCAGGCGTAGGTCTCGC	GTCACAGGCTCTGTCTTTGCG		High GC; PCR MM

Abbreviations: bp, base pair; CG min; minimal CG content; GC max, maximal GC content; Tm min, minimal temperature melting; Tm max, maximal temperature melting; AL min, minimal amplicon length; Ta, annealing temperature; MM, Master Mix; Emerald, EmeraldAmp Max HS PCR Master Mix; ImoMix, ImmoMix™ Red master mix.

.

Table 2. List of primers used to amplify all the exonic regions of *DNAI1* gene (NM_001281428.1), including the respective modifications in primer design and the PCR conditions used.

Exon	Primer Forward	Primer Reverse	Changes in Primer design Parameters	PCR Conditions/ PCR Mixture
1	TGGATTCTATCCTGCAAGGGC	GGTAAGAAACGGTGGATTGGG		Stand-38c;Ta=58ºC PCR MM
2	CTCCTCAAATGCCACAAAGTG	TTATAAGGTGGTAGGTTTAAGGGTG	Tm min: 55ºC, CG min-35%	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	AAAAAAGCCCAATAGCTGGTTG	TCCTTTCCCAAGTTACCCATGA		Stand-35c; Ta=61ºC PCR MM
7	CCTCCCTGCCACAGATTGG	AGGCTGAGGTCTGCTAGGGC	Tm max:63ºC; GCs máx:65%;	High GC; PCR MM
8	CAGAATGGACACAATATGGGC	CAGGGAGGCAAGTAGATGAGG	AL min:380bp; Tm min- 58ºC,	Multi-DMSC; Ta=63ºC Dream
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	TCTGCCCACAGCCCTAACTC	ACTTCCGTGGTGCTGCCTT		Stand-35c; Ta=61º PCR MM
15	GGAGGCAGGGAGCTAAATGG	TCTACAAACAGCCAAGGGTTCA		Stand-35c; Ta=61º PCR MM
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	GGCAGGGATCCAGAATTTACTG	GGGTGGGTTTCTACACAGGTTT		Multi-DMSC; Ta=63ºC Dream

Abbreviations: CG min, minimal CG content; GC max, maximal GC content; Tm min, minimal temperature melting; Tm max, maximal temperature melting; AL min, minimal amplicon length; Ta, annealing temperature; MM, Master Mix; Emerald, EmeraldAmp Max HS PCR Master Mix; Dream, DreamTaq Green PCR Master Mix 2X.

Table 3. List of primers used to amplify all the exonic regions of *AKAP3* gene (NM_001278309.1), including the respective modifications in primer design and the PCR conditions used.

Exon	Primer Forward	Primer Reverse	Changes in Primer design Parameters	PCR Conditions/ PCR Mixture
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(.1.1)	TGAGAGGGCTGATGACATTTGT	TGACTAGATTCGTGAGGCGGTT		Stand-35c;Ta=58ºC PCR MM
5(.1.2)	GGGAAACGTCCAACTCAGGAG	CATTCCTGGAACATGCAGAGAC		Stand-35c;Ta=58ºC PCR MM
5(.2)	GGGAGTTCAGTAGATGAAGTTTCC	CAATGGTTGTGTCTTTCACTTGG	Tm -57ºC	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	GGGAAACGTCCAACTCAGGAG	CATTCCTGGAACATGCAGAGAC		Stand35c;Ta=61ºC; Emerald

Abbreviations: Tm min, minimal temperature melting; Ta, annealing temperature; MM, Master Mix; Emerald, EmeraldAmp Max HS PCR Master Mix.

Table 4. List of primers used to amplify all the exonic regions of *AKAP4* gene (NM_003886.2), including the respective modifications in primer design and the PCR conditions used.

Exon	Primer Forward	Primer Reverse	Changes in Primer design Parameters	PCR Conditions/ PCR Mixture
1	TTAGAGCCCTCCATCTTTGTGC	TTCATTTCCTGCGAGACCTCC		Stand-38c;Ta=58ºC; PCR MM
2	CCCACCAGTTCCTAGCCTAAACC	TTACATCACACCACACCCTGCC		Stand-38c;Ta=58ºC; PCR MM
3	TGCTAATGCTAAGATTCTGTCTGCC	TTCAGAACCAGGGATCTTGGG		Stand-38c;Ta=58ºC; PCR MM
4	TTTGTGAGCATCTTCAGAGTGGG	TCTTACCATGTCTGATCCAGAGTCC		Stand-38c;Ta=58ºC; PCR MM
5.1	CAAACTCCAAAAGGTCAAATCCC	TTCTCCTTGATTTCCTTATGGGC		Stand-38c;Ta=58ºC; PCR MM
5.2	TCAGCTCCTCCAGCCAAACC	CAATCAAATCGGACACAATCTCC		Stand-38c;Ta=58ºC; PCR MM
5.3	AATTCCAGCATCTGTGGTCC	GCCAGTGAATCTGTGGAAGC		Stand-38c;Ta=58ºC; PCR MM
5.4	CACATTCTCAAAGAGGGCCTAACC	AATTGCCCGTTCGCTTGC		Stand-38c;Ta=58ºC; PCR MM
5.5	AGCAAAGCAGCTTCCATGTCC	TGAAACTCGAATGATCTTGGGC		Stand-38c;Ta=58ºC; PCR MM
6	AGATGTGTCAGCAGCCTCTAGTCC	CCAACACTGTATCACCATTCCTGG		Stand-38c;Ta=58ºC; PCR MM

Abbreviations: Ta, annealing temperature; MM, Master Mix.

Table 5. List of primers used to amplify 40 (of a total of 79) exons of *DNAH5* gene (NM_001369.2), and the respective modifications in primer design and PCR conditions used.

Exon	Primer Forward	Primer Reverse	Changes in Primer design Parameters	PCR Condition/ PCR Mixture
3	TCAACAGCTCAACACTTTAGGG	CAGGTCCGTTCTCACAGTAGC	Tm min-56ºC	Stand-35c; Ta=55ºC PCR MM
6	GGCTTTGGGTTGAATTGTGAGG	TGACAGTAAAGGAAAACCACATTGG		Stand-35c; Ta=61ºC PCR MM
7	ATTGGAAGCATGGAATACGC	AGCCTCCAAAGTGATGTGAGG	Tm min-56ºC	Stand-35c; Ta=58ºC PCR MM
8	AGGTCTCGCCTAATGGACTTG	CAATTCAGCTAGTGCAAAAAGG		Multi-DMSC; Ta=58ºC; PCR MM
9	TTAATGCAGATGGAATGGTTAGTC	CATAGGAAAGAAATTCCAAGAGC	Tm min-55C	Stand-35c; Ta=55ºC PCR MM
11	AATTTTGCTGTCGCCTTCAC	AGCACCATTACACAATTCTCAAAC	Tm min-56ºC; CGs min- 36%, GC clamp:1 residue	Stand-35c; Ta=55ºC PCR MM
12	TGATTTCTTTGAATGCAGTAGGGC	AATCATTGCTACACAGTGACAAACG		Stand-35c; Ta=58ºC PCR MM
13	AGTACCCGTGGAACCTGACG	TGACCCAAATTGCCAAGAGAG	Tm min-56ºC; CGs min- 36%, GC clamp:1 residue	Stand-35c; Ta=55ºC PCR MM
14	TGTCTTCCCAAGTTAATAGTTTTGC	GCTTAGATTCAACCCATCTGCC	Tm min-56ºC; CGs min- 36%	Stand-35c; Ta=55ºC Emerald
17	GGATATTTTATTAGCGACAACCAAG	TGATCTGAAATGATGTGTTCCC	Tm min-56ºC; CGs min- 36%, GC clamp:1 residue	Stand-35c; Ta=55ºC Emerald
25	GCACAGTTAACCTCTTGCGG	AGGAAGTTTCACAGAAATTTACCC		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	Tm min-55ºC; CGs min- 35%, GC clamp:1 residue	Stand-35c; Ta=58ºC PCR MM
34	AATACAGGAAACAATGAGAAACGTG	TTCTACTGGGTAAATGCAGATAGTG	Tm min-55ºC; CGs min- 36%, GC clamp:1 residue	Stand-35c; Ta=58ºC PCR MM
35	TGAATAGCTTGGCTGGTGATACAG	AGGCAAAGGAAGCAAGGTCTG	gc clamp: 1residue	Stand-35c; Ta=61ºC PCR MM

Abbreviations: CG min, minimal CG content; GC max, maximal GC content; Tm min, minimal temperature melting; Tm max, maximal temperature melting; AL min, minimal amplicon length; Ta, annealing temperature; MM, Master Mix; Emerald- EmeraldAmp Max HS PCR Master Mix; ImoMix, ImmoMix™ Red master mix; Dream, DreamTaq Green PCR Master Mix 2X.

Table 5. (continuation)

Exon	Primer Forward	Primer Reverse	Changes in Primer design Parameters	PCR Condition/ PCR Mixture
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
45	CCTATTATGTTTAAAGGCAGCTTGC	CGAACGTTTTCATATCTTACAGCC		Stand-35c; Ta=61ºC PCR MM
48	TTCATATCAGTTCAAAGGTTACCAG	CTAACTCCTTGAGTGTTTCCAAAG	Tm min- 55ºC; CGs min- 35%,	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	Tm min- 56ºC; GC clamp:1 residue	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	Tm max- 63ºC	Stand-35c; Ta=61ºC PCR MM
68	AGATATCGTTGTGAGTGATGATGCC	TTCCATGCCCAGGTATACAGTAGG		Stand-35c; Ta=61ºC PCR MM
72	GTGTATACTGGAACAAATGCATGG	GAGCTCTGATTCTTTCACACTGC	Tm min- 55ºC; GC min- 35%	Stand-35c; Ta=58ºC;PCR MM
73	TCTTGTTAGATTGGCAGGAAGTTG	ACGCCAAAGCTAGGAGGTCC	Tm min-57ºC	Stand-35c; Ta=61ºC PCR MM
75	AATTTAGACGGTGCTTATGTTGGC	GGAAGGTTCAGAAAAGCAGTCC		Stand-35c; Ta=58ºC Emerald
76	TTATGGAATAGTAAGAATTTGGGCCT	GGCTTCGTCCCTGTGCAA		Stand-35c; Ta=61ºC PCR MM
77	TTGAAACACTTACGCCAGAATC	AGCTAGCAGCCTGAGAACAAC		Stand-35c; Ta=58ºC PCR MM

Abbreviations: bp, base pair; CG min, minimal CG content; GC max, maximal GC content; Tm min, minimal temperature melting; Tm max, maximal temperature melting; AL min, minimal amplicon length; Ta, annealing temperature; MM, Master Mix.

Table 6. List of primers used to amplify all the exonic regions of *RSPH1* gene (NM_080860.3), including the respective modifications in primer design and the PCR conditions used.

Exon	Primer Forward	Primer Reverse	Changes in Primer design Parameters	PCR Condition/ PCR Mixture
1	CTATGACAGTACGCAAAGGCTGC	CCTCGATGGGAATTAACACAGG		Stand-35c; Ta=58ºC PCR MM
2	ATGAATTGCAAGCTCTGTGTTGG	TGTAGATCCCCTGAAACACATTG	Tm min: 58ªC	Stand-35c; Ta=58ºC PCR MM
3	CAGAACCGTTGACTTGACTGGATG	GGCCACAGGGACAGATTTCC		Stand-35c; Ta=58ºC PCR MM
4	TCTTAGCCCACTGTCCACCG	TGAAAGACTCACGTGCTCTGCC	Tm min: 58ªC	Stand-35c; Ta=58ºC PCR MM
5	TTGTGAGCCTGCAACAGCC	TGTACTGAGATCCAGAGAAACGGG		Stand-35c; Ta=58ºC PCR MM
6	TCCAATCCCATCAGCCACG	GGGAAATGATATTGTTTGCTTCTGG		Stand-35c; Ta=58ºC PCR MM
7	GTCTTTCAAAACTTGTCCTTCCTCC	AAAACAGGATGTCCACAACTCCC		Stand-35c; Ta=58ºC PCR MM
8	ATCAACACCCTCAGGAATGAAAG	GCAACGTTCTGAGATGGCTCC	Tm min: 58ªC	Stand-35c; Ta=58ºC PCR MM
9	CGTATTTTGCTTATTTTATGTCGGG	CCACTCACTGCAACTGCCAC	Tm min: 57ªC	Stand-35c; Ta=58ºC PCR MM

Abbreviations: Tm min, minimal temperature melting; Ta, annealing temperature; MM, Master Mix.

ATTACHMENT 3

In this supplementary information 67 candidate genes selected based on the literature and used for exome analysis are a listed.

Gene Symbol	*Gene full name*	*Locus*
ACVR2B	Activin A receptor, type IIB	3p22
ARMC4	Armadillo repeat containing 4	10p12.1-p11.23
C21orf59	Chromosome 21 open reading frame 59	21q22.1
CATSPER1	Cation channel, sperm associated 1	11q12.1
CATSPER2	Cation channel, sperm associated 2	15q15.3
CCDC103	Coiled-coil domain containing 103	17q21.31
CCDC114	Coiled-coil domain containing 114	19q13.33
CCDC65	Coiled-coil domain containing 65	12q13.12
CFC1	Cripto, FRL-1, cryptic family 1	2q21.1
CRISP2	Cysteine-rich secretory protein 2	6p12.3
DNAAF1	Dynein, axonemal, assembly factor 1	16q24.1
DNAAF2	Dynein, axonemal, assembly factor 2	14q21.3
DNAAF3	Dynein, axonemal, assembly factor 3	19q13.4
DNAH1	Dynein, axonemal, heavy chain 1	3p21.1
DNAH10	Dynein, axonemal, heavy chain 10	12q24.31
DNAH11	Dynein, axonemal, heavy chain 11	7p21
DNAH12	Dynein, axonemal, heavy chain 12	3p14.3
DNAH14	Dynein, axonemal, heavy chain 14	1q42.12
DNAH17	Dynein, axonemal, heavy chain 17	17q25.3
DNAH2	Dynein, axonemal, heavy chain 2	17p13.1
DNAH3	Dynein, axonemal, heavy chain 3	16p12.3
DNAH5	Dynein, axonemal, heavy chain 5	5p15.2
DNAH6	Dynein, axonemal, heavy chain 6	2p11.2
DNAH7	Dynein, axonemal, heavy chain 7	2q32.3
DNAH8	Dynein, axonemal, heavy chain 8	6p21.2
DNAH9	Dynein, axonemal, heavy chain 9	17p12
DNAI2	Dynein, axonemal, intermediate chain 2	17q25
DNAL1	Dynein, axonemal, light chain 1	14q24.3
DPCD	Deleted in primary ciliary dyskinesia	10q24.32
DRC1	Dynein regulatory complex subunit 1	2p23.3
DYX1C1	Dyslexia susceptibility 1 candidate 1	15q21.3
FOXJ1	Forkhead box J1	17q25.1
FSIP1	Fibrous sheath interacting protein 1	15q14
FSIP2	Fibrous sheath interacting protein 2	2q32.1
GAPDHS	Glyceraldehyde-3-phosphate dehydrogenase, spermatogenic	19q13.12
GAS8	Growth arrest-specific 8	16q24.3
HEATR2	HEAT repeat containing 2	7p22.3
HYDIN	HYDIN, axonemal central pair apparatus protein	16q22.2

Gene Symbol	Gene full name	Locus
INVS	Inversin	9q31
KIF3A	Kinesin family member 3A	5q31
KIFC1	Kinesin family member C1	6p21.3
LEFTY2	Left-right determination factor 2	1q42.1
LRRC6	Leucine rich repeat containing 6	8q24.22
NME8	NME/NM23 family member 8	7p14.1
ODF1	Outer dense fiber of sperm tails 1	8q22.3
ODF2	Outer dense fiber of sperm tails 2	9q34.11
ODF3	Outer dense fiber of sperm tails 3	11p15.5
ODF4	Outer dense fiber of sperm tails 4	17p13.1
OFD1	Oral-facial-digital syndrome 1	Xp22
RHPN1	Rhophilin, Rho gtpase binding protein 1	8q24.3
ROPN1	Rhophilin associated tail protein 1	3q21.1
ROPN1L	Rhophilin associated tail protein 1-like	5p15.2
RPGR	Retinitis pigmentosa gtpase regulator	Xp21.1
RSPH4A	Radial spoke head 4 homolog A	6q22.1
RSPH9	Radial spoke head 9 homolog	6p21.1
SEPT12	Septin 12	16p13.3
SEPT4	Septin 4	17q22
SEPT7	Septin 7	7p14.2
SPAG1	Sperm associated antigen 1	8q22.2
SPAG16	Sperm associated antigen 16	2q34
SPAG17	Sperm associated antigen 17	1p12
SPAG4	Sperm associated antigen 4	20q11.21
SPAG5	Sperm associated antigen 5	17q11.2
TCTE3	T-complex-associated-testis-expressed 3	6q27
TEKT2	Tektin 2 (testicular)	1p34.3
ZIC3	Zic family member 3	Xq26.2
ZMYND10	Zinc finger, MYND-type containing 10	3p21.3

ATTACHMENT 4

This attachment contains the primers' sequences and PCR amplification conditions used to confirm, the candidate variants selected from the whole exome sequencing of the patient compatible with Kartagener syndrome.

The selected genes for Sanger sequencing confirmation were: *CCDC103* (NM_001258398), *DNAH10* (NM_207437.3), *DNAH6* (NM_001370.1), *GAS8* (NM_001481.2), *SPAG17* (NM_206996.2), *INSL6* (NM_007179.2), *MYCBPAP* (NM_032133.4).

The primers' standard design parameters were: primer length 18-25 bp; percentage of GC 40%-60%; amplicon length (AL) 400-650 bp; and Tm 59Cº-62Cº. Some primers have slight modifications in the parameters, which are also referred in this supplementary data.

Two PCR conditions were used: 1) **Stand-35c**, with an initial denaturation at 95°C for 10 min, followed by thirty-five cycles of 95°C for 1 min, 58º for 30 s, and 72°C for 2 min, with a final extension at 72°C for 10 min; and 2) **High-GC**, a PCR with an initial denaturation at 95°C for 5 min, followed by forty cycles of 95°C for 45 s, 63°C for 45 s, and 72°C for 1 min, with a final extension at 72°C for 10 min.

The Stand-35c reaction mixture (30 µl) contained: 15µl of PCR Master Mix [(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 MgCl2]; 12 µl of sterile bidistilled water, 1 µl of each primer at 10 pmol/µl (Thermo Fisher Scientific, Einsteinstrasse, Germany) and 1 µl of DNA at 100ng/µl.

For the High-GC PCR condition, the reaction mixture (30µl) contained: 12.5 µl of DreamTaq Green PCR Master Mix 2x (Thermo Fisher Scientific); 1.5 µl of sterile bidistilled water; 5 µl of Betaine 5M (Sigma-Aldrich, Missouri 63103, USA); 1.5 µl of DMSO (Bioline, London, United Kingdom), 1 µl of each primer at 10 pmol/µl (Thermo Fisher Scientific) and 2.5 µl of DNA at 100ng/µl.

Table 7. List of the primers and PCR conditions used to validate the variants identified through the whole exome sequencing analysis.

Gene	Exon	Primer Forward (F) Primer Reverse (R)	Changes in Primer design Parameters	PCR Condition/ PCR Mixture
CCDC103	2	F: AATGCCTCCTCGACTGAAGC R: AAGATCAAACAGTAGACGGTGGG	GC min:35%	Stand-35c;Ta=58ºC PCR MM
DNAH10	47	F:TCCATTGTACGGCTCTGCC R: CACATCCTATAACCAACCGGC	Tm min:58ºC; GC min: 37%	Stand-35c;Ta=58ºC PCR MM
DNAH6	20	F:GAGAGCAATAAGGAGGAAAATGGGA R: ACTGGTCTAGGGAAAATGCTCTG		Stand-35c;Ta=58ºC PCR MM
GAS8	7	F: CTGCGACAGGTGGATAGGTG R: AGGAGAACACGAGGCAGAGA		Stand-35c;Ta=58ºC PCR MM
SPAG17	31	F: AAGAACAGCAGGCCACACTCC R: TGGCATGGGATGTAAAAAGGG		Stand-35c;Ta=58ºC PCR MM
INSL6	1	F: CGCAGACAGGGAGCAGGGT R: GAGCTGTGTACTAGGCACAAATTCC	Tm máx:64ºC; GC máx:70%	High GC; Ta=63º; Dream
MYCBPAP	1	F: CCGCAGGGCTTTCTAGGG R: AGTGCGCTGTAGGCTCTTCCT	Tm máx:60ºC; GC máx:67%;	High GC; Ta=63º; Dream

Abbreviations: F, forward primer; R, reverse primer; CG min, minimal CG content; GC max, maximal GC content; Tm min, minimal melting temperature; Tm max, maximal melting temperature; Ta, annealing temperature; PCR MM, PCR Master Mix (Promega); Dream, DreamTaq Green PCR Master Mix 2X (Thermo Fisher Scientific).