Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
Biochimica, Biologia Molecolare e Biotecnologie

CICLO XXVIII

COORDINATORE Prof. Bernardi Francesco

Molecular Basis of Primary Ciliary Dyskinesia

Settore Scientifico Disciplinare BIO/18

Dottorando
Dott. Olcese Chiara

(firma)

Tutore
Prof. Bernardi Francesco

(firma)

Co-Tutore
Prof. Bartoloni Lucia

(firma)

Anni 2013/2016

Abstract

Primary Ciliary Dyskinesia (PCD) is an autosomal recessive disorder due to a dysfunction of the epithelial cilia that do not move properly. It is often caused by a defect in one of the dynein arms, protein complexes that connect between them the axonemal microtubules. The ciliary malfunction leads to respiratory problems, can cause reversal of the asymmetry of internal organs and, in males, may cause infertility, because the sperm flagellum has the same internal architecture as cilia. The PCD is a genetically heterogeneous disease: mutations in any gene coding for each one of the proteins involved in the cilia ultrastructure, assembly or composition, can alter the ciliary function.

This study aims at assessing the percentage recurrence of mutations in the 34 PCD known genes and to identify new causative genes to explain the part of PCD patients who remain genetically uncharacterized. The entire exome of a cohort of 65 individuals, belonging to 46 families, was sequenced. Using a custom designed HaloPlex DNA selection panel, including 640 ciliary genes, a total of another 96 additional individuals (85 patients and 9 family members) was analyzed. Only 40% of patients have mutations or likely pathogenic variants in known PCD genes. These findings are different from a recent estimation made by another group, in the United States, that assessed the known PCD genes prevalence around 60%.

In 25% of additional families strong candidate genes that will now pass the functional validation were found.

In the 3% of patients mutations in *CFTR* were identified, therefore they suffer of mild Cystic Fibrosis.

In almost 68% of patients mutations in known PCD genes or strong candidates were detected, but a molecular diagnosis couldn't be done for the remaining 32% of patients. Therefore different hypotheses on this remaining 32% can be put forward and the research on ciliopathies still remains an interesting and largely to be studied field.

Abstract

La Discinesia Ciliare Primaria (PCD) è una malattia autosomica recessiva caratterizzata dalla disfunzione delle cilia epiteliali che non si muovono correttamente. Essa è spesso causata da un difetto in una delle braccia di dineina, complessi proteici che collegano tra loro i microtubuli dell'assonema. La disfunzione ciliare produce problemi respiratori, puo' dare origine ad asimmetria degli organi interni e, negli uomini, è talvolta causa di infertilità, in quanto il flagello degli spermatozoi presenta la stessa architettura interna delle cilia.

La PCD è una malattia geneticamente eterogenea: la mutazione in uno solo dei geni che codificano per le proteine coinvolte nella struttura e nel funzionamento delle cilia puo' alterare la fisiologia ciliare.

Questo studio intende stimare la percentuale di ricorrenza di mutazioni in geni già identificati quali responsabili della PCD e di identificarne di nuovi per poter confermare la diagnosi nei pazienti che restano non caratterizzati geneticamente.

L'intero esoma di una coorte di 65 individui appartenenti a 46 famiglie è stato sequenziato. Inoltre, utilizzando un pannello Haloplex personalizzato, contenente 640 geni ciliari quali bersaglio del sequenziamento, un totale di altri 96 pazienti è stato analizzato.

Solo il 40% degli individui presi in esame presenta mutazioni patogeniche nei geni che causano la PCD. Questa percentuale differisce da quella stimata da un altro gruppo statunitense che ritiene invece che ben il 60% dei pazienti si possa caratterizzare grazie ai soli geni attualmente noti.

In alcune famiglie (25%) sono stati individuati dei geni che, codificando per proteine coinvolte nella struttura e nel funzionamento delle cilia, sono ritenuti essere possibile causa di PCD. Essi verranno sottoposti a validazione funzionale.

In circa il 3% degli individui sono state rilevate mutazioni nel gene *CFTR.* Questo porta a concludere che tali soggetti siano affetti da Fibrosi Cistica piuttosto che da PCD.

In conclusione circa il 68% della coorte presa in esame è stato definito geneticamente, dacchè sono state reperite mutazioni patogeniche in geni noti o in buoni candidati. Tuttavia, per il rimanente 38% non è stato possibile raggiungere una diagnosi molecolare.

La presenza di individui non ancora caratterizzati geneticamente (32%) dà adito alla formulazione di plurime ipotesi. La ricerca sulle ciliopatie rimane pertanto un settore di indagine interessante ed ampiamente ed ulteriormente studiabile.

Acknowledgements

I would like to express my sincere gratitude to Lucia Bartoloni, Jean-Louis Blouin, Francesco Bernardi, Pier Giorgio Borasio, Alessandra Stupiggia, Arturo Colamussi, Ximena Bonilla, Emanuele Alberto Zanocco, Federico Lerda, Antonio Walter Riotto, Hannah Mitchison, Mitali Patel and to all researchers and technicians who are members of S.Antonarakis laboratory (CMU, Gènève). A special thank to Lorna and Av Mitchison for the kind welcome they gave me in UK.

Contents

Introduction

Cilia, evolutionarily conserved organelles, were recenty found to play multiple cellular functions by controlling diverse intracellular signaling cascades. Defects in their assembly or function leads to a group of diseases called ciliopathies [1]. Up to date ciliary function is not incompletely understood, as well as cilia interactions with other cellular compartment. But genetics has recently helped to identify genes mutated in human ciliopathies and to understand the functional connections between the encoded proteins. The type of mutation and the subpart of cilia for which the mutated gene encodes determine very heterogeneous phenotypes in patients [2-5].

Primary Ciliary Dyskinesia (PCD) is an autosomal recessive disorder affecting approximately 1 in 30,000 live births. It is caused by axoneme structural defects and dysfunction of the epithelial motile cilia.

The cilium shows a structure composed of hundreds of proteins, encoded by different specific genes often transcribed into different isoforms, derived by alternative splicing [6]. A mutation in any axoneme-related protein could potentially alter ciliary function. Therefore this suggests that PCD is a highly genetically heterogeneous disease.

Motile cilia are present on specialised epithelial cells lining the respiratory airways, brain, epididymis and fallopian tubes. In PCD patients, defective cilia impact on the airways mucociliary clearance, cause chronic respiratory disease, infections and permanent lung damage (bronchiectasis). Moreover defective sperm flagella often lead to subfertility (asthenozoospermia). Patients can display hydrocephalus and about 50% shows laterality defects: situs inversus (in the major part of cases) or severe isomerisms (6-12%) with cardiac defects due to malfunction of embryonic node cilia that are supposed to distribute morphogens during development [7]. The clinical symptoms are very highly variable and can affect different organs [8].

Cilia

The cilium (Figure 1) is a long cellular projection sustained by a complex cytoskeleton structure, named the axoneme, covered by the plasma membrane [9].

The axoneme is composed of nine peripheral microtubule doublets surrounding, in motile cilia, two central microtubules (9+2) and several other structures such as outer dynein arms (ODAs) and inner dynein arms (IDAs), radial spokes, and nexin bridges. These structures are synthesized in the cytoplasm and stocked in a cytoplasmic pool in which there is a pre-assembly of multisubunit structures that are afterwards transported to the site of final assembly by intraflagellar transport (IFT), [6, 10].

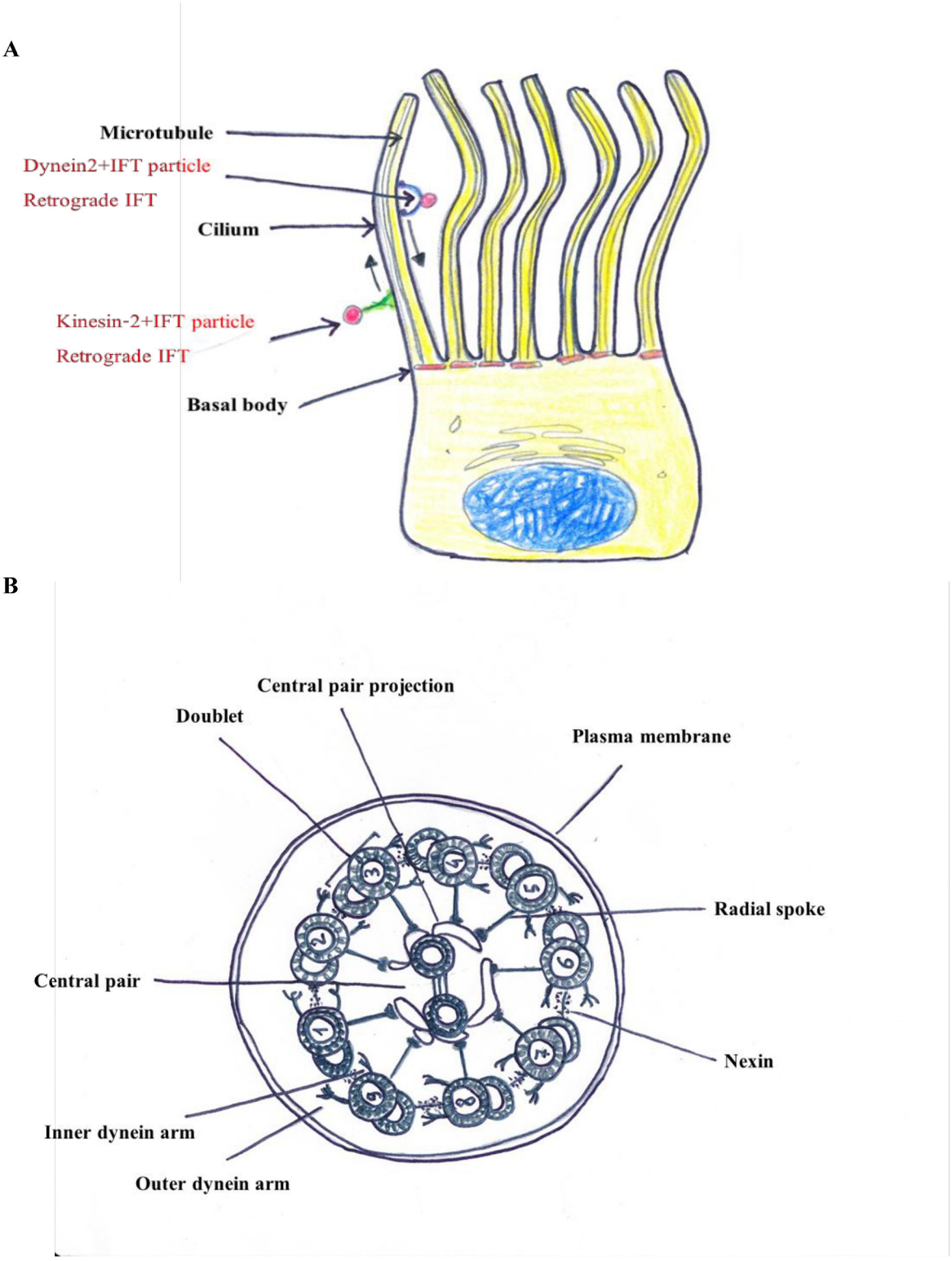


Figure 1. A) Ciliated cell. **B**) Cilium cross-section [Picture Copyright C. Olcese].

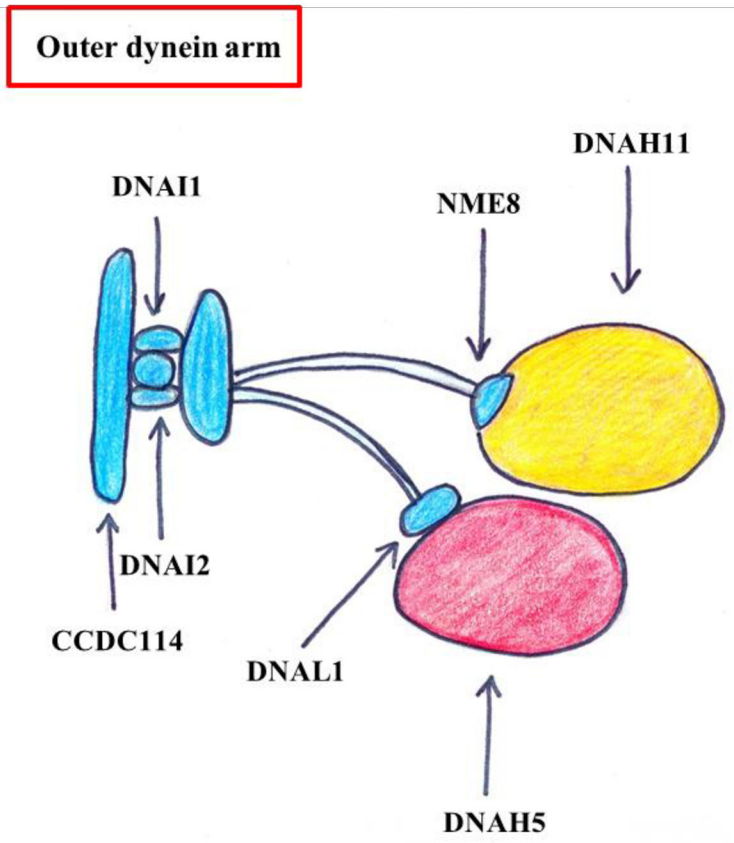


Figure 2. Outer dynein arm structure. The PCD genes which encode its subunit are shown [Picture Copyright C. Olcese].

In human cells cilia are quite ubiquitous. This suggests an evolution of this ancient organelle across species and its development in vertebrates in order to facilitate a broad range of functions, above all the sensory and motility ones [11].

Starting during the two decades between 1960 and 1980 and continuing to the present, primary cilia have been described and localized in many different tissue cells [12], including several neuronal types [13, 14] and myocytes [15].

Nodal and Primary cilia, because of the lack of the central-pair, were defined as 'sensory' cilia, contrary to the mechanical function of 'motile' cilia. They show a 9 +0 pattern (with only 9 peripheral doublets). However, it is now clear that this functional-based classification is far from reality. Infact, both the 9+2 and 9+0 cilia share mechano- and/or chemosensor functions [3, 16-19].

.

Motile cilia

Motile cilia are the most studied ones. They localize on tracheal epithelial cells, on brain ependymal ones and on reproductive tract cells which form the inner coating of the the oviduct and the epididymis. Motile cilia are often massively concentrated on surface cells and they beat together to allow fluid movement (e.g. mucociliary clearance and cerebrospinal fluid movement), the ovum transport along the fallopian tubes and the movement of the spermatozoa flagella.
The typical microtubule pattern of motile cilia is a 9 +2 one, with 9 microtubule peripheral doublets and a central doublet important for motility generation [20]. Each doublet is formed by two microtubules fused together, while the central doublet is formed by two separate microtubules [7]. In the axoneme there are several accessory proteins that serve to maintain the structure that allows elastic movement (e.g the elastic protein nexin that connect doublets one to each other).

The dyneins permit the cilia movement. They are organized in multiprotein complexes that protrude from one doublet to the adjacent one and are divided into inner arms (IDAs) and outer arms (ODAs). Each dynein includes one or more heavy chains (>500kD) and intermediate chains (40-140kD) and 1-10 light chains (8-28 KD).

The ODAs and IDAs neighbouring arms are linked together by filamentous structures called outer-outer-dynein (OOD) linkers or outer-inner (OID). ODAs and IDAs' function is supported by an ATPase activity that allows each microtubule to slide on the adjacent one, causing the ciliary beat where the doublets 1–4 are responsible for the effective stroke and the 6-9 for the recovery one. The ODAs regulate the frequency of beats, while the IDAs control the wavebeat form [20, 21].
Microtubule doublets are linked to the radial spokes (RS), thin stalks, spaced every 96 nm, with a globular head interacting with the central pair projections in cycles of detachment and reattachment. The RS form pairs (*C. reinhardtii*) or triplet groups (*T. termophila*). These 'thin stalks' protrude a right-handed helix along the axoneme. The RSs were hypotized to coordinate the dynein activity [22-26].

Furthermore a subset of six axonemal polypeptides was found to interact with inner arms and this is referred to as the "Nexin-Dynein Regulatory Complex" (N-DRC). The N-DRC was hypothized to be linked to inner arms by two proteins identified in *C. reinhardtii*: an actin and a Caltractin/centrin. Studies *in vivo* and *in vitro* were conducted to observe inner arms' Ca++-sensitiveness and to individuate what other proteins interact with actin and caltractin/centrin [27-29].
The Nexin- Dynein Regulatory Complex (N-DRC), connected to the outer dynein arms (ODAs) via the outer–inner dynein linkers, regulates both ODAs and IDAs [30]. The lack of any ODAs, IDAs and N-DRC subunit impairs the movement of cilia and flagella. RS and DRC, strictly interacting

between them, allow the doublet sliding to become effective by disrupting the symmetry of its pattern.
In fact, *C. reinhardtii* mutants that lack radial spoke heads, show a more asymmetrical and unspecific cilia beating [31].

The role of the nine outer doublets is to perform the reciprocal and dynein dependent, sliding movement, but the function of the central pair of two microtubules (CP) and the radial spokes (RS) is still unclear. The central pair apparatus, the radial spokes and the nexin link-dynein regulatory complex (N-DRC) are thought to be the manager of cilium motors.
Developmental, structural, biochemical, and functional studies suggest that the RS structures function as a sensor network, both mechano (e.g. interactions between the radial spokes and central apparatus) and chemical (e.g. second messengers binding, anchored kinases and phosphatases' enzymatic activities), to control motility in 9+2 cilia and flagella [32, 33]. The motile cilia and flagella's RS, in fact, contain kinases and phosphatases that act on other doublets and are not absolutely required for oscillatory beating [33].
N-DRC may on one hand construct an architectural, useful mechanical connection between all cilia ultrastructure elements, on the other may impact on ciliary and flagellar motility by chemical signals. N-DRC involves cyclic nucleotide monophosphates and calcium second messengers and axonemal proteins acting as receptors and enzymes [20].
But, although the disruption of central structures impair cilia movement [34] [35], suggesting a role in cilia beat management, recent studies have demonstrated that also the 9+0 nodal cilia are motile, and that this motility plays a critical role in establishing left-right asymmetry in developing embryos [36-38].

Primary cilia

Primary cilia have a 9+0 ultrastructure architectural pattern and they are almost ubiquitous in human tissues [39]. By the use of specific antibodies and epifluorescence, primary cilia were observed by light confocal-microscopy [40]. The ultrastructural comparison between the axoneme of primary cilia (referred to as 'sensory' cilia because of being localized in organs for olfaction, hearing and sight) and motile cilia (found in the respiratory tract, reproductive tract, and lining the ependyma) revealed that primary cilia lack the central microtubules pair, required for an active motility. However the composition of each doublet is identical to the motile cilia ones.
Studies in *C. elegans* neurons and in mammalian kidney show that, although of different achitectural pattern, all primary cilia function as either chemo- or mechanosensors by still unclear mechanisms [41].

Nodal Cilia

The nodal cilia are typical of the embryonal nodal cells. They protrude from the embryonal cells surface during the gastrulation-stage and they are thought to play a critical role during early embriogenesis allowing the right body asymmetry. They present the same 9+0 structure as primary cilia. Therefore they were, at first, supposed to be immotile. A deeper characterization showed that they have a rotational motion, differing from the primary cilia 'because they can move in a propeller-like fashion' [36].
Afzelius proposed that these 9+0 motile cilia are required for correct heart positioning in the the embryo, the basis of a subsequent correct situs [42]. This hypothesis is supported by the fact that *situs inversus* is a result of mutant axonemal intermediate or heavy dynein chains. In this case the ciliary immotility causes the inversion of the left-right body axis [43-47].
The origin of the body's left-right (L-R) asymmetry in vertebrates has been deeply investigated. However, the point zero of the symmetry break still remains unclear.
There are two main theories on it:

1_ Nodal cilia have a mechanical role in fluid flow in the definition of the body's L-R pattern. They are required for the transport of extracellular morphogens that are responsible for a correct body axis assess. The evidence was provided by studies on mouse. Mice embryos, both with motile or immotile cilia, were tested with artificial forward (with the same direction of the intrinsic leftward nodal flow) and reverse flows sufficient to generate a reversal situs in wild-type embryos and to properly direct the situs of mutant ones with immotile cilia [38].

2_ Nodal cilia affect the embrional flow and consequently the symmetry through a chemical, not mechanical, mechanism. Whereas two different forms of cilia are present in the node with motile

cilia in the central region and immotile cilia on around cells, Yoshiba and colleagues hypothized that it is the Ca(2+) channel Polycystin-2 (Pkd2), localized in the perinodal cells cilia, that manages the flow. They observed that artificial flows fail to correct the L+R asymmetry in Kif3a mutant embryos which lack all cilia. Interestingly, after the restoration of the mutant cilia, the flows acted properly. These findings suggest that Pkd2 of nodal crown cells cilia is responsible for 'sensing' flows and managing their direction [48, 49].

PKD2 mutations account for 15% of autosomal dominant polycystic kidney disease (ADPKD) families [50]. PKD2 mutated ADPKD patients show *situs inversus* [51]. A heterozygous stop codon mutation alone was found to cause both ADPKD and complete situs inversus [52]. Moreover, although it is a very rare phenotype, association of situs inversus totalis and autosomal recessive polycystic kidney disease has also been reported [51, 53].

Cilia seem to be required also in later stages of development for a correct evolution of different tissues, but this is not the case for nodal ones.

The Intraflagellar Transport (IFT)

Cilia are fixed to cells surface by the basal body wich is also involved in the regulation of proteins movement along the cilium: the so called Intra Flagellar Transport (IFT) mechanism.
The IFT is thought to be involved into the construction and maintainance of both cilia and flagella [54-56]. It was firstly observed in *C. reinhardtii* ciliogenesis. The movement of cilia-structural proteins requires the IFT motors, kinesins complexes and a specialized cytoplasmic dyneins system to allow the anterograde and the retrograde transport along the cilium.
IFT-involved-proteins (about 19), divided into two complexes moving along the outer doublets, are highly conserved in both motile and sensory (e.g mammalian photoreceptor) cilia of ciliated eukaryotic cells [57] and in protistan motile cilia [58]. Algae, nematodes and mammals with an IFT-deficient phenotype and that are mutated in these proteins show a failing set of cilia and flagella.
Since cilia and flagella lack protein synthesis, IFT seems to be the only mechanism accounting for management and transport of proteins involved in cilia axoneme assembly, maintenance and function.
In spite of the studies in *C. reinhardtii* on the precise mechanism by which the proteins are pre-assembled in the cytoplasm and then transported to the axoneme, it remains still unclear [59-63].
One hypothesis is that IFT is based on a kinesin II-motor system to allow the transport of structural proteins from the basis to the tip of the cilium and to permit, viceversa, the return of nonaxonemal dynein in the cell body [59].

Cilia Ultrastructure Assembly and PIH- proteins

Studies in *Chlamydomonas* reveal that ciliary proteins are preassembled in multisubunit complexes (such as ODAs, IDAs, and radial spokes) and pooled in cytoplasm. Then they are sorted and transportated to cilia. The cytoplasmic preassembly was studied aiming to explain the interactions of the proteins coded by PCD causative genes [49, 64].

Analysis of dynein-deficient mutants of *C. reinhardtii* has identified four proteins that are required for cytoplasmic preassembly of dynein complexes: PF13 (also known as Ktu or DNAAF2 [65], ODA7 (also known as LRRC50 or DNAAF1) [66, 67], MOT48 [68] and PF22 (also known as DNAAF3) [69].

Transmission electron microscopy (TEM) and high-resolution immunofluorescence microscopy were performed on respiratory cilia and sperm from DNAAF2 mutated patients. They showed axonemal dynein arms abnormality and absence of DNAH5 and DNAI2 (both component of ODAs) on the distal ciliary axonemes and a residual staining on proximal ones. Moreover DNAI2 was completely lost in sperm, together with DNALI1, orthologue of *C. reinhardtii* p28 a component of various IDAs subtypes.

Mutations in Pf13 produce ciliated cell paralysis in *C. Reinhardtii*. IF and Western blot co-precipitation studies showed that Pf13 impacts on ODAs preassembly and stability. The orthologous Ktu immunoprecipitation was performed in mouse and it resulted to interact with Hsp70, that is, with Hsp90, the most expressed chaperon proteins in mouse testes [65].

Studies in vivo on ida10 *C. reinhardtii* mutant, whose phenotype shows inner arm dyneins deficiency, resulted in the identification of two genes, involved in dynein preassembly: Pf13/Ktu and MOT48.

Interestingly both *C. reinhardtii* Pf13/Ktu and MOT48 are PIH-domain containing proteins. This suggested a specific role of PIH- proteins in preassembly of axonemal dyneins.

PIH1 domain-containing proteins together with the kintoun (*Saccharomyces cerevisiae* PIH1) are members of the PIH1 protein domain family.

PIH1 is a component of the conserved R2TP complex which is Hsp90 associated. The R2TP, highly conserved across species, is required for phosphatidylinositol-3 kinase-related protein kinase (PIKK) signalling and it is involved in RNA polymerase II assembly [70].

Cilia sistemic importance and Ciliopathies

Cilia have different signaling sensibilities and mechanisms: sensory and motile functions can be found alone or blended in each different kind of cilia, in different cell types and at different developmental stages [71-78].

Evidence on cilia function comes from studies on cilia-dependent kydney diseases, where there are primary cilia localizing on kydney tubular epithelial cells. Mouse *Kif3a* disruption results in a phenotype presenting cystic distal-nephron lesions due to the loss of cilia, whereas in the proximal-nephron tubules, expressing normal cilia, the results are noncystic [79]. Similar data were provided by the use of morpholino antisense oligonucleotides to disrupt conditionally various IFT proteins in zebrafish pronephros [80]. Therefore, although the exact physiology of this mechanism remains unclear, the primary cilium is thought to act as a mechanosensor in the renal tubule, measuring the flow rates in the collecting duct and regulating the calcium signaling.

The 'dual' or 'mixed' cilia function is underlined also by ciliophathies overlap: different ciliophaties' phenotypes are often blended, both in human and mouse models. Pathological mutations in genes encoding for structures and processes conserved between motile and primary cilia, result in diseases and phenotypes impairing the function of both types of cilia. Meanwhile in those cases in which the disease is due to the mutation of one specific type of cilia component (e.g. dynein arms in motile cilia) the phenotype spectrum is less wide.

The ciliopathies are a group of heterogeneous diseases often characterized by multi-organ defects (including kidney, brain, limb, retina, liver, and bone). Over a dozen disorders are now considered to be within the ciliopathy spectrum including Joubert syndrome (JBTS), nephronophthisis (NPHP), Senior-Loken syndrome (SLS), orofaciodigital (OFD) syndrome, autosomal dominant and recessive polycystic kidney disease (ADPKD and ARPKD), Leber congenital amaurosis (LCA), Meckel-Gruber syndrome (MKS), Bardet-Biedl syndrome (BBS), Usher syndrome (US), some forms of retinal dystrophy (RD), hydrocephalus and oncogenesis (when cilia defects affect cells division) [81-84].

Ciliopathies can show very rare phenotypes (Bardet–Biedl: OMIM mp (minimum prevalence) 1/ 13500; Meckel–Gruber syndrome: OMIM mp 1/135000, Primary Ciliary Diskinesia OMIM mp 1/30000) or less rare diseases, such as polycystic kidney disease (PKD, OMIM mp 1/1000). Nevertheless the ciliopathies seems to share some symptoms such us renal impairment, retinal problems, obesity, mental retardation, and defects of the left–right body axis [1, 85].

It is to be noted that some mutated genes were found to be common to different ciliopathies. This offers an intriguing matter for study and increases interest in further cilia studies.

Policystik Kydney Desease (PKD) patient phenotypes show the involvement of other organs. Cysts in the pancreas and liver, intrahepatic biliary duct dilatation and fibrosis, vascular deficiency, retinitis pigmentosa, photoreceptors defects, diabetes insipidus, sensorineural hearing loss and

congenital hepatic fibrosis were all observed in humans and mouse in association with PKD [86-90].

In these cases, a convincing explanation is the presence of pathological mutations in genes encoding for structures and processes higly conserved and essential for cilia structure and function.

The polycystin-1 (PC-1) and polycystin-2 (PC-2) genes (associated with human cystic kidney disorders, PKD), nephronophthisis (nephrocystin-1, nephrocystin-2, and nephrocystin-4) and Bardet-Biedl syndrome (BBS 1–8) ones (associated with cysts in non-PKD disorders) encode all for proteins that localize in the primary cilia and/or in the basal body [91].

Several BBS proteins were studied in *C. elegans*. They usually localize at the cilia's base and, like IFT-involved proteins, move bidirectionally along the ciliary axoneme. In particular BBS-7 and BBS-8, were found to be required by other IFT proteins function (OSM-5/Polaris and CHE-11), suggesting a BBS proteins role in the selection and assembly of IFT components [62].

Animal Models

After the association of Polycystic Kidney Disease (PKD) with aberrant primary cilia, a wide range of ciliopathies was related to primary cilia. Therefore *in vitro* and *in vivo* models were designed to study primary ciliary function and diseases [92-97].

Primary cilia are highly conserved structures, allowing the possibility of studying them across a range of species.

The use of lower organism models such as *C. reinhardtii* and *C. elegans,* the availability of Zebrafish (*Danio rerio*) and murine models and the culture of human ciliated cells helped the understanding of the cilia's systemic importance. *C. reinhardtii*

For example Zebrafish and its embryos permit the study of human gene variants *in vivo* and their impact on development and can give a deeper insight into ciliopathies [80, 98, 99]

The injection of translation- or splice-blocking antisense morpholinos [80] and then the 'humanization' of the model by human BBS mRNA are used to rescue the morphant phenotypes and to observe the functional consequences of any introduced variant [99, 100].

Primary Ciliary Diskinesia

The disease

Primary Ciliary Dyskinesia is a genetic disorder characterized by altered ciliary and flagellar motility. It is usually an autosomal recessive disorder that occurs in approximately 1:30000 live births. Because the cilia are present in different tissues, the disease assumes a multisystem character. The standard symptoms affect the respiratory system, such as respiratory infections, that can lead to bronchiectasis [101]. Since the flagellum has a structure similar to the cilia, PCD often lead to infertility or male subfertility (asthenozoospermia). In the 50 % of PCD cases, situs inversus viscerum is present, which means that the organs are reversed, and this is due to the malfunction of the primary monocilio in Hensen's node, which apperars to be fundamental for the distribution of morphogens during development [7, 102]. If in PCD patients there are respiratory infections, bronchiectasis and situs inversus viscerum, the clinical picture takes the name of Kartagener syndrome. Since defects in cilia can cause other types of diseases, PCD is a syndrome that affects mainly the respiratory system and is chronic and progressive. Other clinical manifestations are highly heterogeneous and vary from person to person and may affect different organs [8].

Diagnosis and Therapy

The diagnosis of PCD is often delayed or missed completely. The several tests that can be used differ in effectiveness, invasiveness and easiness of analysis [103, 104].
PCD patients show a wide range of symptomes, more or less severe and numerous. The disease is similar to Cystic Fibrosis and other air ways affections.
Therefore the median age of diagnosis in Europe is about 5.3 years, lower in children with situs inversus (3.5 years vs 5.3 years) [105, 106].
The PCD management and provision care, based on a surveyor from 196 centres in 26 European countries, results to be not centralised in most countries, whereas there is a median of only four patients (IQR 2–9) per centre and a very variable diagnostic approach [107]. However in few cases PCD management is centralised. For example in England the Department of Health recently commissionated four national Paediatric PCD Centres, based in Leeds/Bradford, Leicester, London and Southampton, to take care of all PCD patients from England and Scotland [106, 108].

In theory, the diagnostic process for PCD should at first include the optical microscope analysis of the nasal mucosa, to see if there is deficiency in the ciliary movement, and then, using the *in vitro* neociliogenesis, should check whether the fault of the movement is primary or secondary. At this

point the electron microscope analysis should be carried out to identify the structural defect of the cilia. Having identified the missing protein, the geneticist would have an idea of which gene has to be investigated.
In reality what actually happens is that the doctor sends a sample directly to the geneticist, without ultrastructure data, and the geneticist chooses how many genes and on which he performs the analysis.

Example of a standard diagnostic protocol

-Clinical evidence
In clinical practice the following symptoms have to be carefully assessed: chronic rinithis-sinusitis (100%), recurrent othitis media (95%), neonatal respiratory symptoms (such as upper air-ways infections and bronchectiasis) (73%) and situs inversus (55%). In some cases also hydrocephalus, retinal abnormalities, abnormal sperm flagella and infertility in adult males are detected.

-Nitric Oxide Measurement
Nitric Oxide (NO) level is usually reduced in PCD cases. Therefore the NO measurement can be used as a noninvasive screening test [109, 110].

- Saccharin test
Saccharin test permits an assessment of mucociliary function by measuring the delay in the perception in the mouth of a microtablet of saccharin placed on the inferior turbinate. However this test is difficult to perform in children aged <12 yrs and its results are unreliable. For this reason now a blue dye is added and colouring of the tongue is assessed [111, 112].

-Nasal Brushing
Nasal Brushing is performed on inferior turbinates and it is somewhat tricky for patients. Moreover samples are a blend of ciliated cells, muco and blood. When HSPV and TEM fail (27% of cases) because of poor material, patients understandably refuse to repeate the brushing and the samples undergo a cell-culture protocol that unfortunately has an efficiency of 52% only.

-High Speed Video Microscopy (HSVM)
This is performed on fresh samples containing still alive ciliated cells. It allows to check of both the beat frequency (40-50 beats/second for normal cilia) and the movement waveform. It records 500fps and play back at 60fps. Micro-videos are done on side, top and front of section.

-Immunofluorescence (IF)

Since the wider availability of antibodies to detect assonemal and dynein proteins, IF is now routinely used to study cilia-associated protein expression [113].

The nasal brushing is scraped on slides and allowed to dry overnight. It is then fixed in 4% paraformaldheyde, before being analysed with specific antibodies.

Transmission electron microscopy (TEM)

TEM on cilia obtained by nasal brushing has an important role in PCD diagnosis, allowing observations on different structural alterations. TEM presents some technical limits, for example in detecting the poor electron dense inner dynein arms, and more subtle ultrastructural defects can be missed.

PCD patients usually show ODA defects (37%), ODA+IDA defects (21%), IDA defects (18%), disarrangenment defects (18%), transposition defects (11%) and normal ultrastructure (13%).

It is to be noted that both HSPVM and EM do not succeed in detecting all PCD cases (e.g DNAH11-linked cases show normal axonemal structure).

To release an EM result 300 single cilia are counted and observed.

Electron Tomography (ET)

Electron Tomography (ET) has a much higher resolution that falls in the 2 nm range. This method uses 2D projection images, collected by 'photographing' the targeted object from many different angles, to profile its 3D image [114-116]. ET is used to obtain a detailed picture of patient cilia macromolecular complexes then to compare these with the expected healthy cilia structure. It permits rotating of the sample by 2° until 70° left and 70° right. The sample can then be rotated of 90° manually. At the end a 3D imaging reconstruction is obtained by specific softwares.

Genetics

Primary ciliary dyskinesia (PCD) is a phenotype with a high clinical variability. Insofar that, in a set of monozygotic twins, one twin migh show situs inversus and the other might not [117].

Genome-wide linkage analysis failed to find a single locus or two major loci responsible for the majority of families with PCD [118]. This clinical variance is explained by the fact that PCD should be considered a very heterogeneous autosomal recessive disorder with many genes involved in its phenotype.

All genes encoding proteins that either contribute to the structure or motility of the cilia (e.g the different dynein genes and those encoding the ciliary proteins such as tubulins, connexins, kinesins, nexins and microtubule-associated proteins) or are responsible for ciliary or laterality phenotypes are strong PCD candidates, since they can alter the ciliary function. Approximately 2,500 putative genes and their associated proteins are currently thought to be involved in both morphology and function of cilia: their axonemal and basal body structural biochemistry, the IFT transport, the sensing/signal transduction mechanisms (http://www.ciliaproteome.org [119]).

The first proteomic analysis of the human cilia axoneme resulted in the availability of 1400 peptides'sequence data and in the identification of at least 200 proteins related both structurally and functionally to the axoneme [120]. Each of these different proteins is encoded by a different gene and is present in different isoforms obtained by alternative splicing [121].

Up to date 34 genes are recognised to cause PCD. A loss of outer dynein arms (ODAs) (Figure 1) results from mutations in ODA proteins (DNAH5 [122], DNAI1 [123], DNAI2 [43], DNAL1 [124], DNAH11 [45], NME8 [125], DNAH1 [126]) or in the ODA docking complex (CCDC114 [127], ARMC4 [128]). Loss of both inner dynein arms (IDAs) and ODAs are caused by mutations in dynein assembly genes and attachment factors (CCDC103 [129], DYX1C1 [130], SPAG1 [131], C21ORF59 [132], DNAAF1/LRRC50 [66], DNAAF2/KTU [65], DNAAF3 [69], LRRC6 [133], ZMYND10 [134], HEATR2 [135], CCDC151 [136]). More rare defects affecting the Nexin–Dynein Regulatory Complexes (N-DRC) and causing microtubule disarrangements and inner dynein arm (IDAs) loss, are due to other gene mutations (CCDC39 [137], CCDC40 [138], CCDC65 [135], CCDC164 [139], GAS8 [140]). Mutations of the central microtubules (CCs) and radial spokes causing central pair defects also occur (RSPHA4 [141], RSPH9 [141], RSPH1[142], HYDIN [47], RSPH3 [143]). The ciliary aplasia due to defects in the multiciliated cell differentiation pathway is caused by two genes (CCNO [144], MCIDAS [145]). Two PCD subtypes are X-linked recessive syndromic forms with retinitis pigmentosa (RPGR gene [146]) and oral-facial-digital syndrome 1 symptoms (OFD1 gene [147]). Since genetic screening has not been widely funded for PCD, despite, to date, 34 PCD genes being identified, the known causative-PCD mutations account for only a percentage of the PCD cases, and about 1/3 of PCD patients remain genetically uncharacterized [148-150].

Gene	Function	Name	Locus	Reference	Page
DNAI1	ODA	dynein, axonemal, intermediate chain 1	9p13	Pennarun et al., 1999	22
DNAH5	ODA	dynein, axonemal, heavy chain 5	5p15.2	Omran et al., 2000	23
DNAI2	ODA	dynein, axonemal, intermediate chain 2	17q25	Pennarun et al., 2000	24
NME8	ODA	NME/NM23 family member 8	7p14-p13	Sadek et al., 2001	25
DNAH11	ODA	dynein, axonemal, heavy chain 11	7p21	Bartoloni et al., 2002	25
DNAL1	ODA	dynein, axonemal, light chain 1	14q24.3	Horvath et al., 2005	26
DNAH1	ODA	dynein, axonemal, heavy chain 1	3p21.1	Imtiaz et al., 2015	26
ARMC4	ODA docking complex	armadillo repeat containing 4	10p12.1	Hjeij et al., 2013	27
CCDC114	ODA docking complex	coiled-coil domain containing 114	19q13.3	Onoufriadis et al., 2013	27
RSPH4A	CCs	radial spoke head 4 homolog A (Chlamydomonas)	6q22.1	Castleman et al., 2009	28
RSPH9	CCs	radial spoke head 9 homolog (Chlamydomonas)	6p21.1	Castleman et al., 2009	28
HYDIN	CCs	HYDIN, axonemal central pair apparatus protein	16q22.2	Olbrich et al., 2012	29
RSPH1	CCs	radial spoke head 1 homolog (Chlamydomonas)	21q22.3	Kott et al., 2013	30
RSPH3	CCs	radial spoke head 9 homolog (Chlamydomonas)	6q25.3	Jeanson et al., 2015	31
CCDC39	N-DRC	coiled-coil domain containing 39	3q26.33	Merveille et al., 2010	32
CCDC40	N-DRC	coiled-coil domain containing 40	17q25.3	Becker-Heck et al., 2010	33
CCDC65	N-DRC	coiled-coil domain containing 65	12q13.12	Horani et al., 2013	34
DRC1/CCDC164	N-DRC	dynein regulatory complex subunit 1	2p23.3	Wirschtell et al. 2013	34
GAS8	N-DRC	growth arrest-specific 8	16q24.3	Olbrich et al., 2015	35
DNAAF2	dynein assembly	Dynein, Axonemal, Assembly Factor 2	14q22.1	Omram et al., 2008	35
DNAAF1	dynein assembly	Dynein, Axonemal, Assembly Factor 1	16q24.1	Duquesnoy et al., 2009	36
DNAAF3	dynein assembly	Dynein, Axonemal, Assembly Factor 3	19q13.42	Geremek et al., 2011	37
CCDC103	dynein assembly	coiled-coil domain containing 103	17q21.31	Panizzi et al., 2012	37
HEATR2	dynein assembly	HEAT repeat containing 2	7p22.3	Horani et al.., 2012	38
LRRC6	dynein assembly	leucine rich repeat containing 6	8q24.22	Kott et al., 2012	38
C21orf59	dynein assembly	Chromosome 21 Open Reading Frame 59	21q22.1	Austin-Tse C, et al., 2013	39
DYX1C1	dynein assembly	dyslexia susceptibility 1 candidate 1	15q21.3	Tarkar et al., 2013	39
SPAG1	dynein assembly	sperm associated antigen 1	8q22.2	Knowles et al., 2013	40
ZMYND10	dynein assembly	zinc finger, MYND-type containing 10	3p21.3	Moore et al., 2013	40
CCDC151	dynein assembly	coiled-coil domain containing 151	19p13.2	Hjeij et al., 2015	41
CCNO	cell differentiation	cyclin 0	5q11.2	Wallmeier et al., 2014	41
MCIDAS	cell differentiation	multiciliate differentiation and DNA synthesis associated cell cycle protein	5q11.2	Boon et al., 2014	42
PCD subtypeslinked to other diseases					
OFD1		oral-facial-digital syndrome 1	Xp22	Budny et al .,2006	43
RPGR		retinitis pigmentosa GTPase regulator	Xp21.1	Moore et al., 2006	44

Table 1. PCD known genes. The ID, the name and the chromosomal locus of each PCD-causative gene is showed in column 1, 3 and 4 respectively. The function of the encoded protein is reported in column 2, the first article which reported each gene in column 5, the page where gene detailed information is reported in column 6.

Known genes

DNAI1 (OMIM: 604366)

The Dynein Axonemal Intermediate Chain 1 (*DNAI1*) gene (NM_012144.3, GI: 526479828) encodes an intermediate chain axonemal protein (DNAI1_HUMAN, Q9UI46; NP_036276.1, GI: 6912338), part of the dynein complex of respiratory cilia and belonging to the large family of motor proteins. It contains several 'WD repeats', highly conserved during evolution, and it maps to chromosome 9p13-p21. In 1999 the human cDNA sequence encoding DNAI1 (AF063228; GI: 3493584) was isolated by cross -species PCR, using degenerated primers derived from *C. reinhardtii* hortologous cDNA sequences coding for IC78 and IC69 [123]. This gene disruption, althoug explaining only 2% of Primary Ciliary Dykinesia [151] is associated with PCD type 1 (CILD1) with or without situs inversus and infertility [44, 152]. Patients present nasal NO levels (nl/min) very low compared with controls. TEM studies on tracheal and inferior turbinates epithelial cells show cilia that have an abnormal ciliary ultrastructure characterized by defects in, or by the absence of, the outer dinein arms (ODAs) and/or IDAs, often associated with non-specific defects of microtubular organization [44, 152-154]. The IVS1+2-3insT mutation is the most prevalent pathogenetic change in DNAI1 and it accounts for the 54%-55% of the mutations identified worldwide, above all in white patients [153, 154].

Figure 3. DNAI1 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

DNAH5 (OMIM: 603335)

The human *DNAH5* (dynein axonemal heavy chain 5) gene (NM_001369) encodes for an axonemal heavy chain dynein (NP_001360), that functions with ATPase activity to generate the force for power stroke. It maps to 5p15.2 and was identified by homozygosity mapping and linkage analysis performed in one consanguineous family [122]. The transcript is 79 exons long and results in DNAH5 protein that is considered to be the major molecular motor of ODAs. It functions as a force-generating protein with ATPase activity, whereby the release of ADP is thought to produce the force-producing power stroke. Previous studies identified mutations in DNAH5 accounting together for more than 25% of all PCD cases [46, 155].

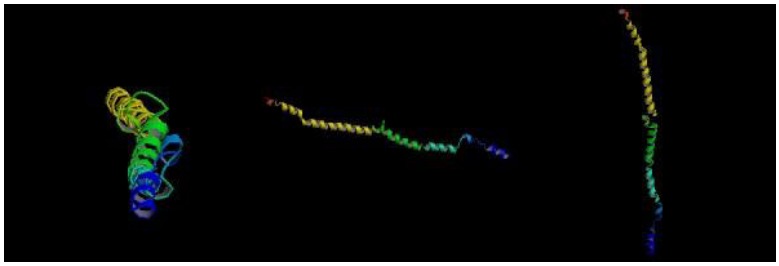


Figure 4. DNH5 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

DNAI2 (OMIM: 605483)

The human axonemal dynein intermediate chain 2 (*DNAI2*) gene maps to chromosome 17q25. Composed of 14 exons, it was identified starting from its *C. reinhardtii* hortologous IC69. It is highly expressed only in trachea and testis [43]. The encoded protein (DNAI2_HUMAN, Q9GZS0) belongs to the dynein intermediate chain family and is part of the dynein complex of respiratory cilia and sperm flagella. Mutations in this gene are associated with primary ciliary dyskinesia type 9 with or without *situs inversus* [43, 156].

Alternatively spliced transcript variants (NM_023036.4, GI: 289802995; NM_001172810.1, GI: 289802996) encoding different isoforms (NP_075462.3, GI: 217416452; NP_001166281.1, GI: 289802997) are present for this gene. In patients lacking DNAI2 expression, high resolution immunofluorescence imaging of both ODAs type 1 (typical of proximal ciliary axoneme, DNAH9 negative and DNAH5 positive) and type 2 (typical of distal cilary axoneme, DNAH5 and DNAH9 positive) reveals an aberrant expression pattern for both these two ODA proteins. DNAH9 is absent and DNAH5 seems to be absent or highly reduced. These observations suggest that DNAI2 is involved in the regulation of the correct assembly of both (distal and proximal) ODA complexes [156]. Otherwise the cells of patients mutated in DNAH5 lack both DNAH5 and DNAI2, revealing that DNAH5 is essential for the correct assembly of DNAI2 within axonemal ODA types 1 and 2.

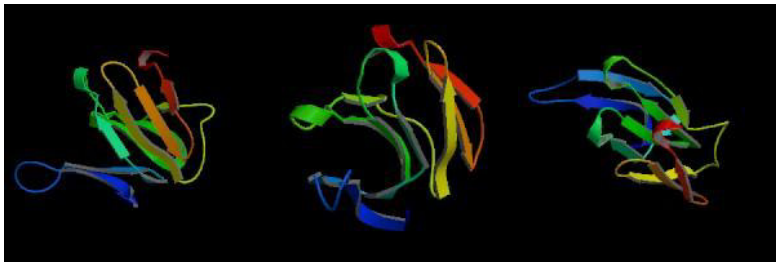


Figure 5. DNAI2 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

NME8/TXNDC3 (OMIM: 607421)

The NME/NM23 family member 8 (*NME8*; previously named thioredoxin domain containing 3, TXNDC3) gene, maps on chromosome 7p14.1 [125, 157]. Its transcript (NM_016616.4, GI: 254911059), derived from 18 exons, encodes for the sea urchin IC1 ortholog protein (TXND3_HUMAN, Q8N427; NP_057700.3, GI: 148839372), a component of sperm outer dynein arms. NME8 is probably involved in the reduction of disulfide bonds of fibrous sheath (FS) at final stages of sperm tail maturation. *TXNDC3* mutations are associated with CILD6. Patients presented PCD clinical symptoms and infertility. TEM analysis revealed ultrastructure anomaly of respiratory cilia.

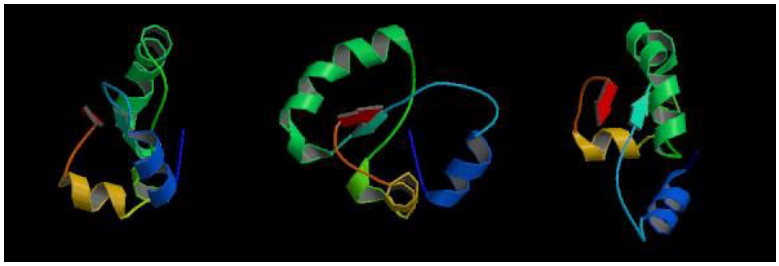


Figure 6. NME8 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

DNAH11 (OMIN: 603339)

The Dynein Heavy Chain Axonemal 11 (*DNAH11*) gene maps to chr7p21. It was firstly identified in patients with situs inversus totalis and probable Kartagener Syndrome (AJ320497, OMIM: 244400, [45]. Only one gene transcript (NM_001277115.1) has been described and it encodes for DNAH11 protein DYH11_HUMAN, Q96DT5 (NP_001264044.1) which is a member of the dynein heavy chain family with an ATPase microtubule-dependent motor function. Mutations in DNAH11 are a cause of primary ciliary dyskinesia type 7 (CILD7, OMIM: 611884). TEM analysis found normal axonemes or with few nonspecific abnormalities [158]. Mouse mutated at the outer arm dynein heavy chain 11 was found to be a good model of human PCD. They show the inversus viscerum, immotile tracheal cilia with normal ultrastructure, reduced sperm motility, gross rhinitis, sinusitis and otitis media, all indicators of human PCD. Therefore DNAH11 gene should be investigated in patients with normal ciliary ultrastructure and static or hyperkinetic cilia [159].

Figure 7. DNAH11 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

DNAL1 (OMIM: 610062)

The Dynein Axonemal Light Chain 1 (*DNAL1*) gene encodes an axonemal dynein light chain protein which is part of the outer dynein arm complexes, acting as a molecular motor to produce an ATP-dependent cilia movement. This gene maps to chr14q24.3, it belongs to the dynein light chain LC1-type family [124]. Alternate splicing results in two transcript variants (NM_031427.3; NM_001201366.1) that encode for two different protein isoforms (NP_113615.2; NP_001188295.1) of which the second one has a shorter N-terminus. The encoded protein (DNAL1_HUMAN, Q4LDG9) interacts physically with DNAH5, via the LC1/γ heavy chain-binding motif, and it regulates DNAH5 function in sperm flagella and in respiratory and ependymal cilia [124]. Furthermore it interacts with tubuline in an unknown manner. Defects in DNAL1 are the cause of primary ciliary dyskinesia type 16 (CILD16) [MIM: 614017]. TEM analysis of cells from mutated patients revealed the absence of ODAs. The patients phenotype was characterized by complete situs inversus, a low nasal NO and a significant decrease in cilia movement. Infertility problems were not reported [160].

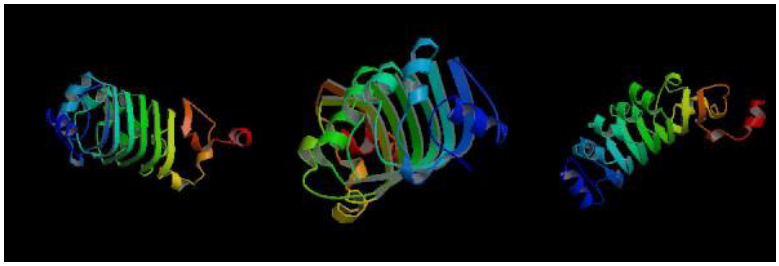


Figure 8. DNAL1 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

DNAH1 (OMIM: 603332)

Homo sapiens dynein axonemal heavy chain 1 (DNAH1) is a 81 exons gene (Gene ID: 25981) mapping on chr 3p21.1. Its transcript (NM_015512.4) encodes the enormous (4265aa) axonemal protein heavy chain dynein 1 (NP_056327.4) which is an inner dynein heavy chain expressed in testis. Studies on mice MDHC7(-/-), lacking the DNAH1 hortologous, revealed a MDHC7-deficient sperm, unable to move. Moreover the beat frequency of tracheal cilia was approximately 50% less in MDHC7(-/-) mice than in wild type. The reduction in both ciliary and flagellar motility is not correlated with any evident defect in the axonemal structure. This MDHC7(-/-) mice phenotype is similar to that observed in some patients suffering from PCD. Therefore it has been suggested that the homologous human gene DNAH1 (HDHC7) is PCD-causative [161]. A cohort of individuals with primary infertility due to multiple morphological abnormalities of the flagella (including absent, short, coiled, bent, and irregular ones) was found to carry a homozygous mutation in DNAH1 resulting in the complete loss of both DNAH1 transcript and protein. A general axonemal disorganization including mislocalization of the microtubule doublets and the loss of the inner dynein arms was also observed [126, 162].

ARMC4 (OMIM: 615408)

The Armadillo repeat-containing protein 4 (ARM4, Gene ID: 55130) gene maps on chromosome 10p12.1-p11.23. The gene's transcript (NM_018076.2, GI: 31657113) encodes for an axonemal protein (ARMC4_HUMAN, Q5T2S8, NP_060546.2, GI: 74934). ARMC4 localizes both within the ciliary axonemes and at the ciliary base in human respiratory ciliated cells. Patients with ARMC4 mutations show PCD symptoms with or without situs inversus and reduced nNO levels. The ARMC4 protein is absent from the ciliary axonemes but it is still detectable at the ciliary base. The ciliary beating is severely impaired [128].

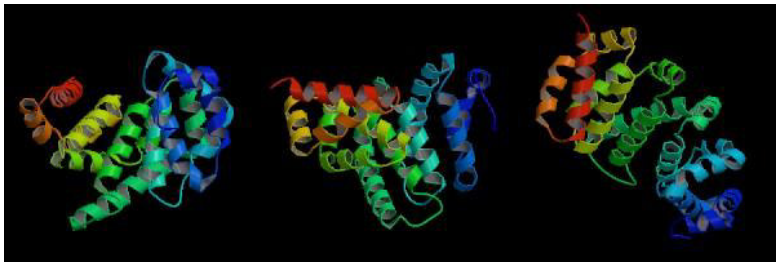


Figure 9. ARMC4 protein 3D structure**.** Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

CCDC114 (OMIM: 615038)

The Coiled-Coil Domain Containing 114 (*CCDC114*) gene (Gene ID: 93233), hortologous to *C. reinhardtii* DCC2, maps on chromosome 19q13.3. Its transcript (NM_144577.3, GI: 93233) encodes for an axonemal ODA microtubule-docking complex subunit (CC114_HUMAN, Q96M63,NP_653178.3,). Gene mutation is associated to PCD with left/right asymmetry and laterality malformations. No analysed patients reported infertility problems. High speed video imaging of their respiratory epithelial cells showed deficiency in ciliary beat frequency versus control cells. TEM data revealed a common loss of the ODAs resulting in immotile cilia. In these cells reduced levels of CCDC114 was also related to a reduction of DNAH5 concentration [127].

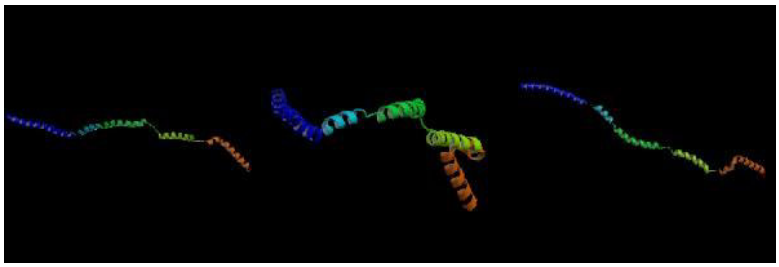


Figure 10. 3D structure of CCDC114 protein. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

RSPH4A (previously RSH3L, OMIM: 612647)

The Radial Spoke Head Protein 4 Homolog A (*RSPH4A*) gene belongs to the flagellar radial spoke RSP4/6 family and it maps to 6q22.1. The gene (NM_001010892, GI:239582733; NM_001161664.1, GI:239582734) encodes a protein (RSH4A_HUMAN, Q5TD94) that appears to be a component of the radial spoke head, as determined by homology with the *C. reinhardtii* RSP4 gene. The transcript, alternatively spliced, produces two different transcript variants; encoding for two different isoforms 716 and 600aa long (NP_001010892.1, GI: 58219006; NP_001155136.1, GI: 239582735). Mutations of the gene are associated with ciliary dyskinesia type 11 (CILD11, OMIM: 612649). Patients clinical phenotype consisted of low nasal NO (<100nL/min) and full-loss of function in respiratory cilia. No one presented situs inversus. TEM observation of nasal ciliary epithelium from patients showed central-microtubular pair defect [141]. The absence of the central pair (9+0 structure) was the most frequently detected defect [141, 163, 164].

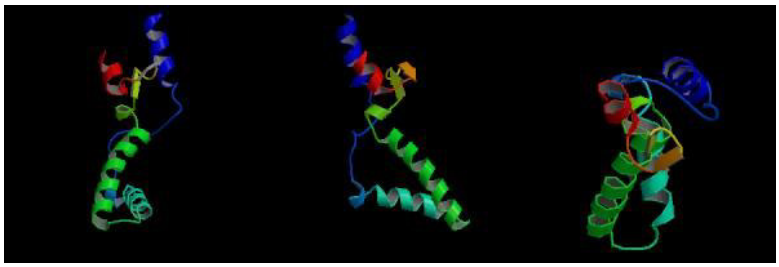


Figure 11. RSPH4A protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

RSPH9 (OMIM: 612648)

The Radial Spoke Head 9 Homolog previously (*RSPH9*) encodes a protein (RSPH9_HUMAN, Q9H1X1) thought to be a component of the axonemal radial spoke head in motile cilia and flagella. This gene which is part of the RSP9 family maps to 6p21.1 [141]. The gene is characterized by an alternative splicing that brings into two different transcripts (NM_152732.4, GI: 301069347; NM_001193341.1, GI: 301069350) that encode for two different protein isoforms (NP_689945.2, GI: 32964825; NP_001180270.1, GI: 301069351). Defects in the gene are the cause of primary ciliary dyskinesia type 12 (CILD12, OMIM: 612650) [165]. Patients mutated in RSPH9 presented low NO. No one showed situs inversus. TEM of nasal brushing cells revealed a generally normal 9+2 structure.

HYDIN (OMIM: 610812)

The axonemal central pair apparatus protein *HYDIN* gene (Gene ID: 54768) is very large (423 kb, 91 exons) and maps on chromosome 16q22.2. In humans its locus was found to be duplicated (recently in evolution history), in a nearly identical 360-kb paralogous segment inserted on chromosome 1q21.1 [166]. The HYDIN gene in chr16q22.2 was identified as candidate gene by homozigosity mapping of PCD-associated loci and by parallel whole exome sequencing [47]. TEM, performed on respiratory cells with loss of function mutations, revealed a normal ODAs and IDAs ultrastructure and a 9+2 axonemal composition, but high resolution Electron Microscopy Tomography showed that the cilia lacked part of the central pair projection. Furthermore High-speed videomicroscopy resulted in reduced amplitude in respiratory cilia beating and stiff sperma flagella [47]. The encoded protein that localizes in cell projections seems to be involved in cilia motility and it contains a MSP (Major sperm protein) domain. Four different transcripts (NM_001270974.1; NM_017558.4; NM_001198542.1; NM_001198543.1) account for four different protein isoforms (NP_001257903.1; NP_060028.2; NP_001185471.1; NP_001185472.1) where the isoform a is the longest.

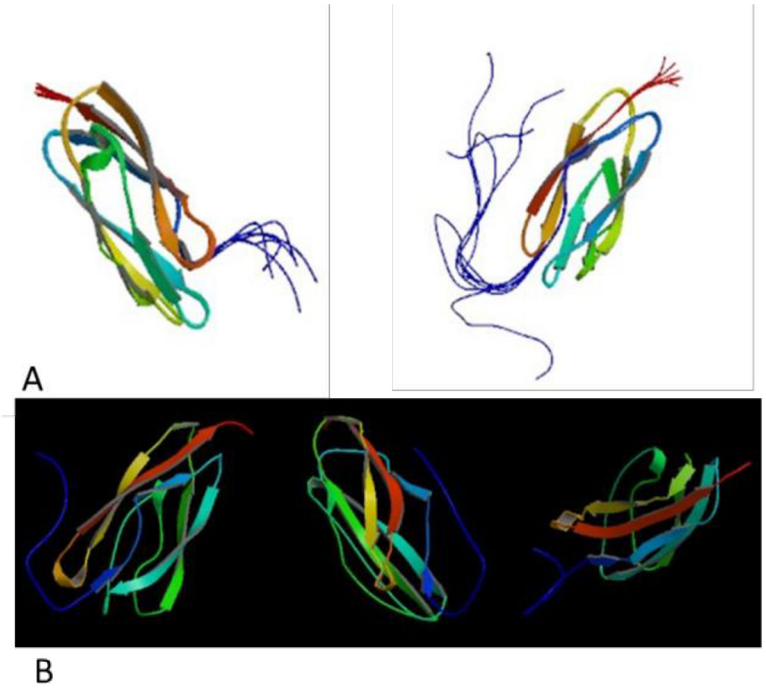


Figure 12. A.HYDIN protein 3D structure. **B.** HYDIN 3D front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

RSPH1 (OMIM: 609314)

The *RSPH1* gene was identified by Whole Exome Sequencing (WES) and it maps on chromosome 21q22.3. Its transcript (NM_080860.2), derived from 9 exons, encodes for the RSPH1 protein (NP_079566.1), part of the RS-head protein family, and it is involved in the anatomy of the central pair of cilia microtubules together with RSPHA4 and RSPH9 and HYDIN [142]. Radial spoke (RS) proteins and the central complex (CC) are thought to control cilia and flagella beating by a sensor activity. The mRNA, is expressed mainly in trachea, lungs, airway brushing and testis cells. Patients' phenotype was characyterized by typical PCD symptoms. They had low levels of nNO and fertility problems. No one presented situs inversus. TEM data revealed central pair defects in patients with mutated RSPH1 (9+1 or 9+0 structure) and high-speed videomicroscopy of airway brushing cells confirmed total ciliar immotility [142].

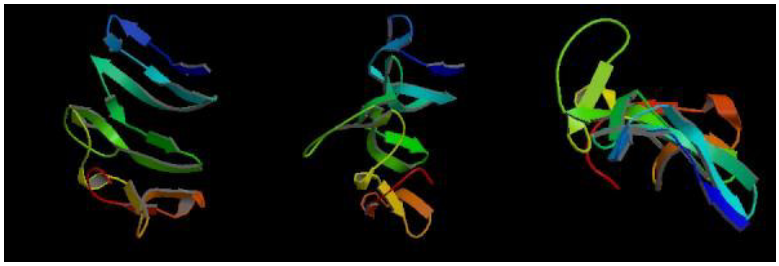


Figure 13. RSPH1 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

RSPH3 (OMIM: 615876)

The human radial spoke head protein 3 homolog (NM_031924.4; Gene ID: 83861), divided in 9 exons, maps on chromosome 6q25.3 and it encodes by alternative splicing for two protein isoforms 560 and 464 amino acids long (NP_114130.3). RSPH3, whose ortholog in the flagellated alga *C. reinhardtii* encodes a RS-stalk protein, is mainly expressed in respiratory and testicular cells. Its protein product, which localizes within the cilia of respiratory epithelial cells, was undetectable in airway cells from an individual with RSPH3 mutations and in whom RSPH23 (a RS-neck protein) and RSPH1 and RSPH4A (RS-head proteins) were found to be still present within cilia. In the case of RSPH3 mutations, high-speed-videomicroscopy analyses revealed the coexistence of immotile cilia and motile cilia with movements of reduced amplitude. A striking feature of the ultrastructural phenotype associated with RSPH3 mutations is the near absence of detectable RSs in all cilia in combination with a variable proportion of cilia with CC defects. Overall RSPH3 mutations contribute to disease in more than 10% of PCD-affected individuals with CC/RS defects, thereby allowing an accurate diagnosis to be made in such cases. It also unveils the key role of RSPH3 in the proper building of RSs and the CC in humans [143].

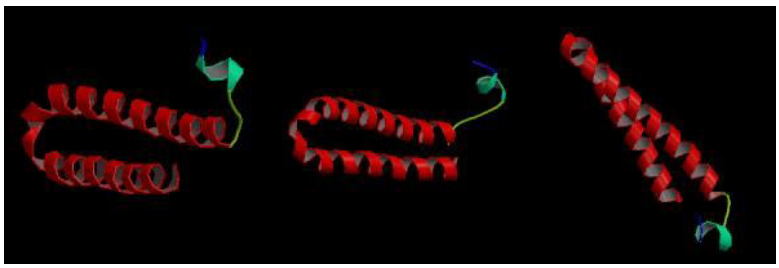


Figure 14. RSPH3 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

CCDC39 (OMIM: 613798)

The Coiled-Coil Domain-Containing Protein 39 (*CCDC39*, Gene ID: 339829) gene maps to 3q26.33 and it codes for a protein (CCD39_HUMAN, Q9UFE4) required for assembly of dynein regulatory complex (DRC) and inner dynein arm complexes (IDAs), involved in the motility of cilia and flagella and in beat regulation [137]. Only one gene's transcript (NM_181426.1, GI: 157785667) encodes for one isoform (NP_852091.1, GI: 157785668).

Defects in *CCDC39* cause primary ciliary dyskinesia type 14 (CILD14, OMIM: 613807). Situs inversus and oligoasthenospermia are often present in patient carrying mutations in this gene. TEM analysis of patients' respiratory epithelial cells revealed disorganization of the peripheral microtubule doublets, absence or shifting of the central pairs and reduction or complete loss of the IDAs. On the contrary ODAs results were normal [137, 167].

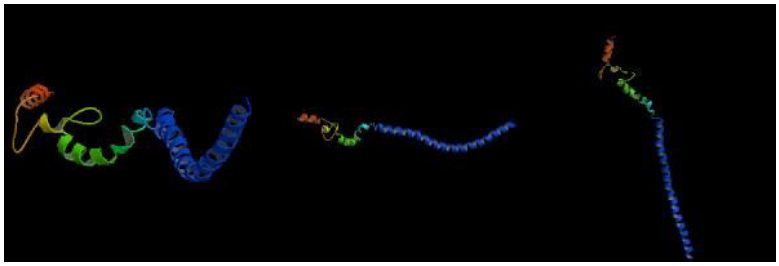
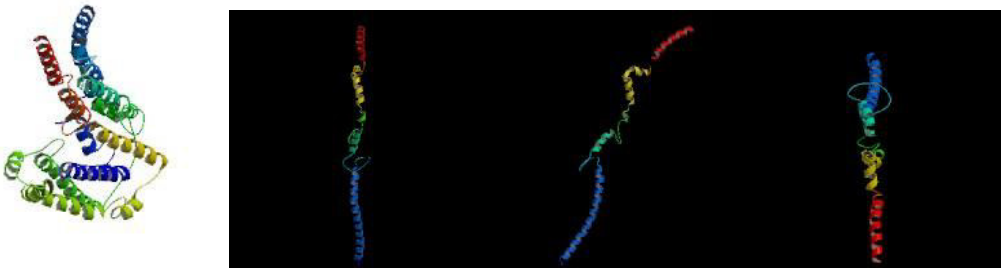


Figure 15. CCDC39 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

CCDC40 (OMIM: 613799)

The Coiled-coil domain-containing 40 (*CCDC40,* Gene ID: 55036) gene maps to 17q25.3. This gene encodes a protein (CCD40_HUMAN, Q4G0X9) with the same function of CCDC39 which is required for the assembly of dynein regulatory complex (DRC) and inner dynein arm complexes (IDAs), involved in the motility of cilia and flagella and in beat regulation [138]. Two transcript variants (NM_017950.3; NM_001243342.1) derive from alternative splicing and they encode for two different isoforms (NP_060420.2; NP_001230271.1). Mutations of *CCDC40* determine PCD variants characterized by misplacement of the central pair of microtubules and defective assembly of IDAs and dynein regulatory complexes. In particular defects in *CCDC40* are the cause of primary ciliary dyskinesia type 15 (CILD15 OMIM: 613808) [138].

TEM analysis found defects in the axonemal structures of mobile cilia, such as displacement of outer doublets, reductions in the mean number of inner dynein arms (IDA) and abnormal radial spokes and nexins links [138], disorganization of microtubule doublets and central pair doublet disorder [167].



A B C D

Figure 16. 3D Structure of CCDC40 human protein. The protein model viewed from the front (B), the top (C) and the side (D). Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

CCDC65 (OMIM: 611088)

The human Coiled-coil domain containing 65 (*CCDC65,* Gene ID: 85478) gene, orthologue of the *C. reinhardtii* nexin-dynein regulatory complex protein DRC2, maps to 12q13.12. It encodes (NM_033124.4, GI: 85478) for CCD65_HUMAN, Q8IXS2 protein (NP_149115.2) that is part of the nexin-dynein regulatory complex (N-DRC). Patients' nNO and fertility are reduced. Situs inversus was never observed. TEM studies on human tracheal airway epithelial cells, silenced for *CCDC65* by RNAi approach, revealed a normal ciliary structure and the absence of Gas8 (a nexin-dynein regulatory complex corresponding to *C. reinhardtii* DRC4). Videomicroscopy on inferior turbinate nasal cells reveales diskinetic (stiff and hyperkinetic) cilia beating. These data confirm CCDC65 role in regulating beating and not in constituting ODAs, IDAs or Ccs [135].

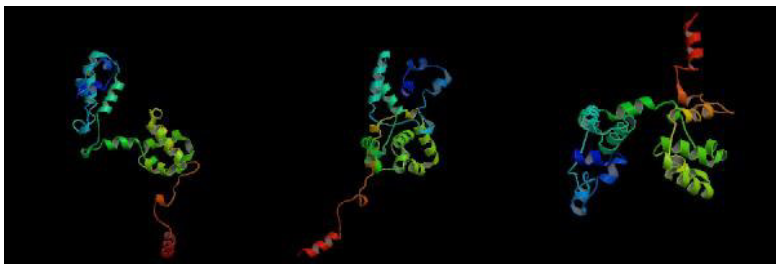


Figure 17. CCDC65 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

DRC1/CCDC164 (OMIM: 615288)

DRC1 dynein regulatory complex subunit 1 (DRC1) 17 exons gene (NM_145038.4, Gene ID: 92749) maps on chromosome 2p23.3. This gene encodes a central component of the nexin-dynein complex (N-DRC) (NP_659475.2 740aa). CCDC164 is primarily expressed in the lung, brain and prostate. Whole-genome and linkage studies in several PCD families identified a positional candidate region on chromosome 2 (13 Mb; Zm-1.20) that contained the *CCDC164* gene. CCDC164 homozygous nonsense mutation, heterozigous in both carrier parents were found in PCD affected individuals. TEM and high-resolution immunofluorescence microscopy showed that only a few cilia present unspecific ultrastructural alterations. The ultrastructures of the outer and inner dynein-arm structures appear normal, which was confirmed by immunofluorescence using antibodies specific for the outer-arm (DNAH5) and inner-arm (DNALI1). However, careful analysis of the cross-sections identified alterations in the N-DRCs showing that nexin links are totally missing in *CCDC164*-mutant respiratory cilia. These findings indicate that *CCDC164* deficiency in humans disrupts assembly of the N-DRC, resulting in changes that resemble those observed in *C. Reinhardtii* pf3 mutants [139].

GAS8/DRC4 (OMIM: 605178)

The growth arrest-specific 8 gene (Gene ID: 2622), a putative tumor suppressor, maps on chromosome 16q24.3. Its three alternative transcripts (NM_001481.2, NM_001286205.1, NM_001286208.1, NM_001286209.1) encode three different DRC4 protein isoforms (NP_001472.1, 478aa isoform a; NP_001273134.1, 395aa isoform b NP_001273137.1, 286aa isoform c; NP_001273138.1, 453aa isoform d). GAS8 recessive loss-of-function mutations, identified by genome-wide exome sequence analysis, result in PCD. GAS8-deficient cilia in PCD patients were found to completely lack LRRC48 (DRC3). Moreover CCDC164 (DRC1) and CCDC65 (DRC2) mutant cilia lack the axonemal GAS8 (DRC4) and present a TEM evident axonemal disorganization that cannot assemble LRRC48 in ciliary axonemes. This finding suggested that GAS8 (DRC4) is part of the N-DRC complex and that its function is essential for axonemal localization of LRRC48. The multimeric N-DRC protein complex does not properly assemble when one of the DRC proteins CCDC164 (DRC1), CCDC65 (DRC2), or GAS8 (DRC4) is mutated [140].

DNAAF2/KTU (OMIM: 612517)

Dynein Axonemal Assembly Factor 2 gene (DNAAF2, previously named 'Chromosome 14 open reading frame 104 ' (*C14orf104,* Gene ID: 55172) maps on chromosome 14q21.3 and encodes the highly conserved PIH1 Family (PF08190) Kintoun protein (KTU_HUMAN, Q9NVR5). The gene encodes for two KTU protein isoforms (NP_060609.2, GI: 145580588; NP_001077377.1, GI: 145580586). KTU is a cytoplasmic 837aa (91114 Da) protein, interacting with Dynein Axonemal Intermediate Chain 2 (DNAI2) and Heat Shock 70kDa Protein 1A (HSPA1A). These are required for cytoplasmic pre-assembly of axonemal dyneins, thereby playing a central role in motility in cilia and flagella. KTU is involved in pre-assembly of dynein arm complexes in the cytoplasm before intraflagellar transport loads them for the ciliary compartment. Defects in DNAAF2 cause primary ciliary dyskinesia type 10 (CILD10) [MIM: 612518] and Kartagener syndrome.

When Ktu (or its orthologous PF13) is absent, both outer and inner dynein arms are missed or defective in the axoneme, leading to a loss of motility [65].

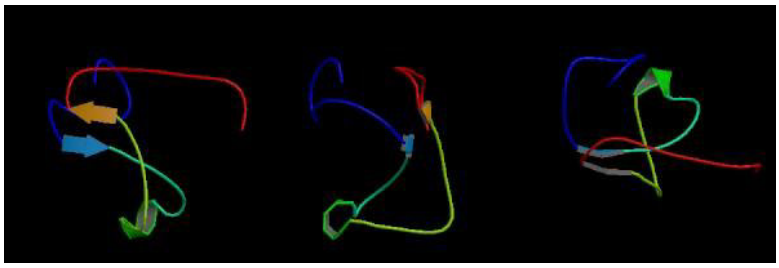


Figure 18. DNAAF2 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

DNAAF1 /LRR50 (OMIM: 613193)

Dynein axonemal assembly factor 1 gene (Gene ID: 123872), located in chr14q21.3, encodes a highly conserved protein involved in the preassembly of dynein arm complexes, that contains leucin – reach-repeat domains. LRR50 is expressed in human adult trachea and testis. Alterations in this gene are related to PCD (CILD1, OMIM 244400) and Kartagener Syndrome (CILD1, OMIM 244400).

Multiple transcript variants encoding different isoforms have been found for this gene: isoform 1 (NM_018139.2, NP_060609.2) which is longer and isoform 2 (NM_001083908.1, NP_001077377.1) which lacks an alternate in-frame exon and results in a shorter protein. Point mutations involving LRRC50 were found to impair assembly of the ODAs and IDAs. Functional analyses showed that LRRC50 deficiency results in immotile cilia. This immotility probably derives from the non-proper assembly of DNAH5- and DNAI2-containing ODA complexes and DNALI1-containing IDA complexes [67].

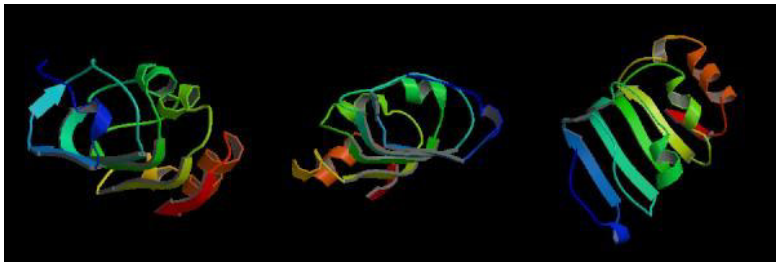


Figure 19. DNAAF1 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

DNAAF3/C19orf51 (OMIM: 6145665)

The Dynein Axonemal Assembly Factor 3 (DNAAF3, previously 'chromosome 19 open reading frame 51' C19orf51, Gene ID: 352909) gene maps to chr19q13.4. The cytoplasmatic protein (DNAAF3_HUMAN, Q8N9W5) encoded by this gene is required for the assembly of outer and inner dynein arms along the entire length of the axoneme (comprising the DNALI1 containing IDA types) and it is involved in preassembly of dyneins into complexes before their transport into cilia [69]. Defects in this gene are a cause of primary ciliary dyskinesia type 2 (CILD2, OMIM: 606763) and are associated with situs inversus and male infertility [69]. Four transcript variants (NM_001256714.1; NM_001256715.1; NM_001256716.1; NM_178837.4) encoding 4 different isoforms (NP_001243643; NP_001243644; NP_001243645; NP_849159 respectively) have been found for this gene. Immunofluorescent analysis of respiratory epithelial cells obtained by nasal-brush biopsy revealed that DNAH5, DNAH9 and DNAI2 (ODA components), and IDA light chain DNALI1 were all absent from the cilia of patients with mutations in DNAAF3. Electron microscopy TEM confirmed an abnormal ultrastructure missing ODAs and IDAs.

CCDC103 (OMIM: 614677)

The Coiled-coil domain-containing protein 103 gene (Gene ID: 300389) encodes a protein (CCDC103_HUMAN, Q8IW40) that functions like a dynein-attachment factor and is required for cilia motility. This gene maps to 17q21.31 and belongs to the CCDC103/PR46b family. Alternate splicing results in three different transcript variants (NM_213607.2, GI: 385719200; NM_001258395.1, GI: 385719201; NM_001258396.1, GI: 385719203) encoding for three different 242aa isoforms (NP_9988772.1, GI: 47106052; NP_001245324.1, GI: 385719202; NP_001245325.1, GI: 385719204) where the variants 2 and 3 differs in the 5'UTR compared to variant 1. Defects in CCDC103 are the cause of primary ciliary dyskinesia type 17 (CILD17 OMIM: 614679). The phenotype of all individuals carrying *CCDC103* mutations is characterized by typical clinical symptoms of PCD. TEM analysis of patients showed either loss or strong reduction in outer dynein arms (ODAs) structure. The high-speed videomicroscopy of respiratory cells from nasal brushing biopsy exhibits cilia paralysis with p.Gly128fs25 mutation and either reduced cilia beat amplitude or loss of beat coordination or cilia paralysis for p.His154Pro [129].

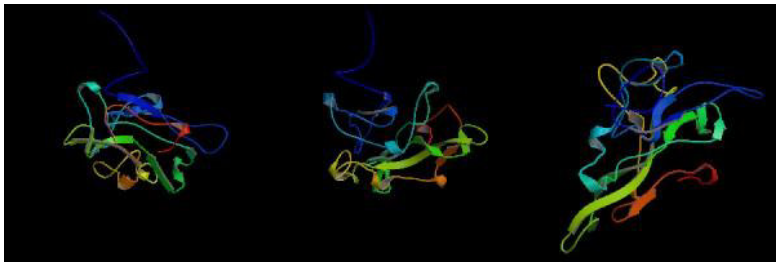


Figure 20. CCDC103 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

HEATR2 (OMIM: 614864)

The HEAT Repeat Containing 2 (HEATR2) gene maps on chromosome 7p22.3. The transcript (NM_017802.3, GI: 157388903) encodes for a cytoplasmatic protein (HEAT2_HUMAN, Q86Y56; NP_060272.3, GI: 157388904). Pathological mutations were first detected by Whole Exome Sequencing. Expression and functional cross-species studies suggested that HEATR2 protein is a highly conserved protein with an essential function in dynein arm transport or assembly. It was detected only in the cytoplasm of ciliated airways epithelial and lungs cells, disposed in a punctate pattern; precise mode of action is still unknown. Inferior turbinate cells from patients presented immotile cilia. RNAi silencing of HEATR2 expression caused asbsence of ODAs and truncated or absent IDAs [168, 169].

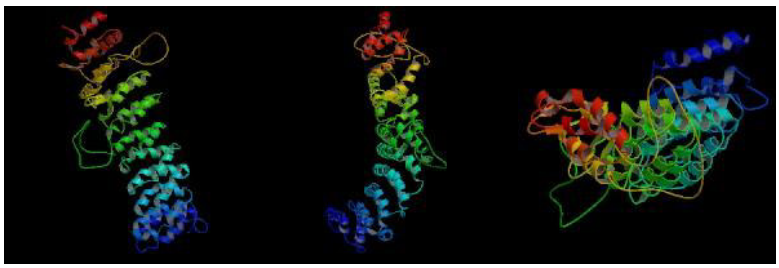


Figure 21. HEATR2 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

LRRC6 (OMIM: 614930)

The Leucine Rich Repeat Containing 6 (LRRC6) gene (Gene ID: 23639) maps on chromosome 8q24.22. Its transcript (NM_012472.4) encodes for a leucine-rich-repeat (LRR)-containing protein (TILB_HUMAN, Q86X45; NP_036604.2) mainly expressed in testis and respiratory cells cytoplasm. Mutations in LRRC6 cause PCD with infertility and laterality defects. TEM data on nasal brushing cells showed the loss of both ODAs and IDAs. LRRC6 is considered to be a regulatory protein crucial for the assemblage of cilia structural proteins. Although the precise LRRC6 action mechanism remains unclear, that protein seems to have a key role in embryonic nodal cilia and it is hypothesized to be part of a chaperone complex with DNAAF1, DNAAF2 and DNAAF3 [170].

C21orf59 (OMIM: 6154949)

Homo sapiens chromosome 21 open reading frame 59 (C21orf59) 7 exons gene (Gene ID: 56683) maps on chr21q22.1. It encodes (NM_021254.2) for a 290aa protein (NP_067077.1) that localizes in the cytoplasm and that seems to play a critical role in dynein arm assembly and motile cilia function. In c21orf59 zebrafish and planaria knockdown the outer dynein arm assembly was blocked. Moreover FBB18, the C21orf59 ortholog in *C. reinhardtii*, is a flagellar matrix protein that accumulates specifically when cilia motility is impaired. Recessive truncating mutations of C21orf59 were found in four PCD families. Similar to findings in zebrafish and planaria, mutations in C21orf59 caused loss of both outer and inner dynein arm components, suggesting its essential role in dynein arm assembly [132].

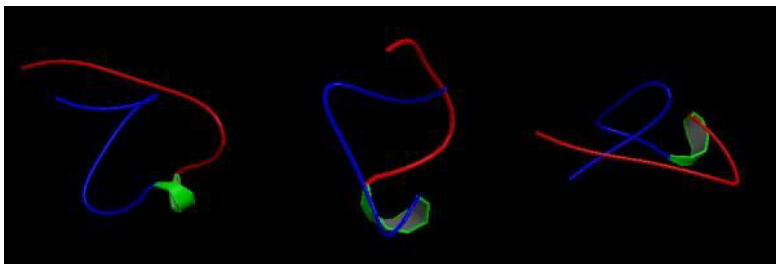


Figure 22. C21orf59 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

DYX1C1 (OMIM: 608706)

Homo sapiens dyslexia susceptibility 1 candidate 1 (DYX1C1) is an 11 exons gene (Gene ID: 161582) mapping on 15q21.3. Three different transcripts (NM_130810.3, NM_001033559.2, NM_001033560.1) encode for three different protein isoforms (NP_570722.2, 420aa isoform a; NP_001028731.1, 376aa isoform b; NP_001028732.1, 381aa isoform c). DYX1C1 is a tetratricopeptide repeat domain-containing protein that localizes both in nucleus and in cytoplasm and that is expressed in several tissues, including brain, lung, kidney and testis. In the brain it localizes to a fraction of cortical neurons and white matter glial cells. This protein also interacts with estrogen receptors and the heat shock proteins Hsp70 and Hsp90. It is involved in neuronal migration during development of the cerebral neocortex and in the estrogenic regulation of neuronal differentiation, survival and plasticity.

DYX1C1 was found also to be an axonemal dynein assembly factor required for ciliary motility. Moreover Dyx1c1 exon2-4 deletion in mice was found to cause a phenotype resembling PCD and morpholino targeting dyx1c1 in zebrafish to produce laterality and ciliary motility defects. Ultrastructural and immunofluorescence analyses of DYX1C1 in PCD patients showed the disruption of outer and inner dynein arms. Functional studies localized DYX1C1 to the cytoplasm of respiratory epithelial cells. Its interactome is enriched by molecular chaperones, and it interacts with the cytoplasmic ODA and IDA assembly factor DNAAF2 (KTU) [130].

SPAG1 (OMIM: 603395)

The human Sperm Associated Antigen 1 (*SPAG1*) gene maps to 8q22.2.

The two alternative transcripts encode for two protein isoforms (NP_003105.2, NP_757367.1, GI: 6674) which contain nine tetratricopeptide repeat motifs (TPRs) and are involved in protein-protein interactions and assembly of multiprotein complexes. Mutations affect the longer transcript that is the one important to ciliated cells function. Patients with *SPAG1* mutations present with typical PCD symptoms, including situs abnormalities and low nNO. TEM and IF data from patient airways epithelial cells revealed that SPAG1 mutations are always associated with defects in both ODAs and IDAs [131].

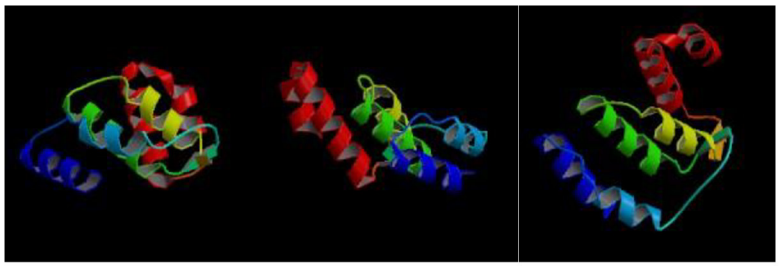


Figure 23. SPAG1 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

ZMYND10 (OMIM: 607070)

Zinc Finger, MYND-Type Containing 10 (ZMYND10) gene (Gene ID: 51364) maps on chromosome 3p21.3. Its two transcripts (NM_015896.3; NM_001308379.1) encode for two protein isoforms (NP_056980.2; NP_001295308.1). Mutations in ZMYND10 are associated with sterility and PCD. TEM data from patients revealed total dynein arms loss in respiratory cells and sperm flagella with partial absence of dynein arm complexes. Furthermore high-speed videomicroscopy shows cilia immotility [134]. ZMYND10 is probably involved in axonemal assembly of inner and outer dynein arms and it is required both for proper axoneme building and for motile ciliary function. It localizes in cytoplasm, cytoskeleton, microtubule organizing center, centrosome and centriolar satellite and it interacts with LRRC6.

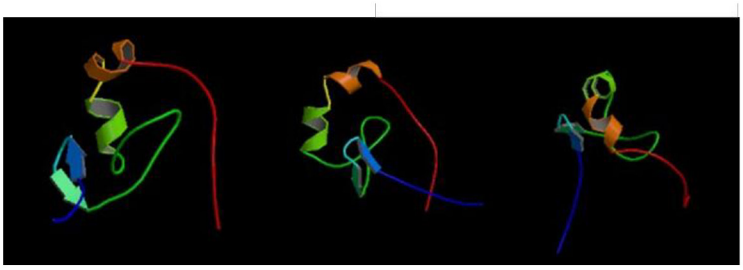


Figure 24. ZMYND10 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

CCDC151 (OMIM: 615956)

CCDC151 coiled-coil domain containing 151 (CCDC151) 14 exons gene (Gene ID: 115948) maps on chromosome 19p13.2. Three different transcripts (NM_001302453.1; NM_001302453.1; NM_001302454.1) account for three different protein isoforms (NP_659482.3 595aa isoform 1; NP_001289382.1 541aa isoform 2; NP_001289383.1 535aa isoform 3). The CCDC151 protein functions in outer dynein arm assembly and is required for motile cilia function [171].

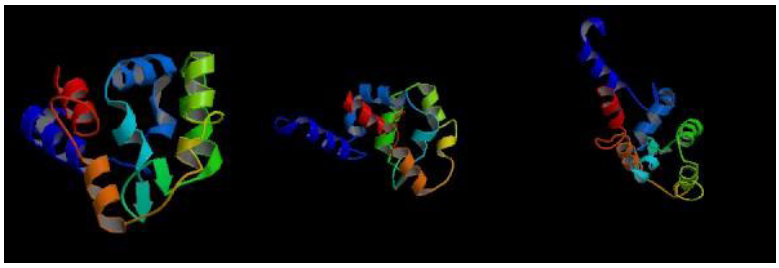


Figure 25. Protein CCDC151 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

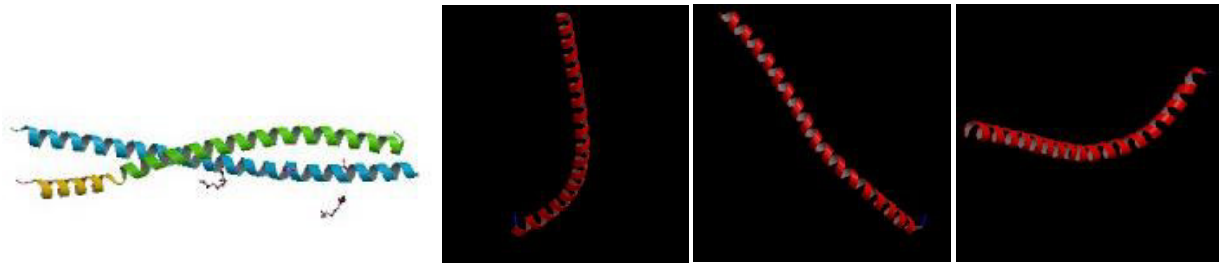
CCNO (OMIM: 607752)

Homo sapiens cyclin O (CCNO) is a three exons gene (Gene ID: 10309) mapping on chr5q11.2. Its transcript (NM_021147.4) encodes the 350aa cyclin-O protein (NP_066970.3), belonging to the cyclin protein family, involved in regulation of the cell cycle. CCNO mutated respiratory epithelial cells showed a marked reduction in the number of multiple motile cilia (MMC) on the cell surface. Subcellular analysis and in vitro ciliogenesis experiments revealed that CCNO is a cytoplasmic protein, expressed in respiratory cells, acting downstream of multicilin (MCIDAS), which governs the generation of multiciliated cells, to promote mother centriole amplification and maturation in preparation for apical docking [144].

Figure 26. CCNO protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

MCIDAS (OMIM: 614086)

The Homo sapiens multiciliate differentiation and DNA synthesis associated cell cycle gene (MCIDAS) is splitted in 7 exons (Gene ID: 345643) and maps on Chr5q11.2. Studies in *X. laevis* showed that the recessive loss-of-function and missense mutations in MCIDAS-encoding Multicilin promote the early steps of multiciliated cell differentiation. MCIDAS mutant respiratory epithelial cells carry only one or two cilia per cell, which lack ciliary motility-related proteins (DNAH5; CCDC39) as seen in primary ciliary dyskinesia. Consistent with this finding, FOXJ1-regulating axonemal motor protein expression is absent in respiratory cells of MCIDAS mutant individuals. CCNO is absent in MCIDAS mutant respiratory cells, consistent with its downstream activity. Therefore Multicilin was identifiable as a key regulator of CCNO/FOXJ1 for human multiciliated cell differentiation [145]. The transcript (NM_001190787.1) encodes the multicilin protein, a transcription regulator specifically required for multiciliate cell differentiation. It localizes in the nucleus and acts in a multiprotein complex that binds and activates genes required for centriole biogenesis. Moreover it is involved in mitotic cell cycle progression by promoting cell cycle exit and it can form homodimers *in vitro*, via coiled-coil domain.



A. **B.**

Figure 27. A.MCIDAS heterodimer protein 3D structure. **B.** MCIDAS front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

OFD1 (OMIM: 300170)

Homo sapiens oral-facial-digital syndrome 1 gene (OFD1) is divided in 23 exons and maps on chrXp22. Its transcript (NM_003611.2) encodes a centrosomal protein (NP_003602.1) that localizes in cytoplasm, cytoskeleton, microtubule organizing center, centrosome, centriole and cilium basal body and is involved in the biogenesis of the cilium, a centriole-associated function. Moreover it plays an important role in development by regulating Wnt signaling and the specification of the left-right axis. The OFD1 mutations effects were studied in mouse, in which the gene is also located on the X chromosome, but not subject to X inactivation. Mutations in this gene are associated with oral-facial-digital syndrome type I and Simpson-Golabi-Behmel syndrome type 2. In a large family a novel X-linked recessive mental retardation (XLMR) syndrome comprising macrocephaly and ciliary dysfunction was found to co-segregate with a frameshift mutation in the OFD1 gene. Respiratory problems and impaired ciliary motility were detected in these patients [147]. Ofd1 may be required for cilia assembly, probably by participation in intrafagellar transport.

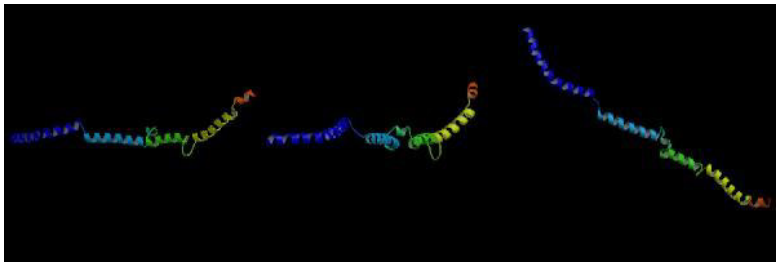


Figure 28. OFD1 protein 3D structure. Front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

RPGR (OMIM: 312610)

Homo sapiens retinitis pigmentosa GTPase regulator (*RPGR*) 19 exons gene (GI: 6103) maps on chrXp21.1. Its transcript (NM_000328.2) undergoes a multiple alternative splicing to produce several isoforms. It encodes a protein (NP_000319.1) with a series of six RCC1-like domains (RLDs), characteristic of the highly conserved guanine nucleotide exchange factors. The encoded protein is found in the Golgi body and interacts with RPGRIP1. This protein localizes into the outer segment of rod photoreceptors and cone outer segments and it is essential for their integrity. Mutations in this gene have been associated with X-linked retinitis pigmentosa (XLRP) and to a form of PCD with retinitis pigmentosa. In fact it is believed to be involved in cilia formation by regulating actin stress filaments and cell contractility. It is expressed in heart, brain, placenta, lung, liver, muscle, kidney, retina, pancreas and fetal retinal pigment epithelium. Isoform 3 is found only in the retina [146].

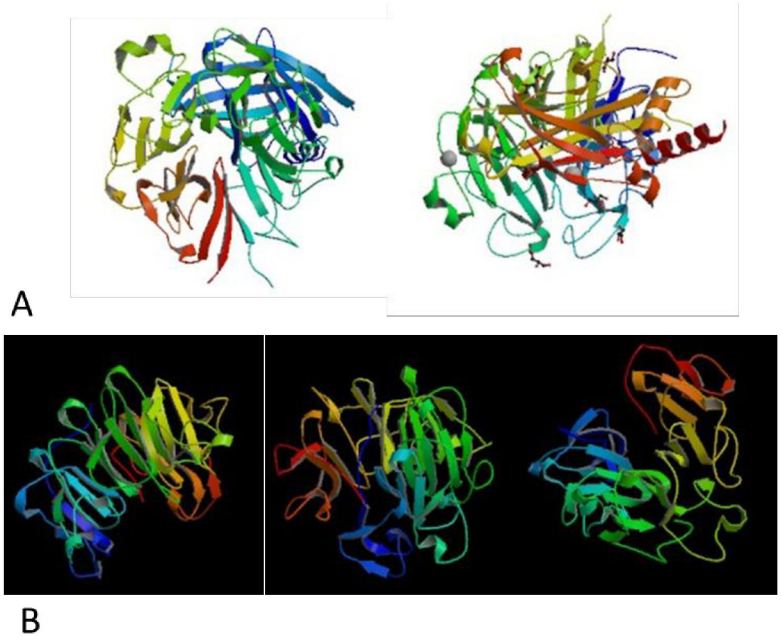


Figure 29. A.RPGR protein 3D structure. **B.** RPGR front, side and top view. Pictures are based on ModBase database prediction (http://modbase.compbio.ucsf.edu/).

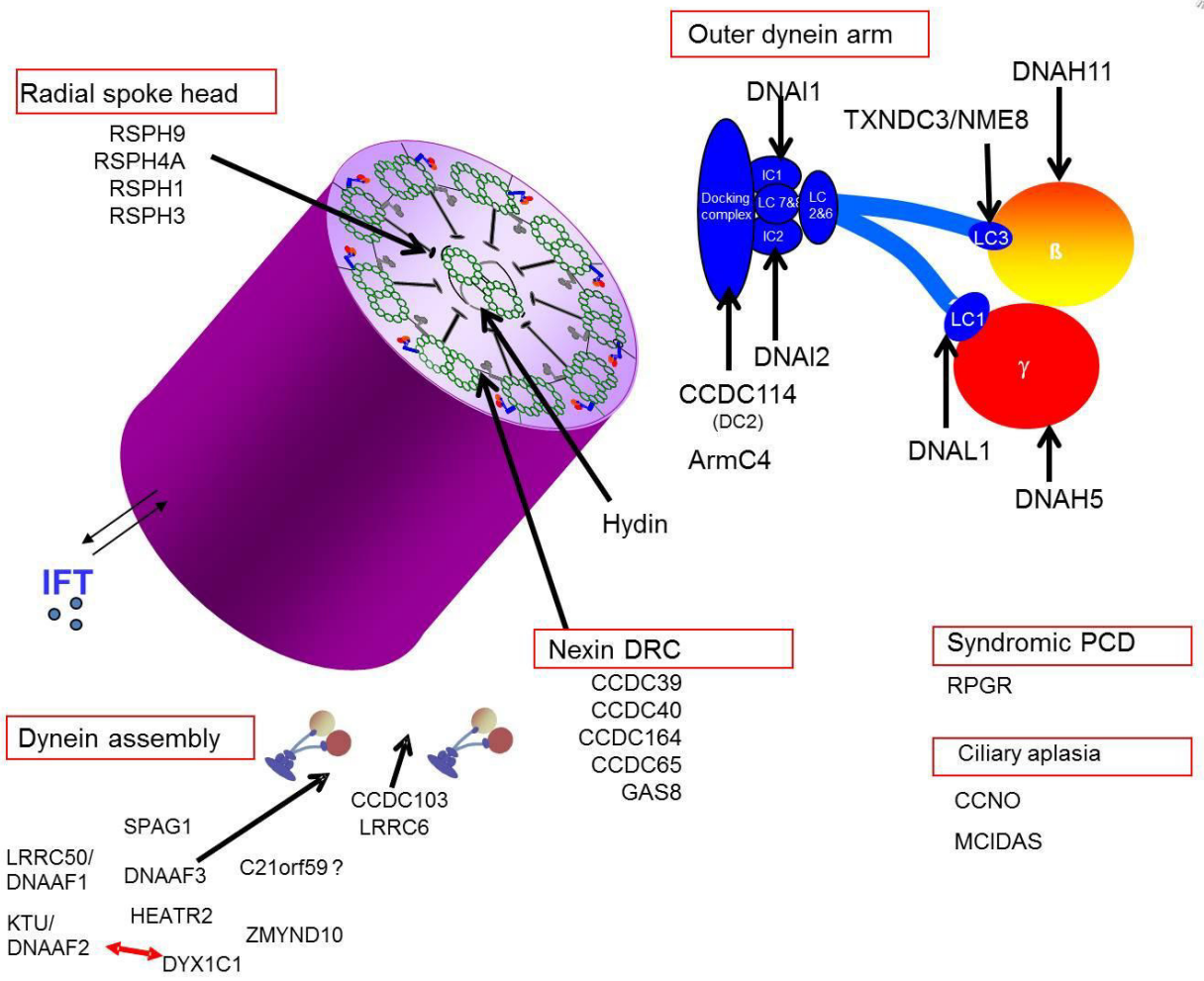


Figure 30. PCD protein involvement in cilia ultrastructure and assembly [Picture Copyright A.Shoemark and colleagues].

Next-Generation Sequencing

The next-generation sequencing (NGS) refers to second-generation sequencing methods, with a cheaper, faster and higher throughput than the chain-termination method developed in 1975 by Edward Sanger [172, 173]. NGS platforms perform massively parallel sequencing: millions of fragments of the same DNA sample are simultaneously sequenced. By this way an entire genome can be sequenced in less than one day. NGS, like each other technique, presents both advantages and limits.

Advantages

-It provides a cheaper and higher throughput alternative to traditional Sanger sequencing.

-It makes easier the discovery of genes and regulatory elements associated with diseases.

Limits

-It is still too expensive.

-Data-analysis is very hard and time-consuming because of the enormous volumes of information and bioinformatics knowledge is required.

-Many variants are still false positive.

The three main applications of NGS are:

1_Whole Genome Sequencing (WGS)

The Whole-Genome Sequencing (WGS) provides a high-resolution base-by-base view of the entire genome. It produces large volumes of data in a relative short amount of time, detecting single nucleotide variants, insertions/deletions, copy number changes, and large structural variants in introns, exons and intergenic DNA.

2_Whole Exome sequencing (WES)

The exome sequencing takes into account only the protein-coding regions, just over 1% of the genome. Although mutation events can occur in gene-coding or control/non-coding regions, WES permits a more cost-effective sequencing and it is a good tool for both gene discovery research and the genetic characterization of disease-phenotypes where the exact genetic cause is not known [174]. A selection of target DNA is required.

3_Targeted sequencing

This is used in cases of suspected disease or identified condition to target-sequence selected genes or genomic regions. Targeted sequencing seems to produce a higher coverage of genomic regions of interest, reducing cost and time and allowing the rapid diagnosis of many genetic diseases. Se-

quencing panels that target hundreds of genomic regions thought to contain disease-causing mutations are often used by clinical researchers.

Variants management and scores

Exome and genome sequencing are now used in disease genetics to increase the power of previous genetic approaches such as linkage analysis and genome-wide association studies (GWASs). But as data acquisition becomes easier, interpretation becomes harder. The challenge presented by 'next-generation genetics' is to identify causal variants in a multitude of detected, filtered and candidated variants [175].

A priori estimation of variant causality is very different from its *a posteriori* functional confirmation, therefore strong convincing data are required to support the hypothesis that a given candidate variant is truly causal. The estimation of the phenotypical importance of any given variant is helped by computational databases and softwares that provide predictions and assessments of variant function.
Rare (frequency < 0.01%) non-synonymous variants are the preferred candidates, whereas synonymous or non-coding variants are usually ignored.

Deleteriousness is also estimated by using an evolution approach: sequences that are common across different species and that have not been removed by natural selection, are probably coding for important proteins and physiological functions. Conserved positions and domains that evolved too slowly to be non significant are identified by quantifying evolutionary changes in genes or genomes. However, sequence conservation is not a deleteriousness predictor *per se* and sometimes the same locus doesn't account for the same biology function in different species, lowering the power of this approach.
Nonsense and frameshift mutations, predicted to impact on protein function, are considered the best disease-causal candidates. However sometimes mutated proteins are still functional and a protein loss does not cause disease. Multivariate Analysis of Protein Polymorphism (MAPP) method [176]), SIFT and polyPhen [177-179] are examples of severity predictions based on protein amino-acidic and nucleotidic sequences.

The homologous proteins alignments used to estimate the sequence conservation level, together with biochemical data, including amino acid properties (such as charge), sequence information (such as the presence of a binding site) and structural information (such as the presence of a β-sheet) can significantly improve deleteriousness prediction accuracy. An example of methods to infer nucleotide-level constraints in genomic sequence alignments is Genomic Evolutionary Rate

Profiling (GERP) [180, 181]. This method considers base by base and assigns the highest scores to the highest conserved sites with fewer substitution events receiving higher scores.
Experimental analyses are generally performed to add support to a limited number of candidate causal variants identified using other information which can provide powerful support of causality for a given phenotype. For example, this may consist of the in vitro demonstration of molecular consequences (such as disruption of expression or protein folding) or the in vivo recapitulation of the human phenotype in a model organism.

It is important to recognize that experimental predictions, as with computational predictions, are informative but often not definitive. For negative results, an inevitable concern is whether the experiment was performed in the appropriate context, including: genomic context (flanking sequence or chromatin state); developmental context (cell type or developmental stage); or organismal context (functional consequences of a mutation that are species-specific). Conversely, positive experimental results do not necessarily establish a better understanding of the relative value of Coding versus non-coding variants, evolutionary annotations versus functional annotations, and the development of a unified framework, accounting for both computational and experimental sources of information, will be important for accurate assessments of deleteriousness.

Furthermore, a large collection of true positive (deleterious) and true negative (neutral) mutations, the improvement of databases resources such as Online Mendelian Inheritance in Man (OMIM), Swiss-Prot [182] and the Human Gene Mutation Database (HGMD, [183, 184]) and the knowledge of interactions between individual variants and between variants and the environment are clearly relevant to accurate genotype–phenotype predictions.

Although in this case comparative genomics becomes fundamental in deleteriousness prediction, the phylogenetic is trickier for non-coding sequences than for proteins.
It is to be noted that the majority of human genetic mutation are non-coding variants and that GWAS studies showed that ~88% of trait-associated variants of weak effect are non-coding [185].

WES in human genetics

The fast acceleration of whole genome sequencing of patients and healthy individuals increases the possibility of getting prospective information on patient genomes. Since Mendelian pathological mutations have to be identified in the background of the entire genome, it is mandatory to find predictive scores of mutations impact on phenotype. Interestingly each disease, PCD included, can be expressed in a big variety of phenotypes and mutations in the same gene can also give rise to radically different clinical symptoms [84].
The heterogeneous and the digenic/poligenic models attempt to explain this phenotype variability by calling on the presence of many different causative genes or the interaction between two or more genes to explain the clinical symptoms. By using the knowledge of mutants-gene-encoded proteins function and reciprocal interactions with variable penetrance, hypotheses are put forward on mutations role in disease raising and evolution. Unfortunately the activity of alleles at a single locus or in pairwise relationships can't be considered free from the genetics background of the genome and epigenome, where pretending to describe accurately such relationships is an oversimplification of biological reality.

There is some evident bias in human genetics, due to the difficult interpretetion of certain clinical genotypes and to the impossibility of providing structural or functional evidence in mutated proteins suggested to be disease-causative. Some patient show mutations not sufficient to cause disease (at least from a theoretical point of view). Moreover bona fide pathogenic heterozygous mutations at a single locus, and intronic, splicing and UTRs are under reported.

On one hand whole genome studies will help to produce a more complete overview of the genome, but on the other they will accentuate any interpretive challenge by suggesting the presence of many genes-interactions shades in contrast to the classical difference between causal or non-causal mutations.

In the last years Whole Exome Sequenced became usually performed in PCD patients to quickly detect mutations in known-genes and to eventually search for new genes. (e.g. [149, 150, 186-190]).

The ciliopathies represent a good example for studying by NGS their genetic complexity. In fact, up to date, a considerable number of proteins required for cilium structure, assembly, maintainance and function are known [39, 119].

Moreover the cilium represents a semi-closed system that can be studied in its (near) entirety. Its subcellular functions have been intensively studied and combined with biology and biochemistry, allowing observation of this organelle expression and functionality in cells and in whole organisms.

Therefore the functional validation on mutations that affect ciliary gene/protein is now possible by studies on model mutants and cell cultures.

To increase the success of this kind of analysis it is necessary to get a precise functional and structural profile of the proteins required for ciliary biogenesis and function in the context of different cell types [191]. A great effort could be made by the generation of a cilio-proteins context-specific network containing both developing phenotype and gene mutations information from patients [84].

Material and Methods

Patients and Sample Collection

Patients DNA sample collection is ongoing since 1996.
Several european phisicians, after patients consent, sent to the Medical Genetics laboratory of the University Medical School in Geneva (Prof. S.E. Antonarakis) fresh blood or DNAs samples of patients for who they had hypthized PCD diagnosis, together with their phenotype description and the results of the clinical tests that they performed to confirm PCD diagnosis.
The recruitment effort to identify new patients and/or families with more than one PCD affected was also maintained through the Swiss Registries for Interstitial and Orphan Lung Diseases (SIOLD).
To date the Geneva cohort of PCD patients consists of DNAs from 539 individuals belonging to 239 families.
The recruitement is ongoing.

DNA extraction

The DNA extraction was performed in column (NucliSENS®) or by authomated platforms such as Arrow Autogen and easyMag®. However the oldest samples or the bloods not in excellent conditions were manually extracted.
5ml of blood and 15 ml of Red Cell Lysis solution are gently mixed for 10 minutes. RCL Solution is 20mM Tris HCl at pH 7.6.
After 10' at 3500 rpm (875xg) centrifugation the surnatant is discarded. The pellet is resuspended by vortexing and 5 ml of White Cell Lysis solution are added and the sample is vortexed to be sure of the complete cells lysis. WCL solution is obtained by mixing 10mM Tris Hcl pH 8, 1mM EDTA and 0.1% SDS.
After 37°C incubation for 30 minutes, 2.6 ml of Protein Precipitation Solution are added and all is strongly vortexed for 25 seconds. A13000 rpm (12000xg) centrifugation is performed for 10'. The surnatant that contains the DNA is collected in a 50ml tube. 2 volumes of 100% ethanol are added at room temperature and the tube is mixed until the DNA appears as a filamentous. Ethanol is removed after a 5' 3500 rpm (875xg) that allows DNA precipitation. 3ml of 70% ethanol at 4°C are added and then removed after a 5' 3500 rpm (875xg). The DNA is resuspended in ddH2O and conserved at +4°C.

DNA quantity and quality

3µg of DNA are required for WES and 2 µg for HaloPlex. The 260/280 ratio has to be of 1.7-2.0. The 260/230 ratio should be >2.
Nanodrop was used to a preassessment of DNA concentration and quality. The nanodrop uses the UV-absorbance method: a spectrophotometer measures the natural absorbance of light at 260 nm (for DNA and RNA) or 280 nm (for proteins). The more DNA, RNA or protein in the sample, the more light is absorbed.

Afterwards Qubit [1] was used to assess the precise concentration of selected DNA samples. The Qubit fluorometer uses fluorescent dyes to determine the concentration of nucleic acids and proteins in a sample. Each dye is specific for one type of molecule: DNA, RNA or protein. These dyes have extremely low fluorescence until they bind to their targets (DNA, RNA or protein). Upon binding, they become intensely fluorescent. Once added to a solution of DNA, the Qubit DNA dye binds to the DNA within seconds and reaches equilibrium in less than two minutes. At a specific amount of the dye, the amount of fluorescence signal from this mixture is directly proportional to the concentration of DNA in the solution. The Qubit fluorometer picks up this fluorescence signal and converts it into a DNA concentration measurement using DNA standards of known concentration. The Qubit dsDNA BR (Broad Range) Assay with a 2–1,000 ng assay ranger and a 100 pg/µl–1 µg/µl sample starting concentration range was used. 2µl of DNA are added to 198µl of BR DNA Qubit buffer solution and measured after a standard assessment prepared with 190µl of BR DNA Qubit buffer solution and 10µl of the two probes provided.

Thanks to the specificity of dyes for each type of molecule, by the use of Qubit it is possible to get a precise quantification of DNA even in the presence of other bio-molecules. Meanwhile it is to be noted that NanoDrop measurement is inaccurate. In fact the absorbance is a natural property of many other molecules than nucleic acids and proteins. Therefore these particles can contaminate the sample and result in false results. Moreover, the absorbance method does not distinguish between DNA, RNA, protein or free nucleotides or amino acids in the sample.

Whole Exome Sequencing (WES)

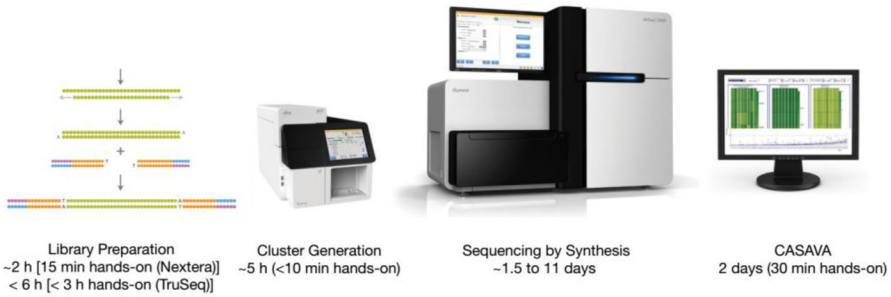


Figure 31. Whole Exome Sequencing Semplified [Picture Copyright: Illumina].

Todate DNAs of 46 families (56 affected and 9 non-affected individuals) have been entirely sequenced and bioinformatically analyzed for the whole exome.

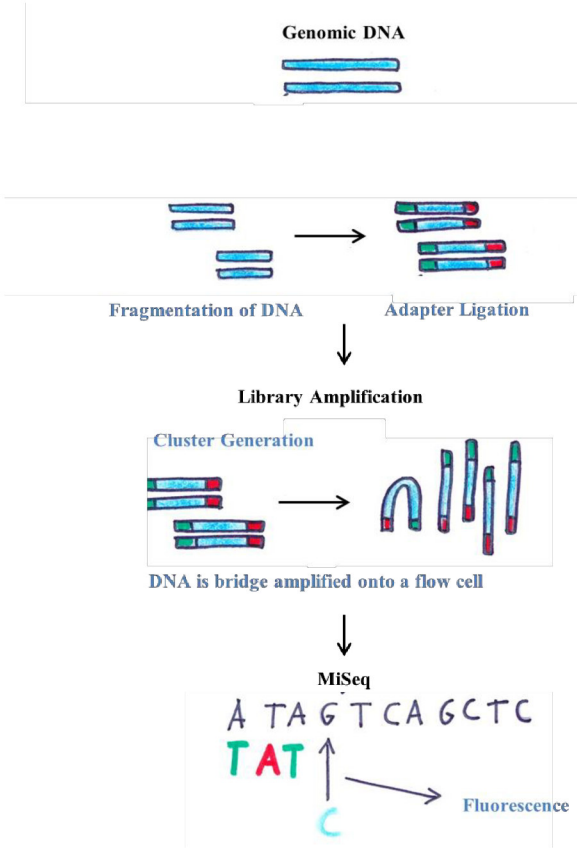


Figure 32. Template preparation for WES performed by Illumina cBot HiSeq 2000 [Picture Copyright C. Olcese].

cBot cell-flow Cluster Generation was used to perform the Whole Exome Sequencing.

This automated clonal amplification system is based on three steps:

1. Exome enrichment and libraries preparation from nucleic acid samples (Illumina Sureselect XT Target Enrichment System);
2. amplification to produce a clonal cluster from each single DNA molecule of the library;
3. sequence using massively parallel synthesis (HiSeq2000 sequencing system).

Illumina Sureselect XT Target Enrichment System

The Sureselect XT Target Enrichment System for Illumina Paired-End Sequencing Library [Sureselect XT Target Enrichment for Illumina Multiplexed Sequencing] was used.
All experiments were performed using the manufacturer's recommended protocols without modifications. Each Illumina version differs for probes. The fragmentation method (sonication) and the protocol [2] remain the same.
The Library preparation protocol proceeded with many subsequent steps :

1-Sonication: the DNA, is fragmented into fragments of 150 to 200 by the Covaris S-Series S220 Focused-ultrasonicator which is computer controlled by SonoLab™ Software. To obtain a good exome library 3ug of high-quality DNA are required. The sonication has to take place at 4-7°C and to last 6*60 seconds.

2-Purification by magnetic beads: the protocol steps are separated by one or more round of purification by Agencourt AMPsure XP beads. They allow the selection of fragments of interest by a length filter: they remove all fragments < 100 bp.

3-Agilent 2100 Bioanlyzer: the DNA quality is assessed by this chip-based electrophoresis to check the presence of a peak between 150 to 200 nucleotides-long fragments.

4-Ends Repair: the DNA fragments ends are repaired by using the DNA-polimerasi I from E.Coli, the T4 polymerase from phage, the polynucleotide-kinase T4, their buffer and a dNTP mix.
The repair step is followed by magnetic-beads based purification that filters for blunt-ends fragments by removing the sticky-ends ones.
The two polymerases' exonucleic activity removes the 3'-overhangs, meanwhile the polymerasic one fills the 5'-overhangs. At that point all fragments have 5'-phosporilated blunt-ends.

5-Adenine insertion at the 3'-end: an Adenine base is added at 3'end to increase the ligation efficiency: the oligonucleotide-adapters present, in fact, a T base at 3'end. A purification step is then performed.

6-paired-end adapter ligation: the DNA fragments are ligated to the Indexing-specific Paiered-End Adapter oligo by the DNA ligase. These adapters are complementary to the oligonucleotides anchored to the flow cell. The ligation product undergo a purification step.

7-adapter-ligated library amplification: a PCR reaction is performed on the previously obtained 'adapted' fragments. A purification step is then performed.

8- Agilent 2100 Bioanalyzer: the DNA quality is assessed to check the presence of fragments with a lenght between 250 to 275 nucleotides.

9-Hybridization: each DNA library is submitted to liquid phase capture against whole human exome. All coding sequence (about 200'000 exons of 20'000 genes to/for a total of 51.542.882 pb) are selected by SureSelect biotinylated RNA library 'baits' (Human All Exons v3 reagents, Agilent Inc®).

10-Hybridized fragments capture: the DNA strands hybridized with complementar biotinylated RNAs (Invitrogen Dynal MyOne Streptavidin T1 magnetic beads) are incubated with and captured by streptavidin coated magnetic beads. The unbound DNA factions are then discarded, the beads washed and the RNAs digested.

11-Library amplification: each library undergoes an Herculase II PCR using specific primers for the exomic sequences previously hybridized. These fragment-specific (Forward and Reverse) primers add a 6bp sequence that differs in each DNA sample (index-tags). A purification step is then performed.

12- Agilent 2100 Bioanlyzer: the DNA quality is assessed to check the presence of fragments with a lenght between 300 to 400 nucleotides.

13- The concentration (nM) of each index-tagged captured library is determined by the use of Agilent QPCR NGS Library Quantification kit or by Qubit. The libraries preparation is now ended.

14-Pool preparation for multiplexed sequencing: each index-tagged library has to be present in equimolar amounts in the pool.

Bridge Amplification

The template (the result of the Sureselect XT Target Enrichment for Illumina Multiplexed Sequencing) was denaturated and the cBot flow cell was prepared.

A pool of all libraries (each one with a specific tag and equimolarly present) plus 4ul of the PhiX control library was twice collected. Two eppendorf with the same content were charged each in one of the eight cBot [3] flow cells.

The cBot allows hundred of millions of singlestrand templates to be hybridized to a lawn of oligonucleotides covalently bound to the flow cell surface. At first the templates are copied, starting from the hybridized primers by 3' extension using an high-fidelity DNA polymerase. Then the original templates are denatured leaving the copies immobilized on the flow cell surface.

After that the bounded molecules are amplified by successive isothermal 'bridge amplification' to form clonal clusters each containing more than 1,000 copies of the starting molecule. The templates loop over to hybridize to adjacent lawn oligonucleotides, forming dsDNA bridges, which are denatured to form two ssDNA strands. These two strands loop over and hybridize to adjacent oligonucleotides and are extended again to form two new dsDNA. The process is repeated to create millions of individual, dense clonal clusters containing ~ 2000 molecules.

Each cluster of dsDNA is denaturated, the reverse strand is removed by specific base cleavage, leaving the forward DNA strand.

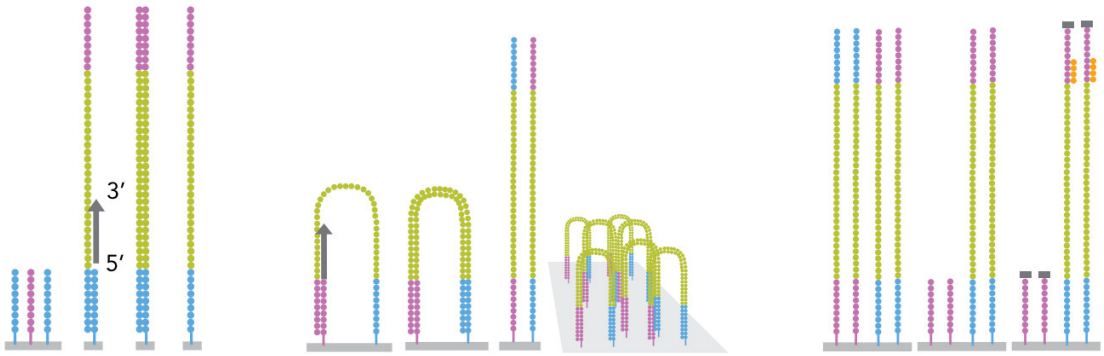


Figure 33. Isothermal Bridge Amplification. Cluster generation from single-molecule DNA templates occurs within the Illumina flow cell on the cBot instrument. It involves immobilization and 3' extension, bridge amplification, linearization and hybridization [Picture Copyright: Illumina].

Illumina HiSeq 2000

Massive parallel sequencing was performed in the Illumina HiSeq2000 high throughput apparatus available in Genetic Medicine and Development (CMU, Gènève). Illumina HiSeq2000 Genome Analyzer in 3 to 12 days generates 100-200 million fragments sequenced in parallel (7.5-35 Gbases data output). The read length is 100-200bp, paired end.

The 3'-ends of the DNA strands and flow cell bound oligonucleotides are blocked and the sequencing primer is hybridized to the complementary sequence on the illumina adapter on unbound ends of the templates in the clusters.

The polymerase incorporates fluorescence-labeled nucleotides which are 3' terminated and therefore elongation stops after just one nucleotide. Red and green lasers excite the fluorophores and four images are taken through different filters for each of the 100 tiles that the Genome Analyzer II has per lane. A washing step removes the phluorophores and the 3' terminators and a new elongation step begins. The more the cycles, the more the sequence lenght.
After sequencing, the images are integrated and extracted. The base caller software, Bustard, discriminate between A and C as well as G and T that have similar emission spectra, and recovers phasing and pre-phasing, noise due to incomplete removal of the 3' terminators and fluorophores, or incorporation of nucleotides without effective 3' terminators. Increasing the number of cycles, the fraction of sequences per cluster affected by phasing increases, hampering the identification of the correct base [192].

Illumina Sequencing Analysis Pipeline

Illumina pipline software for Illumina's data analysis transforms primary imaging output from the Genome Analyzer into meaningful results. It contains some integrated algorithms that perform the primary data transformation steps (image analysis, intensity scoring, base calling and alignement) to compose billion of bases of raw sequencing data into discrete aligned strings of bases.

Data are processed with Illumina pipeline version 1.7.0 (within SCS 2.8).

Two types of files are created containing the sequences (reads). The seq.txt file is a file containing only the sequence information. The qseq file contains both sequence and base by base quality information. In general only the seq.txt files are provided to users, but qseq files are available on request. Details regarding the Illumina Sequencing Analysis Software are provided by the Illumina Sequencing Analysis Software User Guide for Pipeline Version 1.3 [4].

Per-cycle BCL basecall files are the Illumina's primary sequencing output. bcl2fastq software [5]

convertes these BCL files to obtain the output raw data under files of short sequence reads (FastQ format) of 2x74 to 2x95 bp for both ends of each fragment of the libraries.

FastQ file format

A FASTQ file contains thousands of sequences. It normally uses four lines per sequence as follows:

line 1	@HWUSI-EAS490:1:1:1:910#0/1
line 2	ATTAAAATTGGGACTAGATATGACACCACTTAAGAAGGCACTGCCT
line 3	+
line 4	aaa^Z^\aaaaa_``_a]^Ya`^_\_^_\`_Y__PV]\`]_]]T]Z

Line 1 begins with a '@' character and is followed by an identification sequence (@SEQ_ID).

For Illumina:

@HWUSI-EAS490:1:1:1:910#0/1	
@HWUSI-EAS490:	the Illumina instrument name
1:	flowcell lane
1:	tile number within the flowcell lane
1:	'x'-coordinate of the cluster within the tile
910	'y'-coordinate of the cluster within the tile
#0	index number for a multiplexed sample (0 for no indexing)
/1	the member of a pair, /1 or /2 (paired-end or mate-pair reads)

Line 2 reports the raw sequence letters.

```
ATTAAAATTGGGACTAGATATGACACCACTTAAGAAGGCACTGCCT
```

Line 3 begins with a '+' character and is optionally followed by the same sequence identifier (and any description) again.

```
+
```

Line 4 encodes the quality values for the sequence in Line 2, and must contain the same number of symbols as letters in the sequence.

```
aaa^Z^\aaaaa_``_a]^Ya`^_\_^_\`_Y__PV]\`]_]]T]Z
```

The character '!' represents the lowest quality while '~' is the highest. In the exemple there are the quality value characters in left-to-right increasing order of quality (ASCII):

Custom designed Haloplex Libraries (Targeted NGS)

Agilent's SureDesign tool at www.agilent.com/genomics/suredesign was used to design a customer ciliome panel before ordering a HaloPlex Target Enrichment System Reagent Kit.
An Haloplex panel of 640 potentially related to PCD genes was created (gene list attached in appendix) to be run on Illumina platform. This is the summary of our custom panel:

Target Summary
640 Target IDs resolved to 663 targets comprising 10040 regions.
0 Target IDs were not found.
Region Size: 1.824 Mbp

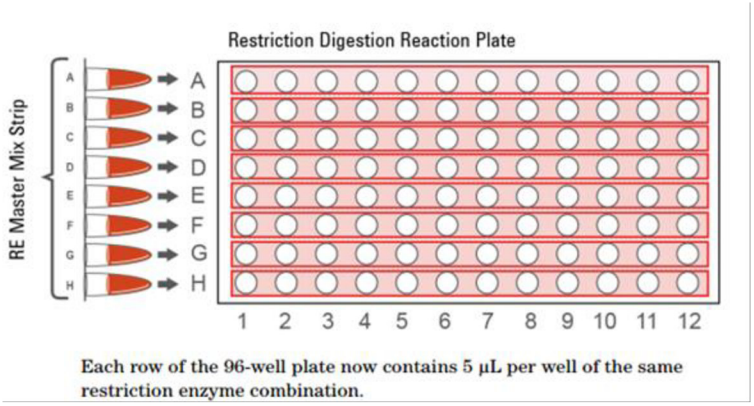
Amplicon Summary
Total Amplicons: 128774
Total Target Bases Analyzable: 1.79 Mbp
Total Sequenceable Design Size: 4.69 Mbp
Target Coverage: 98.05 %

Target Parameters
Databases: RefSeq, Ensembl, CCDS, Gencode, VEGA, SNP, CytoBand
Region: Coding Exons
Region Extension: 10 bases from 3' end and 10 bases from 5' end.
Allow Synonyms: Yes

#High Coverage: Number of regions where analyzable amplicon overlap >= 90%.
#Low Coverage: Number of regions where analyzable amplicon overlap < 90%

Figure 34. Detail of the 640 PCD genes Haloplex Design.

The libraries were prepared following the HaloPlex Target Enrichment System Library preparation protocol [6] that proceed with the following steps:
1- Genomic DNA digestion with restriction enzymes: gDNA samples are digested with 16 different restriction enzymes to create a library of gDNA restriction fragments. Successful enrichment requires samples containing 5µg of 5ng/µl high-quality DNA.



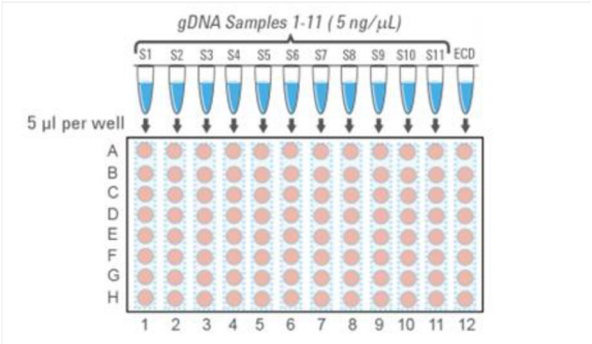


Figure 35. Each row of the 96-well plate will contain 5ul per well of the same restriction enzyme combination. Each column will be completed with 5µl per well of DNA samples (5ng/µl) [Picture Copyright: Illumina].

2- Agilent 2100 Bioanalyzer: The effectiveness of the DNA digestion is checked.

3- Digested DNA hybridization to HaloPlex probe for target enrichment and sample indexing: gDNA restriction fragments are hybridized to the HaloPlex probe capture library. HaloPlex probes are designed to hybridize selectively the fragments originating from target regions of the genome and to direct circularization of the targeted DNA fragments. During the hybridization process, sequencing motifs, including a barcode sequence, are incorporated into the targeted fragments.

4- Capture of target DNA. The circularized target DNA-probe hybrids, containing biotin, are captured on streptavidin beads.

5- Ligation of the captured, circularized fragments. DNA ligase is added to the capture reaction to close nicks in the circularized probe-target DNA hybrids.

6- PCR Master Mix preparation.

7- Elution of the captured DNA with NaOH.

8- PCR amplification of the captured target libraries.

9- Purification of the amplified target libraries. The amplified target DNA is purified using AMPure XP beads.

10- 2100 Bioanalyzer microfluidic analysis of enrichment and quantification of the enriched target DNA in each library sample.

11- Collection of a pool of samples with different indexes for multiplexed sequencing.

The two main differences between WES and Haloplex libraries preparation are that:

- Haloplex protocol includes only one purification step (against WES with a purification after each step);
- Haloplex DNA fragmentation is performed by restriction enzyme and not by sonication.

WES and Haloplex analysis

Putty interface

Putty [7] was used as a Unix-like interface to work on the Vital-IT platform that was created by the Swiss Bioinformatic Institute (SIB). SIB is a non-profit foundation recognised as a public utility which was established in 1998. It coordinates research and education in bioinformatics throughout Switzerland and provides high quality bioinformatics services to the national and international research community. Data are stored and analyzed in this network.

Pytlines

The Illumina's raw data outputs (.FastQ files) were analyzed using a custom bioinformatic pipeline developed on Python script in GEDEV department (Drs F. Santoni, S. Nikolaev, K. Popadin), which utilizes published algorithms in a sequential manner (BWA for mapping the reads, SAMtool for detection of the variants, Pindel for the detection of indels, ANNOVAR for the annotation of all sequence changes). The entire coding sequence of human genome corresponding to the RefSeq coding genes was used as the reference for the calculation of coverage and reads on target. All variant were annotated including functional prediction (Polyphen-2, SIFT, MutationTaster), allele frequency (dbSNP, 1000-genomes). The data are stored in the Geneva cluster of the VITAL-IT high performance computer resources (Prof. I. Xenarios, Dr. C. Iseli, SwissmInstitute of Bioinformatics).

The pytline run takes around 24 hours. Therefore the `nohup` and `&` commands were often used to continue to run the pytline in the remote server, even if the Putty interface was closed.

The informatics analysis of Illumina *.Fastq(.gz) outputs, proceeding through the selected sequence of pythlines, ends in the production of an Annovar annotate .csv file output.

```
> nohup python Illumina.pytline.GVA*.py &
[Illumina.pytline.GVA*.py
import sys
sys.path.append('/home/fsantoni/tools')
import __builtin__ as Param
from modules import *

##Personal Settings
Param.index                                                              =
'/scratch/cluster/monthly/fsantoni/index/hg19/hg19.unmasked'
Param.reference                                                          =
'/scratch/cluster/monthly/fsantoni/index/hg19/hg19.unmasked.f$
#Param.region =  ''

##Modules Settings
#Param.sampileup = {'q': 20, 'Q': 13}
Param.annovarsnpcall = {'coverage': 3, 'SNPquality': 10} ignorate le
spv in numero di meno di tre e con qualità inferior a 10

##Project variables
myDIR = "/scratch/ug/monthly/jlblouin/PCD/CHIARA_ANALYSES"
P = Project('FamilyGVA01',myDIR)
P.set_samples_dir('/home/jlblouin/scratch/PCD/CHIARA_ANALYSES/GVA01_
PCDexomefev$
P.pipeline = "bwa_aln_paired > bwa_sampe > sambam > samrmdup >
[pindel |sampile$
Programmi elaborazione dati della Pipeline standard per illumina

##Pipeline START point

P.run()]
```

Figure 36. e.g of pipline script**.**

Pytlines algorithms

0-bwa_aln_paired: The Burrows-Wheeler Aligner (BWA) alignement software for short sequences [8], permits the reads mapping versus the human reference genome. Genome sequence hg19GRCh37 release was used, even if in 2015 was released the new version hg38GRCh38. For each DNA fragment, made of two paired-end reads, sequence data are provided from both ends. The sequences are therefore stored in two separate files (one for the data from each end). At the end of this step a .sai file is obtained for each reads end [193].

1-bwa_sampe: The .sai files are converted in SAM files, technically human-readable. Bwa combines togheter the information from the two separate ends of each sequence. The mapping proceeds now by reads couples. The output is a *.SAM (Sequence Alignment Map) file. The .sam format permits the storage of the aligned reads. These reads can be then further analysed by using SAMtool options [194].

2-sambam: Samtool compresses the *.SAM file in the binary format *.BAM (Binary Alignment Map). In a *.BAM file each read is identified by its own label name, its sequence and its alignment locus on the reference genome. This file contains also information of alignment quality and coverage [9] [194].

3-samrmdup: SAMtools removes the duplicated reads in order to save only the reads that maps with an high quality to the reference genome. The duplicates are identical reads derived by the PCR performed during library preparation. Their presence could impair the sensibility and specificity of variants identification by altering their real relative frequency [195].

4-pindel | 5-sampileup: SAMtools and Pindel works simultaneously to analyse the reads and to find the variants. SAMtools identifies the Single Nucleotide Polimorphisms and nucleotidic insertions or deletions. Pindel detect breakpoints of large deletions, medium sized insertions, inversions, tandem duplications and other structural variants at single-based resolution [10] [196].

6-annovarsnpcall: It creates a single *.VCF (Variant Call Format) file containing all the variants by pindel and mpileup. It is the input used by ANNOVAR.

7-annovarannotate: The bioinformatics ANNOVAR tool [11] allows the functional annotation of the previously detected and selected genomic variants. It returns a *.CSV (comma-separated values) file, microsoft- excel compatible, that contains all the variants found in the sequenced part of the genome and also useful information on them (ANNOVAR [197]).

The Annovarannotate .csv, Microsoft Excel compatible, file reports for each variant: the genomic position (chromosome and nucleotide), the name of the gene that contains it, its nature and the predicted effect on the relative protein, the alignment quality (assigned automatically by Illumina) and the coverage, its frequency based on dbSNP137 and NHLBI- Exome Sequencing Project 6500

allele frequencies databases, some scores predictive of protein function and mutational pathogenicity (SIFT [177, 178, 198]; Polyphen2 [179]; MutationTaster [199]. It reports also some scores of conservation rate (PhyloP [12]; GERP ++ [180] and some information stored in the laboratory database (based on previous WES data) and in the NHGRI GWAS [13] Catalogue, a catalogue of published genome-wide association studies and in the Database of Genomic Variants (DGV). DGV [14] is a collection of mutations found in health controls, in the Human Gene Mutation Database (HGMD®)[15] and in Orphanet [16], the databases of rare diseases. It is an attempt to order and link all published gene mutations responsible for human inherited diseases (HGMD® [183, 184]; NHGRI GWAS [200]; DGV [201]).

Bioinformatics tools

Stats_v1.py

The first step which has to be performed on each AnnovarAnnotate output is the check of its quality. "stats_v1.py" (developed on Python script by Dr Periklis Makrythanasis) is a pythline that was developed to quickly control the quality of .csv outputs. A good quality .csv file contains about 22.000 variants, subdivided in exonic, intronic, upstream, downstream, UTR3', UTR5'.

Awk

Awk [17] is a data driven language useful to create precise and selected filtered files.

The `-awk` command allows the search of files for raws (or other units of text) that contain choosen words or/and values. When a line matches the user's request, awk recalls that line in an output file, and it continues to process input lines in this way until it reaches the end of the input files.

Sed

The `-sed` command (stream editor) is one of the early Unix commands developed from 1973 to 1974 by Lee E. McMahon of Bell Labs. It was built for command line processing of data files. `-sed` is a line-oriented text processing utility: it reads text line by line from an input stream or file, applies one or more operations which have been specified via a sed script and reorders each line (modified as requested by the user) in an output file.

Samtools tview

Samtools tview is a set of utilities that permits to open and manipulate alignments in the BAM format [18] [194, 202]. It is mainly used to check the real presence of interesting mutations selected in the .csv files.

Variant Master

The Variant Master Software* (VM Manual attached as Appendix), developed by the Medical Genetic bioinformatics team of the University Medical School of Geneva, allows to analyse together Annovar Annotate outputs from the same family permitting to compare them on the basis of the familiar pedigree features and filter options inserted by the user.

Sanger Sequencing

Primers design

The primers were manually designed with the help of Primer3 Input [19], checked by Multiple Primer Analyzer|Thermo Fisher Scientific [20] and blatted on ucsc browser [21].

Polymerase Chain Reaction (PCR)

AccuPrime ™ Taq DNA Polymerase System (protocol attached in Appendix) was used to amplify exons containing the mutation of interest.
AccuPrime™ TaqDNA polymerase contains anti-TaqDNA polymerase antibodies that inhibit polymerase activity, providing an automatic "hot start" and the room temperature set-up [203]. The PCR was performed on 25µl volume wells: 3µl DNA (30-50ng/µl), 3µl Oligo forward (20Nm), 3µl Oligo reverse (20Nm), 16µl PCR mix (2.5µl 10X AccuPrime™ PCR Buffer II, 0.5µl AccuPrime™ TaqDNA polymerase, 13µl ddH2O.
35 cycles of PCR amplification were performed as follows:
Denature: 94°C for 15-30 s, Anneal: 65°C -55° for 15-30 s, Extend: 68°C for 1 min per kb
The PCR product was maintained at 4°C after cycling.
Multiplex and nested PCR were often performed due to the poor quality of DNA samples.

Gel electrophoresis

The PCR effectiveness was verified by the amplification products electrophoresis in 2% agarose gel and their visualization by SYBRsafe staining (SYBR® Safe DNA Gel Stain, cat N: S33102, datasheet attached). In case of double band PCR products the electrophoresis was repeated by NuSieve Agarose 3:1 (NuSieve™ 3:1 Agarose, datasheet attached) for double checking and nested PCR.

Pcr product purification

MicroCLEAN by Web scientific (instructions attached) was used to purifyPCR products.
20µl of MicroClean, a half spin DNA clean-up reagent, was added to each well. Then the plate was centrifuged at 3000rpm for 40'. The wells where then dried with a pump. 20µl of ddH2O was added to resuspend the amplified DNA.

Standard sequencing

The cycle sequencing PCR was performed by BigDye® Terminator v3.1 Cycle Sequencing Kit (ThermoFisher Scientific [22]): 2µl of each purified PCR product were prepared in a plate wells with 2µl of one of the two PCR primers (forward or reverse), 3µl of H2O, 1µl of Big Dye Terminator v3.1, 2µl of Big Dye Terminator v3.1 5x Sequencing Buffer. The plate was put in a thermal cycler, set to the correct volume. The PCR sequence program was run as follows:
rapid thermal ramp to 96 °C (initial denaturation) and 96 °C for 1 min; 25 cycles repeating: Rapid thermal ramp (1 °C/second) to 96 °C- 96 °C for 10 sec-rapid thermal ramp to 50 °C- 50 °C for 5 sec-rapid thermal ramp to 60 °C-60 °C for 4 min-rapid thermal ramp to 4 °C.
The products were then purified by EtHO washing (protocol attached): 1µl of NaAc/EDTA plus 40µl of 95% ethanol were added to each well to create the DNA pellet. The pellet was precipitated by 3000rpm 30 minutes centrifugation and the surnatant was then removed. 100µl of 70% EtHO were used to wash the wells. The DNA was resuspended in 12µl Hi-Di formaldehyde.
After PCR and EtHO washing, Sanger was performed on 12µl samples in Sanger 3500xL Genetic Analyzer by Applied Biosystems®.

Sequences check by Mutation Surveyor and Staden

The check of Sanger sequences was performed by Mutation Surveyor and Staden [23] softwares.

Immunofluorescence Analyses

Respiratory cells were obtained by trans-nasal brush biopsy of patients and resuspended in 5ml RPMI 1640 medium, without supplements.

The cells were then spreaded onto glass slides. The spreading was repeated until the whole solution was over. The slides were air dried and stored at -80°C until use.

The Immunostaining proceeded as follows:

1-On the day of the experiment, take out the slides and dry it.

2-After slides dried out, mark around the cells with a fat marker pen.

3- Treat samples with 500µl of 4% paraformaldehyde for 10'.

4- Wash 5 times with Dulbecco's Phosphate Buffer Saline (PBS).

5- Treat with 500µl of 0.5% Triton X in PBS for 5-10 minutes.

6- Wash 5 times with PBS.

7- Block the slides for 1hr with a 5% albumin bovine serum (BSA) prepared in PBS.

8- Incubate overnight at room temperature with primary antibodies prepared in BSA [i.e mouse anti-y [*gamma*] -tubulin (IgG1), mouse anti-acetylated-a [*alfa*]-tubulin (IgG2) in 1000ul dilution and primary antibody for protein of interest produced in rabbit (ex. Anti-rabbit-PIH1D3 in 1:500 dilution)].

9- Wash 5 times with PBS.

10- Add the secondary antibodies, prepared in PBS [i.e Alexa Fluor 488 (donkey) anti mouse (Green) and Alexa Fluor 594 (goat) anti rabbit (Red)]. Close the slides in a slide box to prevent light exposure and incubate at room temperature for 1hr.

11- Wash 5 times with PBS.

12- Add anti fading agent Alexa fluor and put on the cover slip.

Each Immunostaining experiment was twice performed on control and patient slides.

the Abs used in slides immunostaining are reported in table 2 (Primary Antibodies) and in table 3 (Secondary Antibodies).

Primary Abs					
target	**host**	**2Ab target**	**clonality**	**# ID**	**# ref. web**
Anti-PIH1D3	rabbit	IgG rabbit	polyclonal	Atlas Antibodies HPA051099	http://www.biocompare.com/9776-Antibodies/3286770-Anti-CXorf41/
Anti-Acetylated Tubulin	mouse	IgG mouse	monoclonal	T7451 Sigma	http://www.sigmaaldrich.com/catalog/product/sigma/t7451?lang=en&region=GB&cm_sp=Insite-_-prodRecCold_xorders-_-prodRecCold2-1
Anti-γ-Tubulin	mouse	IgG mouse	monoclonal	T5326 Sigma	http://www.sigmaaldrich.com/catalog/product/sigma/t5326?lang=en&region=GB&cm_sp=Insite-_-prodRecCold_xorders-_-prodRecCold2-1
Anti - GM130	mouse	IgG1, *kappa* mouse	monoclonal	BD Transduction Laboratories material number 610823	https://www.fishersci.com/shop/products/anti-gm130-clone-35-bd-2/bdb610823
DNAI2 (Mo1), clone 1C8	mouse	IgG2a, *kappa* mouse	monoclonal	Abnova Catalog #: H00064446-PW2	http://www.abnova.com/products/products_detail.asp?catalog_id=H00064446-PW2
Anti-DNALI1	rabbit	IgG rabbit	polyclonal	Sigma Prestige Antibodies HPA028305	http://www.sigmaaldrich.com/catalog/product/sigma/hpa028305?lang=en&region=GB
Anti-DNAH5	rabbit	IgG rabbit		Sigma Prestige Antibodies HPA037470	http://www.sigmaaldrich.com/catalog/product/sigma/hpa037470?lang=en&region=GB
Anti human GNT46	sheep	IgG sheep	polyclonal	AbD serotec AHP500	https://www.abdserotec.com/human-tgn46-antibody-ahp500g.html

Table 2. Primary Antibodies used in slides immunostaining. For each Ab: the target protein, the host specie, the target for the secondary antibody, the type of clonality, the Ab catalog number and the online datasheet are indicated in column 1, 2, 3, 4, 5 and 6 respectively.

Secondary Abs					
Fluorescence wavelenght	**host**	**target**	**clonality**	**# ID**	**# ref. web**
Alexa Fluor® 488	goat	anti-Rabbit IgG (H+L)	polyclonal	Life Technologies A-11034	https://www.thermofisher.com/order/genome-database/antibody/Goat-anti-Rabbit-IgG-H-L-Secondary-Antibody-Polyclonal/A-11034
Alexa Fluor® 594	goat	anti-mouse IgG1 (γ1)	polyclonal	Invitrogen A-21125	https://www.thermofisher.com/order/genome-database/antibody/Goat-anti-Mouse-IgG1-Secondary-Antibody-Polyclonal/A-21125
Alexa Fluor® 594	goat	anti-mouse IgG2b	polyclonal	Thermo Scientific A-21145	https://www.thermofisher.com/order/genome-database/antibody/Goat-anti-Mouse-IgG2b-Secondary-Antibody-Polyclonal/A-21145
Alexa Fluor® 594	goat	anti-Rabbit IgG (H+L)	polyclonal	Thermo Scientific A-11037	https://www.thermofisher.com/order/genome-database/antibody/Goat-anti-Rabbit-IgG-H-L-Secondary-Antibody-Polyclonal/A-11037
Alexa Fluor® 488	goat	anti-Mouse IgG2b	polyclonal	Thermo Scientific A-21141	https://www.thermofisher.com/order/genome-database/antibody/Goat-anti-Mouse-IgG2b-Secondary-Antibody-Polyclonal/A-21141
Alexa Fluor® 488	goat	anti-Mouse IgG1	polyclonal	Thermo Scientific A-21141	https://www.thermofisher.com/order/genome-database/antibody/Goat-anti-Mouse-IgG1-Secondary-Antibody-Polyclonal/A-21121
Alexa Fluor® 488	donkey	anti-Mouse IgG (H+L)	polyclonal	Thermo Scientific A-21202	https://www.thermofisher.com/order/genome-database/antibody/Donkey-anti-Mouse-IgG-H-L-Secondary-Antibody-Polyclonal/A-21202
Alexa Fluor® 488	donkey	anti-Sheep IgG (H+L)	polyclonal	Thermo Scientific A-11015	https://www.thermofisher.com/order/genome-database/antibody/Donkey-anti-Sheep-IgG-H-L-Secondary-Antibody-Polyclonal/A-11015

Table 3. Secondary Antibodies used in slides immunostaining. For each Ab: the fluorescence wavelength, the host specie, the target primary antibody, the type of clonality, the Ab catalog number and the online datasheet are indicated in column 1, 2, 3, 4, 5 and 6 respectively.

Western Blot

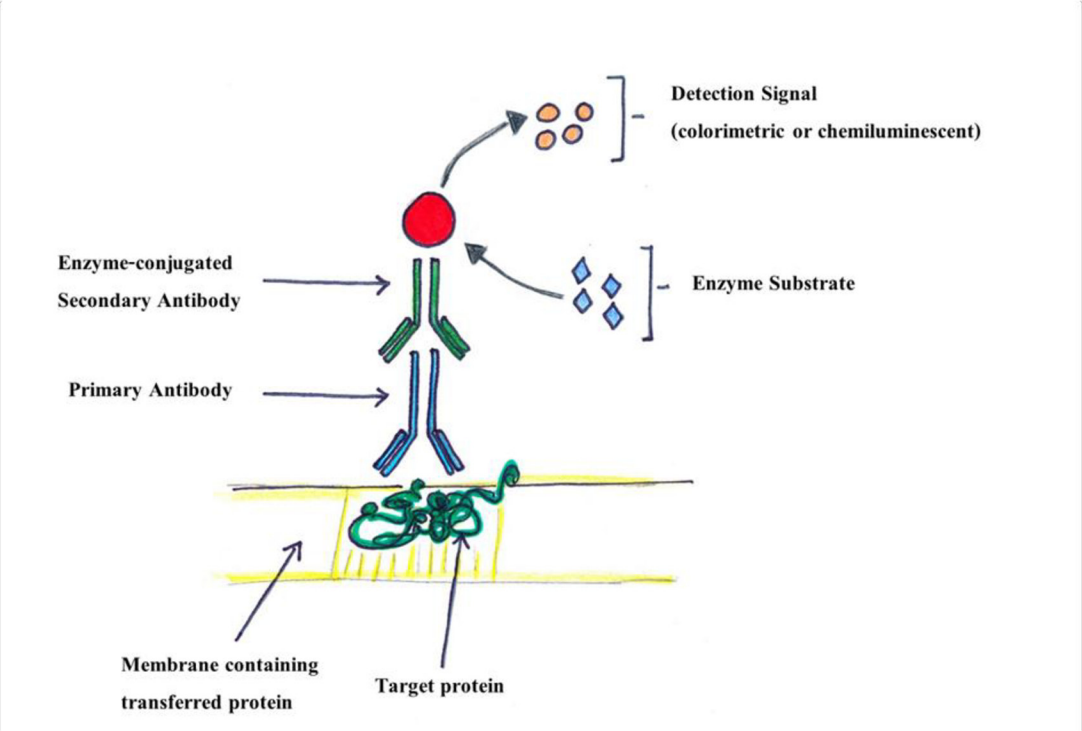


Figure 37. Mechanism of detection in Western Blot [Picture Copyright C. Olcese].

The WesternBlot protocol proceeds as follows:

1- Make the polyacrylamide gels (resolving and stacking gel). They are composed as showed in the table behind (table 4):

	Resolving gel (running) 10%	**Stacking gel**
1.5M Tris HCl ph8.8	2.5ml	-
0.5M Tris HCl ph6.8	-	1.3ml
30% acrylamide/bis	3.3ml	660µl
10%SDS	100µl	50µl
ddH2O	4ml	3ml
10%APS (stored at 4°C)	150µl	100µl
Temed	15µl	10µl*

Table 4. Resolving gel and stacking gel for WesternBlot composition.

2- Prepare the running machine and top it up with the running buffer (1L: 800ml of ddH2O; 100ml of tris-glycine 10x, 5ml of 10% dodecyl sulfate sodium-SDS). Mix the chosen amount of samples with a half part of Laemmali sample buffer. Add to the marker and the samples 1µl of beta-mercaptoethanol and incubate for 5 minutes at 95° to allow protein denaturation.

Use the Color Prestained Protein Standard, Broad Range (11–245 kDa) (NEB # P7703 [24])
as marker (figure 38).

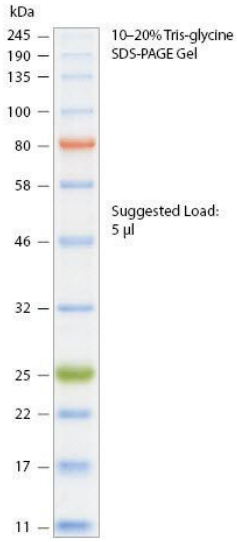


Figure 38. The Color Prestained Protein Standard after running.

3- Apply 80V until the samples stack. Then performe the running at 200V for 45 minutes.

4-After the samples run, prepare the transfer gel-membrane:

a-Mount the proper machine and fill it with the transfer buffer (1L: 700ml of H2O; 200ml of methanol, 100ml of tris-glycine 10x).

b-Set up the 'transfer sandwich'. It has to be filled, from the black side to the white side of the support gride, as follows: sponge-3 layers of filter paper-gel- transfer membrane-3 layers of filter paper –sponge.

c- Perform the transfer at 300 mA for 1:30h.

5- Block the membrane for 1hr at room temperature. The blocking buffer is prepared in washing buffer (1L of PBS Tween: 999.5ml of ddH2O; 10 of PBS tablets; 500µl of Tween 20) with 5% of skimmed milk and 3% of BSA.

6-Incubate the membrane overnight at 4°C with the primary antibodies which had been previously diluted in blocking buffer. The used antibodies are reported in table 5.

7-Wash 5 times the membrane for 5 minutes by the washing buffer.

8-Incubate the membrane with the secondary antibodies for 50' at room temperature. The used

antibodies are reported in table 5.

9- Wash 5 times (5 minutes each) by the washing buffer.

10- Add the reaction mixture Pierce™ ECL Western Blotting Substrate [25] and leave it on for 1 minute.

12-Develop the pictures.

Primary antibodies				
Anti-Glyceraldehyde-3-Phosphate Dehydrogenase Antibody, clone 6C5	Mouse monoclonal	IgG1 mouse	Millipore MAB374	http://www.merckmillipore.com/GB/en/product/Anti-Glyceraldehyde-3-Phosphate-Dehydrogenase-Antibody%2C-clone-6C5,MM_NF-MAB374?bd=1
Anti-PIH1D3	Rabbit polyclonal	IgG rabbit	Atlas Antibodies HPA051099	http://www.biocompare.com/9776-Antibodies/3286770-Anti-CXorf41/
Secondary antibodies				
Horseradish peroxidase conjugated species-specific whole antibody	donkey	Anti-rabbit IgG	GE HealthCare Life Sciences NA934	http://www.gelifesciences.com/webapp/wcs/stores/servlet/productById/en/GELifeSciences-uk/25800685
Horseradish peroxidase conjugated species-specific whole antibody	sheep	Anti-mouse IgG	GE HealthCare Life Sciences NA931	http://www.gelifesciences.com/webapp/wcs/stores/servlet/catalog/en/GELifeSciences-uk/products/AlternativeProductStructure_16827/25005173

Table 5. Antibodies used in Western Blotting.

Nasal brushings, Light Microscopy and High Speed Video Method

Nasal brush biopsies were collected from the inferior turbinate using a modified 3-mm bronchial cytology brush [204].

Nasal brushings were placed in Medium 199 (pH 7.3) which contained antibiotic solution (streptomycin 50 μg/ml, penicillin 50 μg/ml, Gibco, UK). Ciliated strips of epithelium were suspended in a chamber created by the separation of a cover slip and glass slide by two adjacent cover slips. The slide was placed on a heated stage (37°C) of a Leitz Diaplan microscope mounted on an anti-vibration table (Wentworth Laboratories Ltd, UK).

Beating ciliated edges were recorded using a digital high speed video camera (Kodak Ektapro Motion Analyser, Model 1012) at a rate of 400 frames per second, using a shutter speed of 1 in 2000. The camera allows video sequences to be recorded and played back at reduced frame rates or frame by frame. The precise movement of individual cilia may be observed during their beat cycle.

The path taken by a cilium during the power and recovery strokes was plotted on acetate paper overlying the high resolution monitor. Viewing the cilia beating towards the observer (fig 1B), the precise position of the cilium during the forward power stroke was plotted frame by frame. As the cilium moved backward during the recovery stroke its position during this movement was again plotted frame by frame. An angle could be derived from a line drawn through the plane of the power stroke and a line joining the point of maximum deviation of the cilium during the recovery stroke (fig 1B). This was defined as the beat angle α. For each cilium studied the mean angle for five complete beat cycles was measured by image analysis (Scion image, Scion Corporation, Frederick, Maryland, USA). This was repeated when viewing the ciliary beat pattern from above (fig 1C) and a beat angle β was calculated by image analysis.

Measurements of ciliary beat frequency were made from ciliated epithelial strips at least 50 μm in length viewed in sideways profile (fig 1A) using the digital high speed video, photomultiplier, and photodiode methods. The order in which measurements were made by the three different techniques varied equally to help to exclude any confounding effect of the order of measurement.

Ciliary beat frequency (CBF) may be determined directly by timing a given number of individual ciliary beat cycles. Groups of beating cilia were identified and the number of frames required to complete 10 cycles recorded. This was converted to ciliary beat frequency by a simple calculation (CBF= 400/(number frames for 10 beats) × 10). Ten measurements of beat frequency were made along each ciliated strip [204].

Electron Microscopy and Tomography

Nasal brush biopsies were fixed in 2.5% glutaraldehyde in cacodylate buffer and processed as previously described [205]. Briefly, cells were washed in sodium cacodylate buffer, post-fixed with 1% osmium tetroxide and centrifuged in 2% agar to generate a pellet. Using a series of increasing concentrations of methanol followed by propylene oxide, cells were dehydrated before embedding in Araldite resin. For traditional TEM single or serial 70- to 90-nm-thick sections were cut, mounted onto copper grids and stained with 2% methanolic uranyl acetate and Reynolds lead citrate. For tomography 150 and 200 nm sections were cut and 10 nm gold particle solution applied prior to the final staining [206].

Data Acquisition

Cilia originating from healthy strips of epithelium and positioned in the center of the grid were selected for electron tomography. Data was acquired on a JEOL JEM 1200EX using 150nm. Each sample was tilted about one axis with 2° increments over a tilt range of ±60°. Subsequently the grids were rotated 90° and perpendicular tilt series were collected. Images were recorded at 15,000× at a pixel resolution of 1024 by 1024 (0.786 nm/pixel). A Teitz charge-coupled device (CCD) camera was used to record the images and the defocus was configured to between −8 and −12 μm so that the first minimum in the Fourier transform fell at 1/5 nm. Tomograms were taken from five longitudinal and five transverse cilia.

Image Processing

The tomograms were generated and serial tomograms joined together using IMOD [26]. For longitudinal models three tomograms were generated per cilium in order to incorporate the entire ciliary length. For transverse models three serial section tomograms were joined together. The data was modeled in IMOD and visualized using UCSF Chimera [207].

Subtomographic Averaging

Using PEET part of the IMOD software package, similar areas within the tomographic data were extracted and by employing a rotational and translational search, were matched-up and combined to give an averaged density.

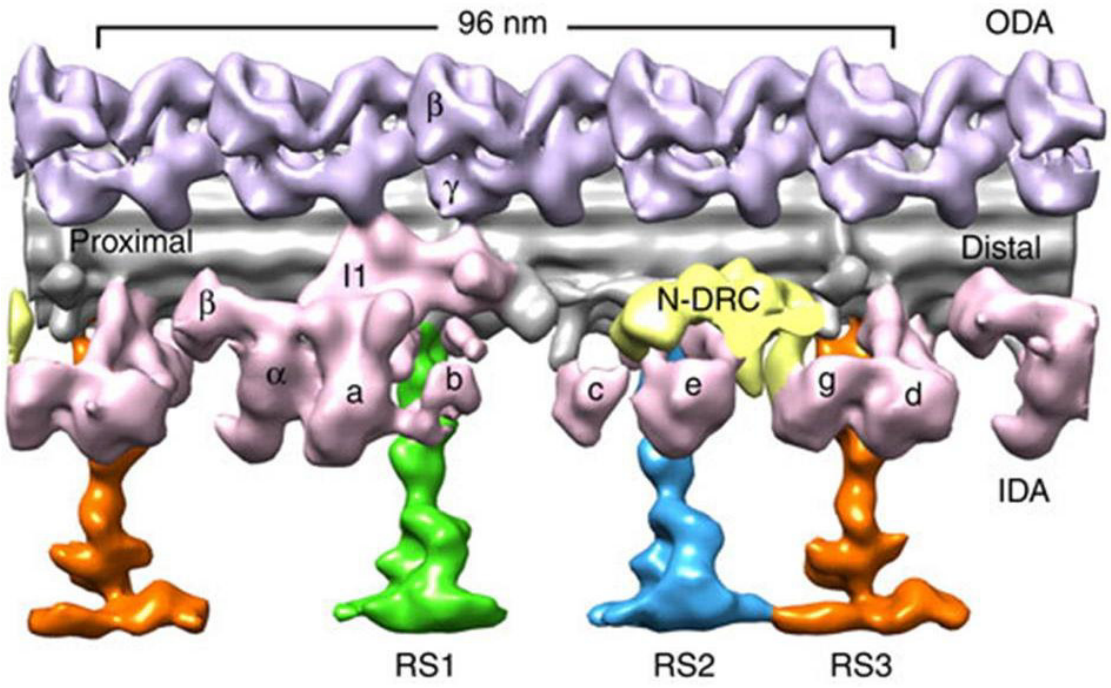


Figure 39. Microtubule pair tomogram. The picture shows how the tomogram of a microtubule pair appears. The microtubule pair is coloured in gray. ODAs are purple, IDAs are pink, the N-DRC complex is yellow, the Radial Spokes (RS) are showed in orange, green and blue [Copyright: A. Shoemaker and colleagues].

Results

DNA samples Collection

The sample collection managed by the Medical Genetics laboratory of the University Medical School in Geneva is conserved and continously increased. Close relationships with the Swiss and foreign phyicians who sent their patient samples to Geneva are mainteined and patient data are continuously updated.

DNA extraction was progressively performed. In case of fresh blood, the DNA was extracted by in column NucliSENS® protocol or by the authomated platforms Arrow Autogen and easyMag®.
For older blood samples the manual extraction resulted to be the only effective one.

DNA samples were archival and subjected to repeated cycles of thawing and freezing over prolonged period. This resulted often in DNA degradation.
Since 3µg of DNA are required for WES and 2µg for HaloPlex, all DNA samples were re-measured by both Nanodrop and Qubit.
Measurement by Nanodrop is somehow inaccurate. In fact the absorbance is a natural property of many other molecules than nucleic acids and proteins and they can contaminate the sample and result in overestimates. Moreover, the absorbance method does not distinguish between DNA, RNA, protein or free nucleotides or amino acids in the sample. By the use of Qubit it is possible to get a precise quantification of DNA even in the presence of other bio-molecules.

DNAs in best condition were selected for WES and Haloplex. WES was performed in individuals with an interesting phenotype and belonging to a family (confirmable segregation). Only 65/539 DNAs were chosen for WES and 96/539 for Haloplex.
The DNAs which presents a lower volume or/and a lower quality than requested by these methods can't be analysed until the biotechnological development will permit to obtain NGS from smaller amount of DNA or degraded molecules.

WES and Halopex DNA libraries preparation

All libraries were successfully prepared for selected DNAs following Illumina preparation protocols.
The sequencing priority was given to those families in which pedigree structures and phenotypes were the most probable to identify with disease causing variants. The first criteria applied was that each family should have at least two affected members. The next-generation whole-exome sequencing (WES) was used to analyze the two affected individual and at least one healthy sibling.

In a second round, trios (single affected with parents) and some singletons were also sequenced. A total of 65 individuals (56 patients and 9 family members), belonging to 46 families, were analyzed by WES. Another 96 additional individuals (85 patients and 9 family members) underwent the Haloplex panel customer designed for 640 ciliome genes.

WES and Haloplex analysis

Phytlines and .csv files quality check

The pythlines were all successfully launched. The quality check by stat.v1 script was performed for each of them. They all showed a good quality. A good quality file contains about 22.000 variants, subdivided in exonic, intronic, upstream, downstream, UTR3, UTR5. An exemple of quality check output is shown in figure 40.

```
> [out/annovarannotate_PCD_GVA59_8176_ATTGGCTC_L007_L008
synonymous SNV  11242   48.4214153422%  11080   98.5589752713%
synonymous SNV_HC       11048   49.9389775347%  10922   98.8595220854%
nonsynonymous SNV       10083   43.4293836413%  9806    97.2528017455%
nonsynonymous SNV_HC    9789    44.2480676219%  9563    97.6912861375%
stopgain SNV    100     0.430718869794% 78      78.0%
stopgain SNV_HC 79      0.357094426615% 71      89.8734177215%
stoploss SNV    14      0.0603006417711%        14      100.0%
stoploss SNV_HC 14      0.0632825566153%        14      100.0%
unknown 632     2.7221432571%   506     80.0632911392%
unknown_HC      561     2.53582244723%  484     86.2745098039%
frameshift deletion     247     1.06387560839%  21      8.5020242915%
frameshift deletion_HC  122     0.551462279076% 18      14.7540983607%
frameshift insertion    189     0.81405866391%  43      22.7513227513%
frameshift insertion_HC 93      0.42037698323%  35      37.6344086022%
frameshift substitution 2       0.00861437739587%       0       0.0%
frameshift substitution_HC      2       0.00904036523076%       0       0.0%
nonframeshift deletion  302     1.30077098678%  33      10.9271523179%
nonframeshift deletion_HC       185     0.836233783845% 27      14.5945945946%
nonframeshift insertion 393     1.69272515829%  148     37.6590330789%
nonframeshift insertion_HC      224     1.01252090584%  120     53.5714285714%
nonframeshift substitution      13      0.0559934530732%        0       0.0%
nonframeshift substitution_HC   6       0.0271210956923%        0       0.0%
        0       0.0%    0       NA
_HC     0       0.0%    0       NA
Total exonic    23217           21729   93.5909032175%
Total exonic HQ 22123           21254   96.0719613072%
Splicing        1327    5.71563940216%
Splicing_HQ     983     4.44333951092%
chrX synonymous SNVs hom|total|ratio]
```

Figure 40. e.g of quality-check output for patient 8176 (GVA059).

Filtering of variants

Among the crowd of only exonic and splicing variants (±10bp of the intron-exon boundary) identified in each individual exomes (10'000 non-synonymous in average), the putative causative variant(s) that explain the disease in each family had to be isolated. Non-synonymous, frameshift, splicing, and indel variants were favoured, according to novelty, quality score, and putative pathogenicity.

Variants were at first filtered to remove 1) synonymous variants, 2) variants with allele frequency bigger than 0.01 in dbSNP version 135, 1000Genomes, Exome Variant Server [27], Exome Aggregation Consortium (ExAC [28]) and Geneve local database, 3) variants found within segmental duplications of the genome. The predicted pathogenicity scores provided by SIFT, PolyPhen2 and Mutation Taster were also taken in account.

Awk command to work out filtering

The .csv files can be opened and analysed in Excel by saving them in .Excel Worksheet format. However Excel based manual filtering is time consuming.

Therefore the basic function of *awk* was used to write a more effective filtering script.

the`-sed` command to group data

The `-sed` command was used to group individual .csv files in a single big file before filtering, adding the individual file name in the first column to differentiate the variants. This method permitted better visualization of the data and improved filtering productivity. In fact, it helps to rule out the variants present in all patients that can therefore be considered non pathological.

Known genes screening

An `awk` script was used at first to create a .csv file containing only the variants found in the 34 PCD- known genes. For each individual (often in sed files), the presence of pathological mutations in 34 known genes was checked.

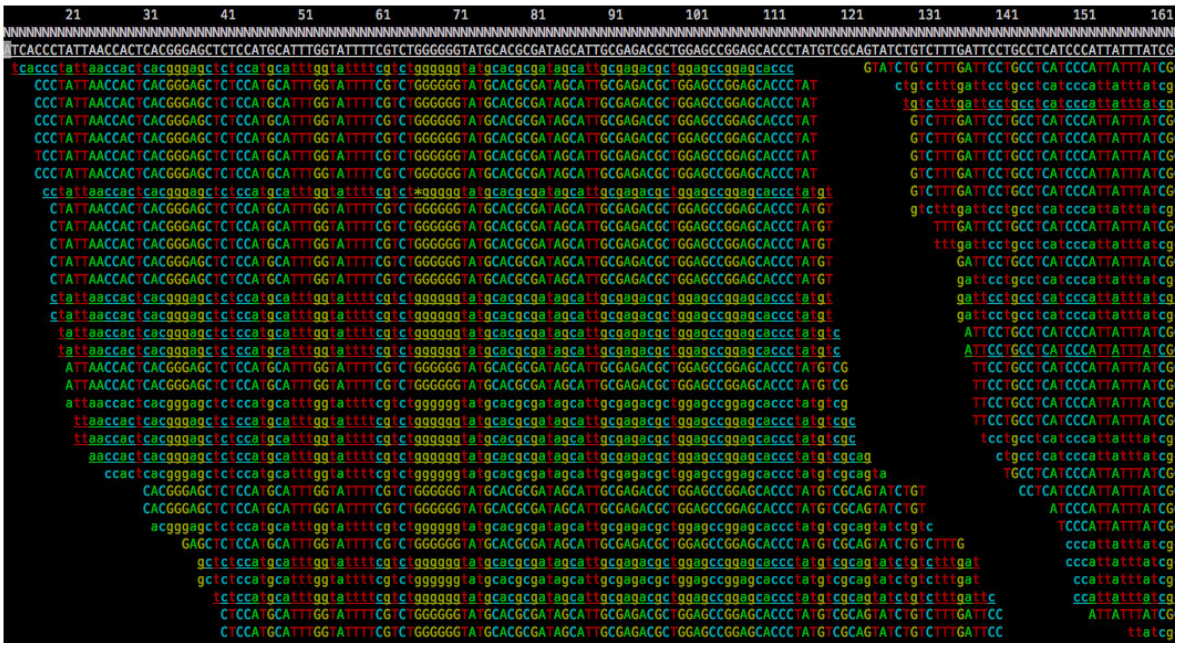
Variants were filtered for 1000g2012apr_all >= 0.01. Frameshift (deletion, insertion), exonic stopgain and nonsynymous mutations and splicing disruptions were favoured.

The first screening was reserved to homozygous exonic synonymous SNPs which were filtered for ljb2_pp2hvar, ljb_gerp++ and snp137. If they were valued interesting, they were analysed also by

considering ncbi SNPs, Exome Variant Server and cilioproteome.org database.

The most interesting variants were checked by Samtools tview. Samtools tview allows to reveal false positives and the presence of uncovered exomic areas (expecially with haloplex probes).

A



B

Figure 41. Samtools tview. A) Reads alignement in 'coloured bases' mode. **B**) How true single-base mutations appear.

The second step, if strong homozygous mutations were not found, was to consider also heterozygous frameshift, stopgain and nonsynymous mutations, looking for a compound heterozygosity disease explanation.

Several variants in the 34 known-genes were found by focused bioinformatics analysis of the WES and Haloplex sequencing. These variants were verified by Sanger and, in individuals who were part of a family, the segregation in all available family members was checked.
In 40% of patients likely pathogenic variants in known genes were confirmed. The most frequent mutated gene in the selected cohort is DNAH5 (12.42%), followed by CCDC40 (6.83%), HYDIN (4.35%) and CCDC39 (3.73%). Mutation in other PCD genes accounted each for less than 3%: DNAAF3 (2.48%), DNAH11 (1.86%), LRRC6 (1.86%), ZMYND10 (1.24%), ARMC4 (1.24%) DNAAF1 (1.24%), DYX1C1 (0.62%), DNAI1 (0.62%), SPAG1(0.62%), RSPH9 (0.62%), RSPH4A (0.62%).

The families and individuals that didn't show new or already reported mutations in known PCD causative genes were selected for further analyses.

One example of mutations found in known genes is reported.

Family GVA024: one novel mutation found in a known gene: DNAH5

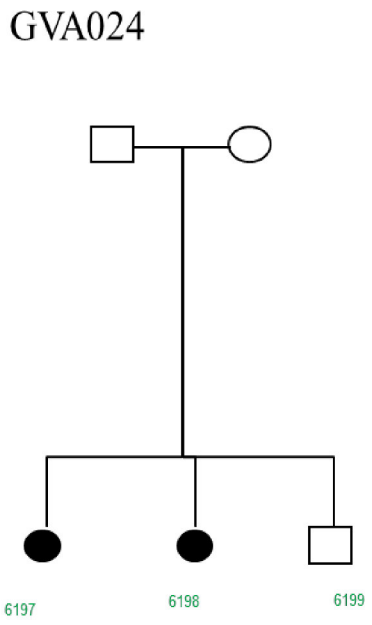


Figure 42. GVA024 genealogical tree.

Family GVA024 is from USA with British ethnic origin. The 3 children (1 male, 2 females) were born from unaffected individuals. The two females reported recurrent infections of upper air ways and situs inversus. For individual 6197 the diagnosis was done by nasal biopsy and the electronic microscopy was performed: the dynein arms resulted to be absent.
For individual 6198 nor biopsy nor EM were done and the diagnosis was based on situs inversus and on the diagnosis of the affected sibling.
DNA was collected from all the familymembers, including mother and father (6201, 6200). WES NGS on the 3 children (6197, 6198, 6199) DNAs was performed.
A new nonsynonymous homozygous mutation was found in DNAH5 of the two affected sisters [DNAH5:NM_001369:exon32:c.5177T>C:p.Leu1726Pro].
DNAH5 is well known to be a PCD causative gene and the Annovar Annotate scores for this variant known as rs138890576 are convincing. The coverage and quality scores resulted appropriate and the allele frequency in the ESP6500 database resulted to be of 0.000077. The observed SNV was checked (rs138890576) by dbSNP ncbi database (ID: 1767) and by the Exome Variant Server (EVS). It resulted validated by cluster and by frequence. The MAF is not reported, but the allele G was found in heterozigosity in 0,001%. It was never observed in homozygosity, as reported in EVS (All Genotype #: GG=0/GA=1/AA=6502). The mutation is not present in ExAC.

The substitution in exon 32 of the Timine in position 5177 by a Citosine is predicted to cause an AA switch from Leucine in position 1726 to Proline (both non-polar, hydrophobic amino acids).
PolyPhen2 indicates this SNP as a probably damaging SNV with a score of 1.0 (the maximum since PolyPhen2 give a damaging score between 0 and 1) and the GERP Conservation score is 4.74. This means that the mutation is located in a very highly conserved domain.
The segregation of c.5177T>C mutation in family 24 was confirmed by Sanger. It is homozygous in all affected and heterozygous in the unaffected siblings and in the father. The mother DNA was found to be degraded, therefore un-suitable for PCR.

Ciliome gene (Haloplex) panel analysis

In families and individuals without convincing mutationd in known genes, -awk was used to create a set of .csv file selected for Ciliophaties related genes (the same gens selected for the Haloplex panel). These files were filtered as previously described.

The analysis of the 640 ciliome genes analysed both in WES and in haloplex panel gave the result shown in table 6. In 25% of families pathogenic variants were found in very interesting candidate genes.
Haloplex panel analysis individuated also three patients with mutations in CFTR. Cystic Fibrosis clinical phenotype, in the mild cases, might overlap with PCD.

In the sequencies obteined with the Haloplex selection tool, the samtools check revealed a lot of false positive and the presence of a lot of uncovered exomic areas, more than extimated during the haloplex probes design check. The main difference between WES and Haloplex selection methods is the way in which DNA fragments are created. The sonication method (used in WES) results in the creation of 200 bp DNA fragments with random cut-areas, meanwhile the digestion by restriction enzyme (use in Haloplex) produces mandatory cuttings in the restriction enzymes consensus sequences.
Therefore, WES in frame reads alignement is more effective than Haloplex targeted sequencing and its 'in column' fragment disposal (figure 40). WES sequences permits to have a major probability to cover all reference basepairs: where one sequence doesn't align, another does. On the contrary it is impossible by Haloplex to infer anything on DNA regions that fall outside 'standard' fragments 3' and 5' ends.

WES analysis

Variant Master was extensively used to identify new candidate genes in patients who underwent WES and who were not solved by the known genes and by the ciliary panel screening.
The Variant Master Software allows the analysis of all the Annovar Annotate outputs from the same family permitting to compare them on the basis of the pedigree features and filter options inserted by the user. The recessive filter which considers only the variants with a confident quality score, that are in exons or in splicing sites and excludes the synonymous variations, was applied. The recessive filter selects mutations which are common to all family members, indicating as ‘passed’ only the ones which are homozygous in the affected individuals and heterozygous in the health consanguineous. The Variant Master launch appears as follows:

```
> python ~fsantoni/tools/VariantMaster_REL2/variantMaster

Usage: variantMaster -c <cfg_file> -t <tfam_file> -o <output_prefix>
[-s <csv_folder> |-v <vcf_folder>] -b <bam_folder> -H <hap_file
[optional]> <options>
```

A .tfam and a.cfg files were generated, following VM manual, to be applied on .bam and .csv files. The .tfam file describes the genealogical tree and the .cfg states the parameters of the used filter (e.g recessive filter, de novo filter, etc). The .bam and .csv files are those previously generated during the pytline running.

```
Fam030 2068 0 0 1 1
Fam030 2069 0 0 0 1
Fam030 6270 2068 2069 1 2
Fam030 6271 2068 2069 1 2
Fam030 6273 2068 2069 0 1
```

Figure 43. e.g of a .tfam file

```
#General variables
CARR_THR :=    0.95
HOM_THR  :=    0.75
reference :=   '/data/hg19/hg19.unmasked.fa'
#Filter definitions
Filter_RecessiveFilter
conditional_select    INFO  in    PINDEL       QS   >=    600
conditional_select    INFO  in    INDEL        QS   >=    200
select Func     in    exonic;splicing
select ExonicFunc     !=    synonymous SNV
remove segdup   -     -
select 1000g    <=    0.01
select esp6500  <=    0.01
End_Filter
```

Figure 44. e.g of a .cfg file

The examples of two families analysed by Variant Master are reported.

Family GVA030: individuation of a new PCD gene: PIH1D3

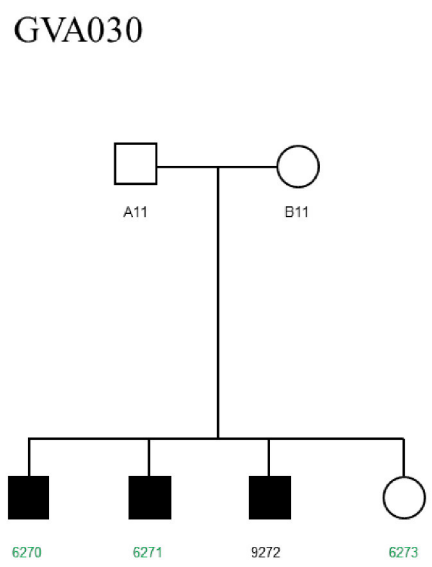


Figure 45. GVA030 genealogical tree.

Family GVA030 has Spanish ethnic origin. The 4 children (3 males, 1 female) were born from unaffected individuals. Three (all three males) of them reported Kartagener's syndrome (situs inversus) and the absence of dynein arms (outer and inner). Furthermore individual 6270 spermatozoa are completely immotile.
DNA was collected from all the affected and non-affected members of this family including mother and father.
Whole Exome Sequencing was performed on 2 affected brothers (6270, 6271) and the unaffected sister (6273) DNAs.

After a first step, in which the presence of new or already reported mutations was excluded in the 34 so far known PCD genes, the analysis was then extended to the remaining genes of entire genome.
The Variant Master Software was used, set for the recessive transmission mode. VariantMaster considered 'recessive' also the variants on chromosome X that were hemizygous in the males and absent in the female.
A novel deletion in the X-linked PIH1D3 gene was found in exon 6 and further investigated [PIH1D3: NM_173494: exon6: c.489_492delCAAT: p.Ile164Leufs*9].

The PIH1 Domain Containing 3 [PIH1D3] gene (Gene ID: 139212) is located on chromosome Xq22.3 and it is transcribed in two alternative mRNAs (NM_001169154, NM_173494) that differ for the inclusion/exclusion of the untranslated exon 2. The resulting protein is 214aa residues long (NP_001162625.1 and NP_775765.1). Homologous genes in animal models seems to be involved

in ciliary assembly [64, 80], therefore PIH1D3 is an excellent candidate PCD gene, and the analysis was expanded to other families and functional studies were performed.

Two highly conserved domains are part of PIH1D3. SMART and Pfam programs locate the human PIH1D3 PIH1 domain between amino acid residues Glutamine 51 and Methionine 202. The shorter alpha crystalline domain (ACD) is then included in the PIH1 domain length. ACD is found in alpha-crystalline-type small heat shock proteins (sHsps) and it is similar to the domain of p23 (a cochaperone for Hsp90) and other p23-like proteins.

The deletion of CAAT (c.489_492delCAAT) falls in both the high conserved domains, breaks out the codons encoding for Thr163 and Ile164 and results in a frameshift starting with the substitution of the Leucine164 by an Isoleucine. The frameshifts exits in a premature TAA stop codon after nine residuals, at aa 173 (figure 47). Since a stop codon is introduced before the last exon, presumably the mRNA is subject to Nonsense Mediated Decay and therefore the mutant protein might not be translated.

Sanger sequencing was performed in all the members of the GVA30 family to verify the presence of the mutation in PIH1D3 and its X-linked segregation. Sanger sequencing confirmed that the affected males (6270, 6271, 9272) are all three hemizygotes for the CAAT deletion and that the mother (6269) is an obligated heterozygous carrier. The father (6268) has the wild-type allele and the helthy sister heredited the wild-type allele from both the parents.
Once the X-linked transmission was hypotized in the family, the analysis of the pedigree was extended to the maternal side. The mother has two healthy sisters and three brothers, one of which shows recurrent bronchitis and has not children. But clinical examination excluded PCD in this family member.

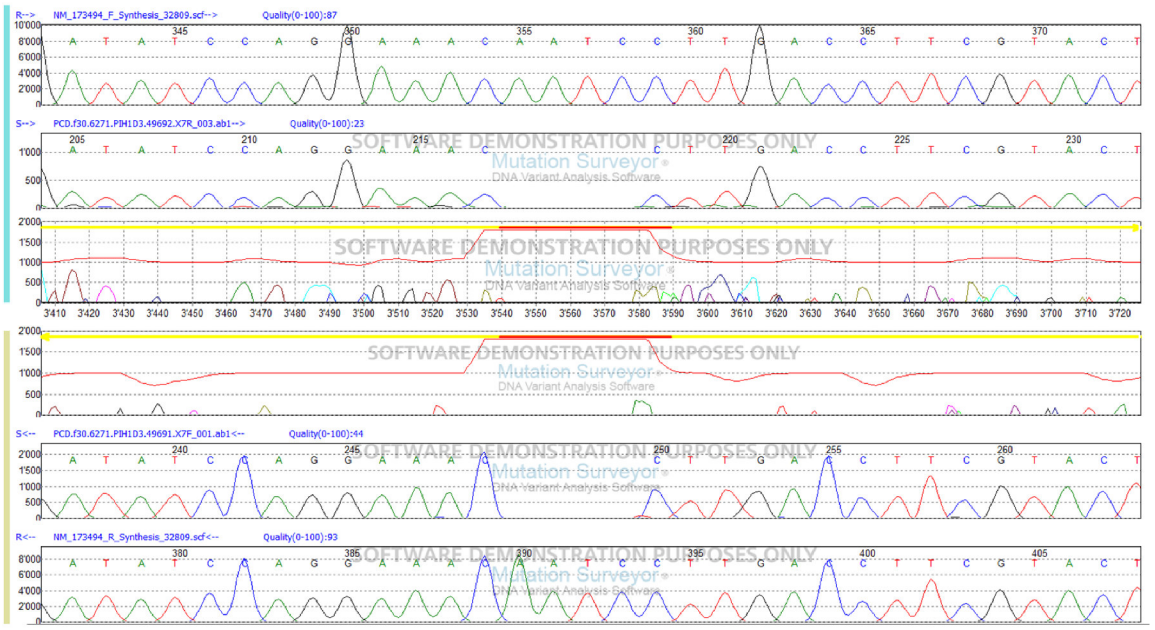


Figure 46. Family GVA030 Sanger sequence chromatogram of patient 6271.

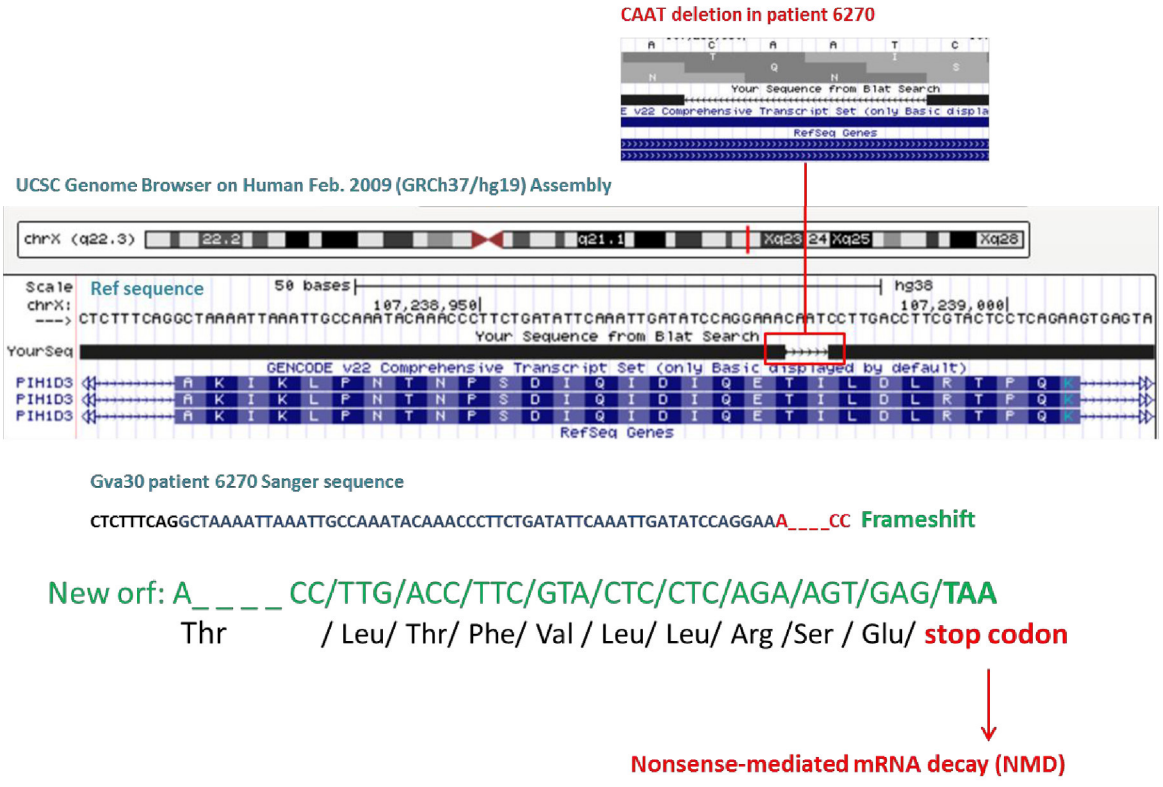


Figure 47. CAAT deletion and frameshift consequence in GVA30 patient 6270. The Ref sequence and the patient Sanger sequence are shown and compared. The deletion position and its result are indicated in red, the frameshift and the new open reading frame (orf) are in green.

The Sanger analysis of 12 additional families with a potentially X-linked segregation pointed out a missense mutation in one patient and another rare variant in the splicing site of one of the not translated exons, supporting the role of PIH1D3 in PCD. Two additional patients were found with PIH1D3 mutations by our collaborators in UK.

On chromosome X there are two genes involved in PCD/KS associated to other clinical symptoms: RPGR with Retinite Pigmentosa and OFD1 associated to oral-facial-digital syndrome type-I (REF). PIH1D3 is the first X-linked gene found to account for isolated PCD.
To date very rare families with X-linked Kartagener Syndrome without additional symptoms have been reported [208].

Immunofluorescence Analyses
The Abs that are usually used in PCD diagnostic practice individuate specific cilia components and are as follows:
-acetylated tobulin as cilia marker;
-DNAH5 as ODA marker;
-GAS8 as dynein regulatory (nexin-link component) marker;
-DNALI1 as IDA intermediate chains marker;
-RSPH1, RSPH4A and RSPH9 as radial spoke head proteins markers.
They can be differently combined to perform interaction studies.

Antibodies reported in table 2 (Primary Antibodies) and in table 3 (Secondary Antibodies) were used in slides immunostaining with the following results:

-PIH1D3 alone in wild-type nasal epithelial cells

The PIH1D3 results to clearly localize in the cytoplasm. This confirms the hypothesis that this protein is involved in cilia assembly. It is so close to the nucleous that it could seem a Golgi staining.

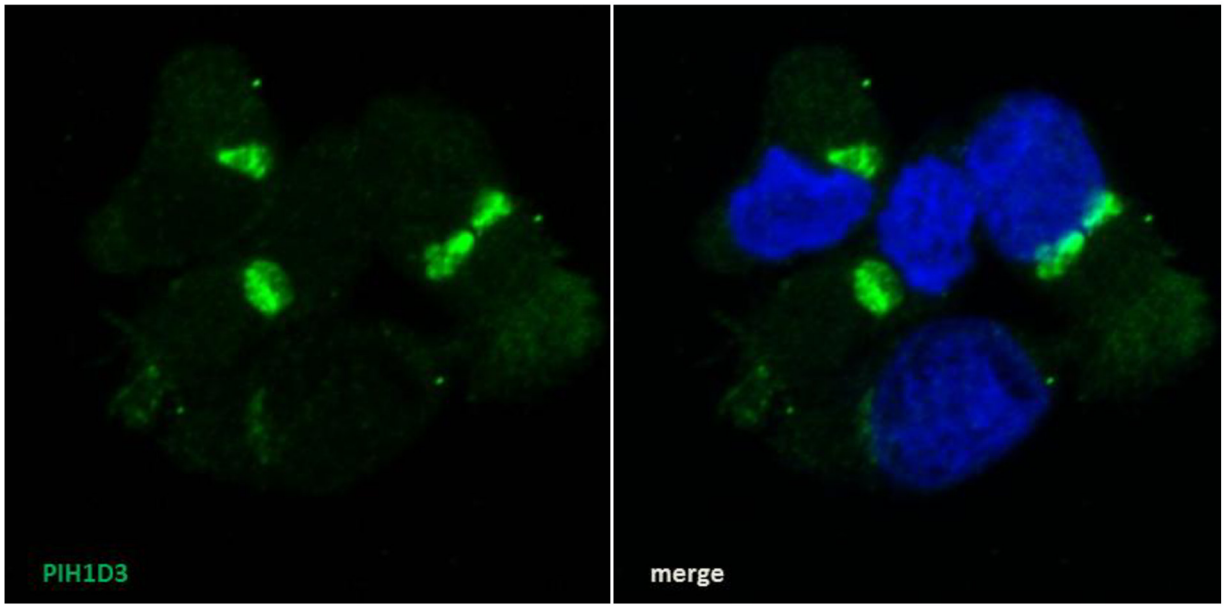


Figure 48. PIH1D3 staining in wild-type nasal epithelial cells. PIH1D3 is shown in green (Alexa Fluor 488), nucleous are in blue (DAPI).

-DNAI2 (ODAs marker) + PIH1D3 in wild type nasal epithelial cells

DNAI2 (an outer dynein arm marker) is found in the axoneme as expected. While the PIH1D3 results to clearly localize in the cytoplasm.

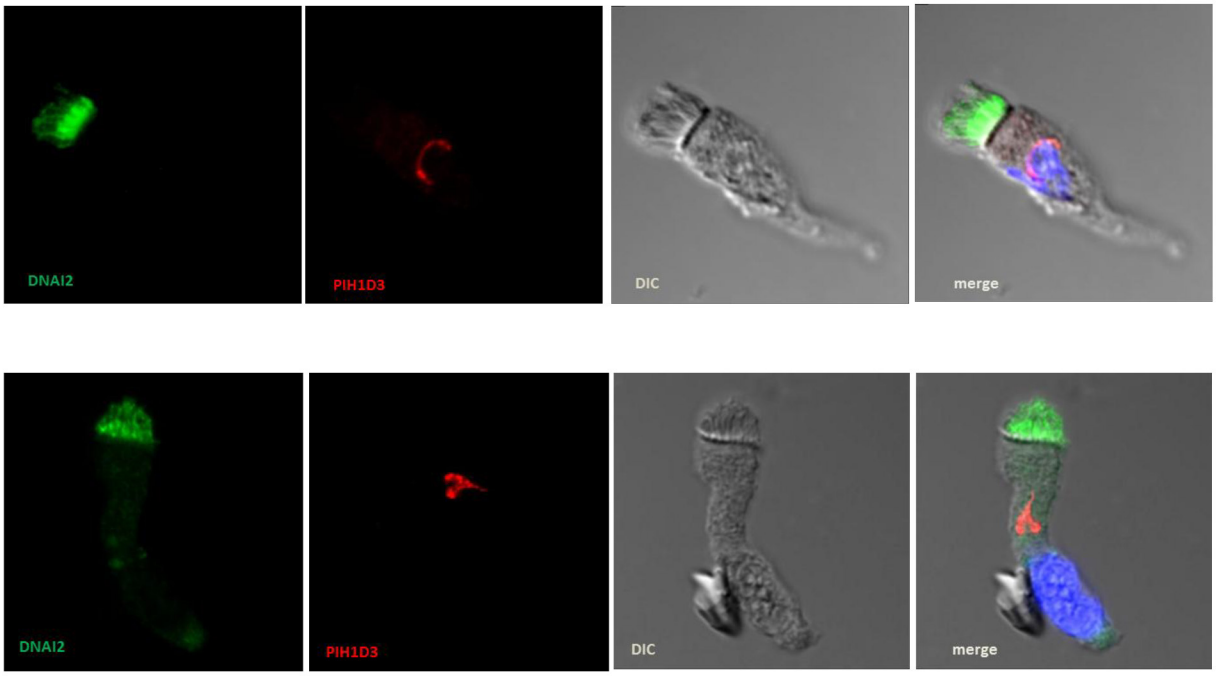


Figure 49. DNAI2 + PIH1D3 staining in wild type nasal epithelial cells. DNAI2 is labelled in green (Alexa Fluor 488), PIH1D3 is in red (Alexa Fluor 594), nucleous are in blue (DAPI).

-Tubulin (axonemal marker) +PIH1D3 in wild type nasal epithelial cells

Tubulin strongly stains in the axoneme as expected. While the PIH1D3 results to localize in the cytoplasm and to be very close to the nucleous.

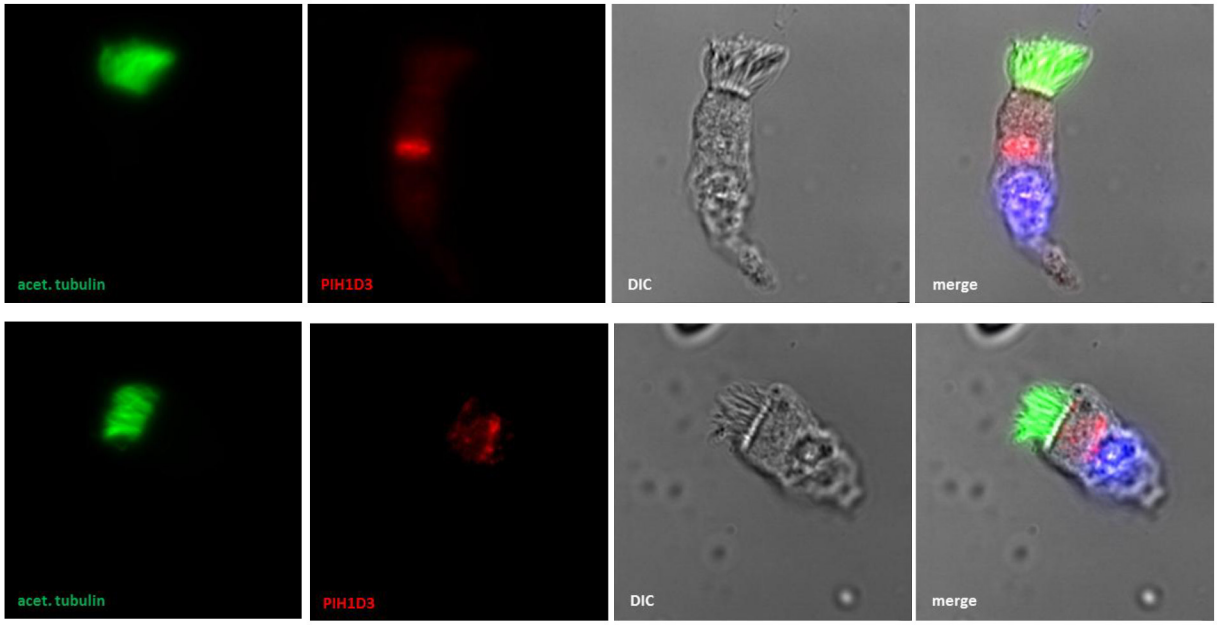


Figure 50. Tubulin + PIH1D3 staining in wild type nasal epithelial cells. PIH1D3 is labelled in red (Alexa Fluor 594), tubulin is in green (Alexa Fluor 488), nucleous are in blue (DAPI).

-Tubulin (axonemal marker) +PIH1D3 in patient 6270 nasal epithelial cells (GVA030)

Tubulin strongly stains in the axoneme as expected. The PIH1D3 is completely missed in patient 6270.

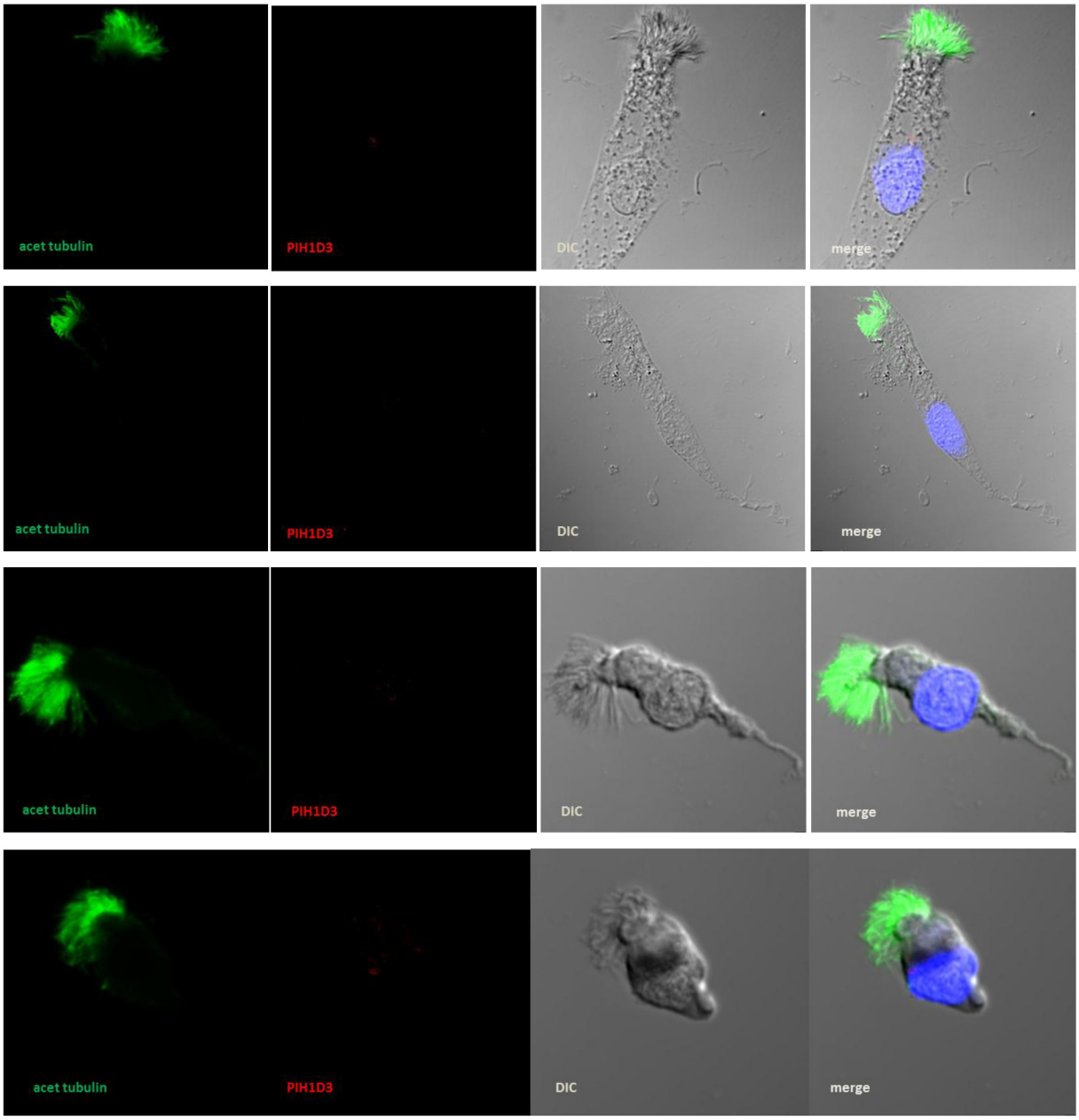


Figure 51. Tubulin + PIH1D3 staining in nasal epithelial cells from patient 6270. Tubulin is labelled in green (Alexa Fluor 488), PIH1D3 is in red (Alexa Fluor 594), nucleous are in blue (DAPI).

-DNAI2 + DNALI1 (IDAs marker) in wild type nasal epithelial cells

DNAI2 strongly stains the outer dynein arms as expected as well as DNALI1 does with inner dynein arms.

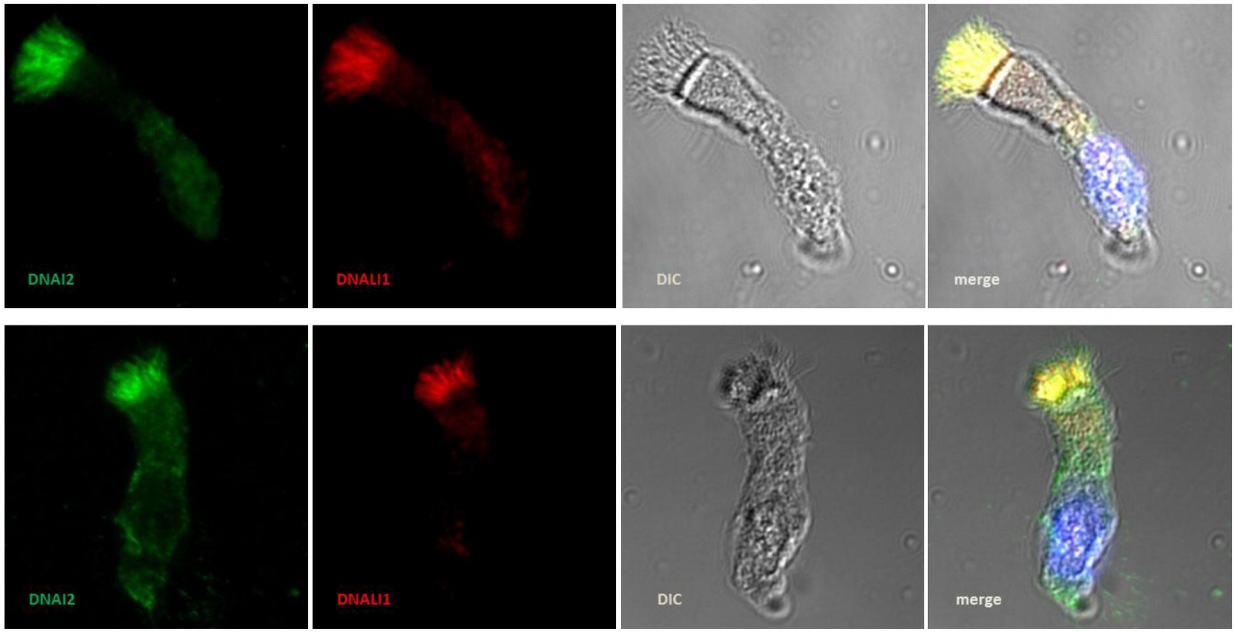


Figure 52. DNAI2 + DNALI1 staining in wild type nasal epithelial cells. DNAI2 is labelled in green (Alexa Fluor 488), DNALI1 is in red (Alexa Fluor 594), nucleous are in blue (DAPI).

-DNAI2 + DNALI1 (IDAs marker) in patient 6270 (GVA030) nasal epithelial cells

Both DNAI2 and DNALI1 are completely absent in patient 6270.

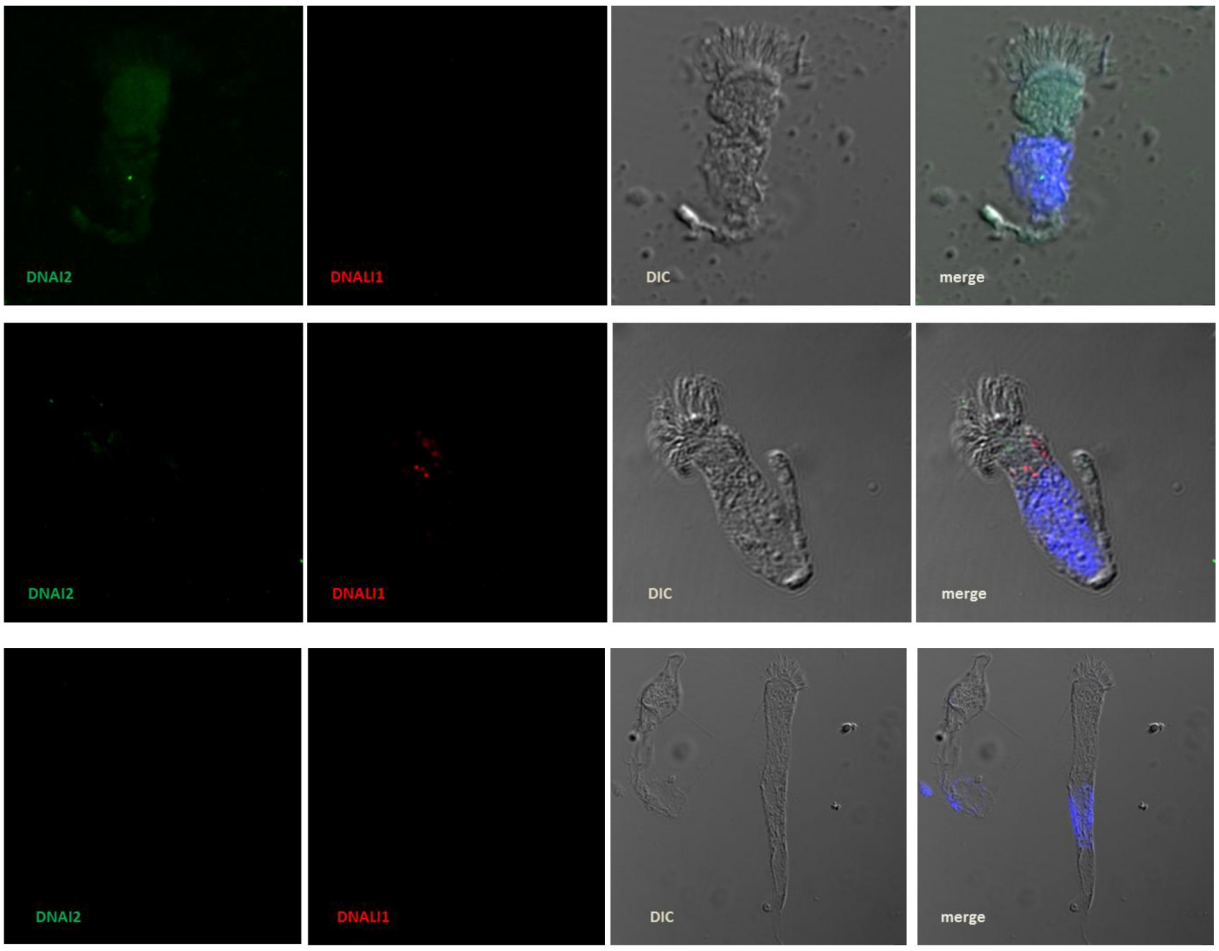


Figure 53. DNAI2 + DNALI1 staining in nasal epithelial cells from patient 6270. DNAI2 is labelled in green (Alexa Fluor 488), DNALI1 is in red (Alexa Fluor 594), nucleous are in blue (DAPI).

-Tubulin + DNAH5 (ODAs marker) in wild type nasal epithelial cells

Tubulin strongly stains the axoneme and DNAH5 ODAs as expected.

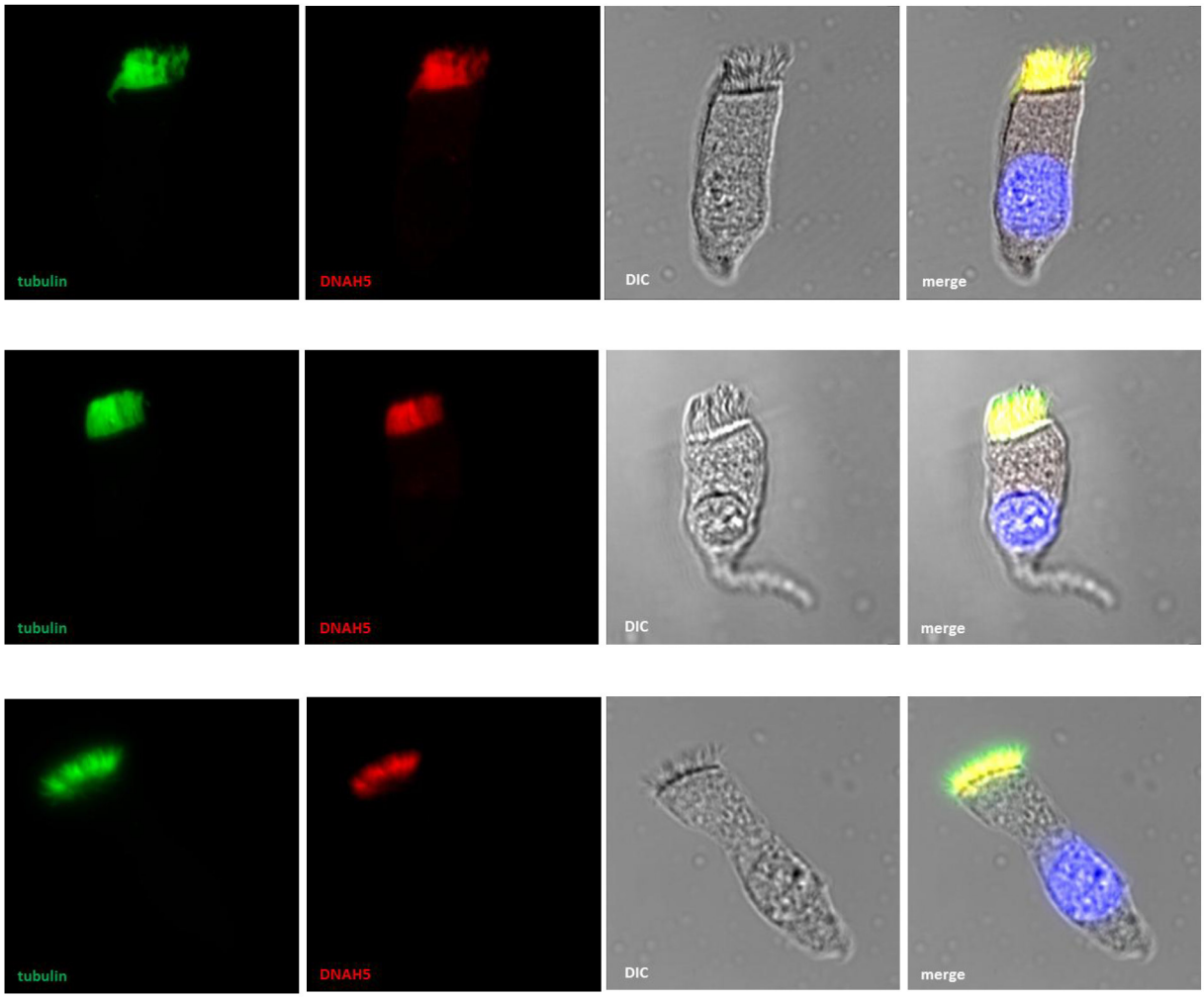


Figure 54. Tubulin + DNAH5 staining in nasal epithelial cells from patient 6270. Tubulin is labelled in green (Alexa Fluor 488), DNAH5 is in red (Alexa Fluor 594), nucleous are in blue (DAPI).

-Tubulin + DNAH5 (ODAs marker) in patient 6270 (GVA030) nasal epithelial cells

Tubulin strongly stains the axoneme as expected. DNAH5 (ODAs marker) is completely absent in patient 6270.

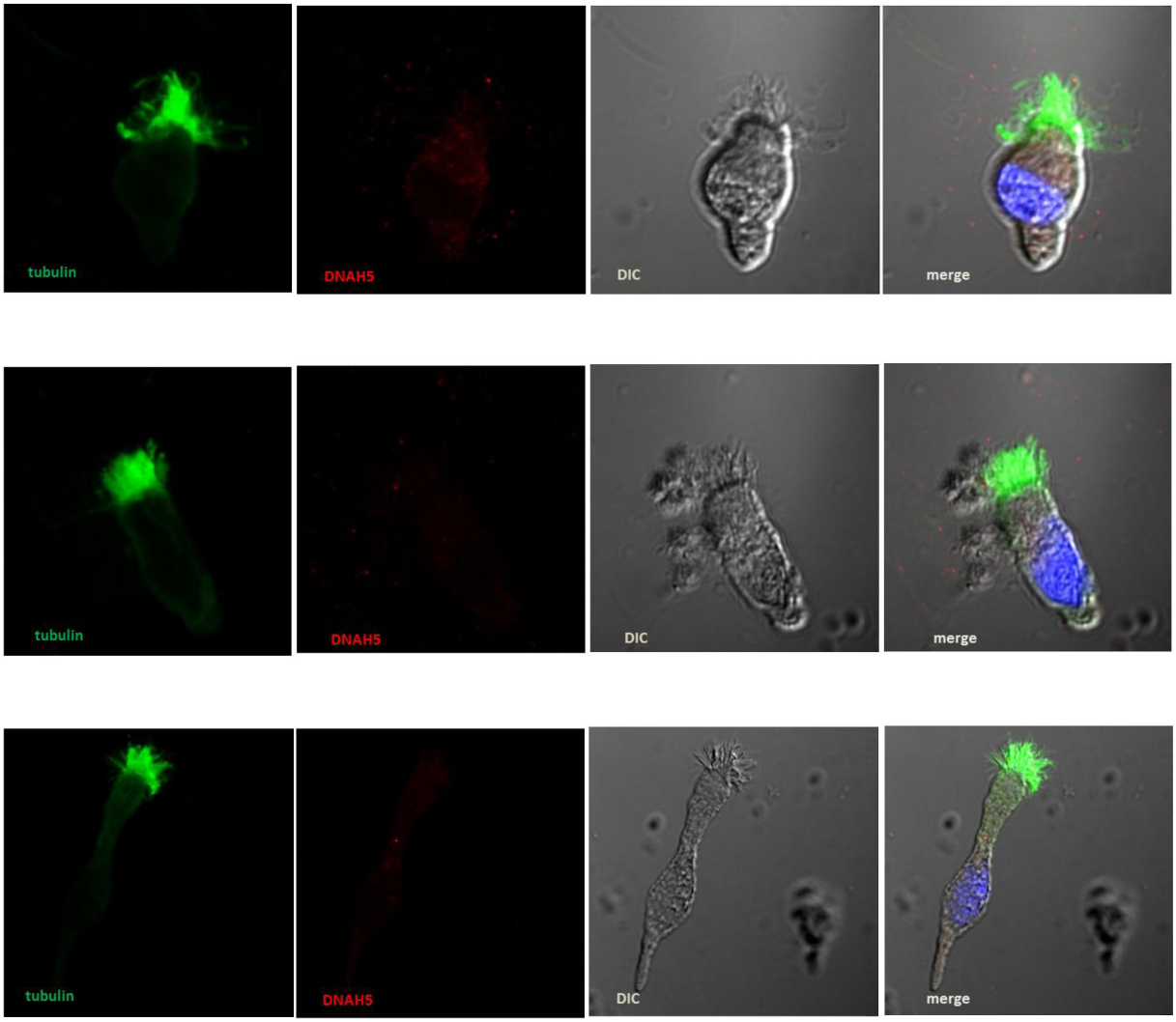


Figure 55. Tubulin + DNAH5 staining in nasal epithelial cells from patient 6270. Tubulin is labelled in green (Alexa Fluor 488), DNAH5 is in red (Alexa Fluor 594), nucleous are in blue (DAPI).

-GM130 (cis-Golgi marker) + PIH1D3 in wild type nasal epithelial cells

GM130 strongly stains the cis-Golgi as expected. PIH1D3 results definitely not co-localize with the cis-Golgi.

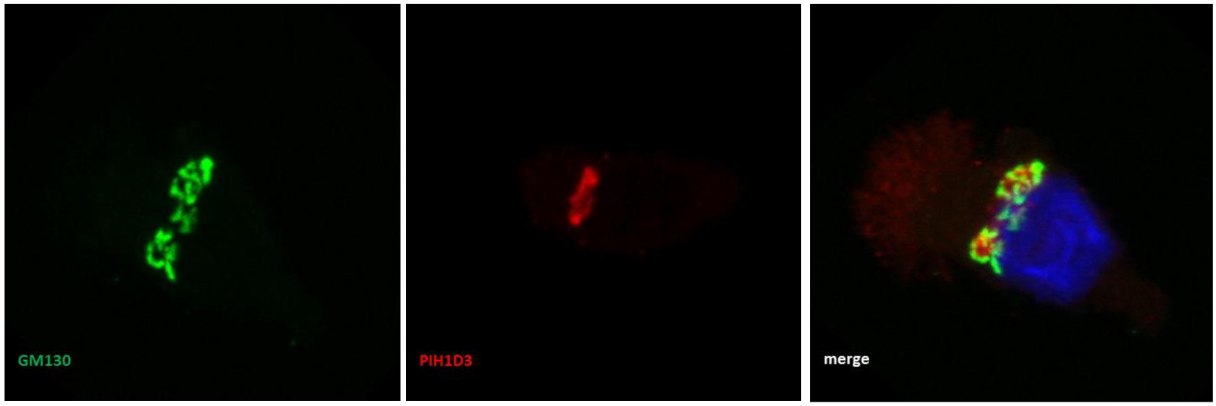


Figure 56. GM130 + PIH1D3 staining in wild type nasal epithelial cells. GM130 is labelled in green (Alexa Fluor 488), PIH1D3 is in red (Alexa Fluor 594), nucleous are in blue (DAPI).

-TGN46 (trans-Golgi marker) + PIH1D3 in wild type nasal epithelial cells

TGN46 strongly stains the trans-Golgi as expected. PIH1D3 results to exactly colocalize with it in the trans-Golgi.

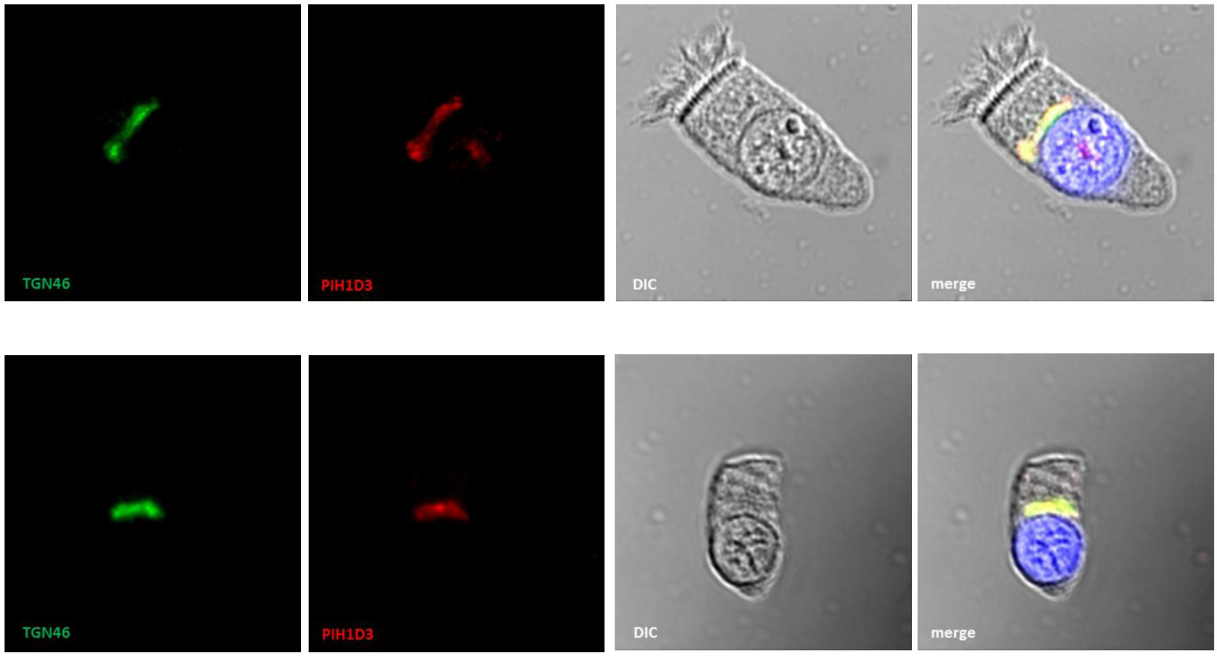


Figure 57. TGN46 + PIH1D3 staining in wild type nasal epithelial cells. TGN46 is labelled in green (Alexa Fluor 488), PIH1D3 is in red (Alexa Fluor 594), nucleous are in blue (DAPI).

-TGN46 + PIH1D3 (trans-Golgi marker) in patient 6270 (GVA030) nasal epithelial cells

TGN46 strongly stains the trans-Golgi as expected. PIH1D3 is totally missed in patient 6270.

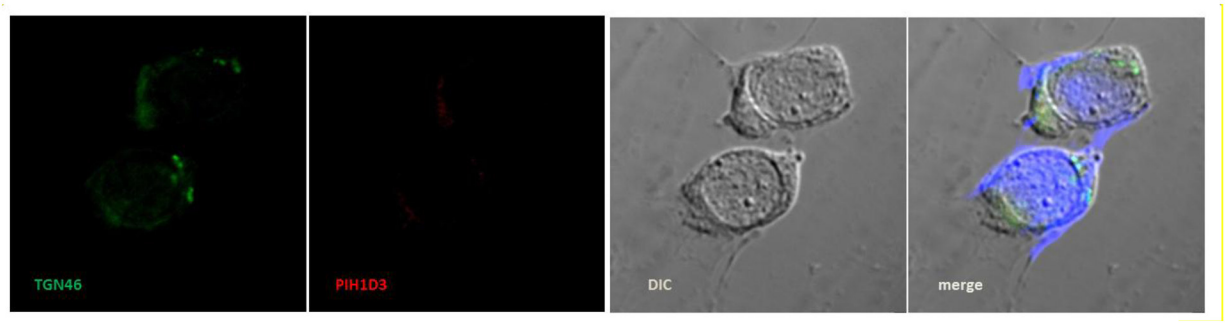


Figure 58. TGN46 + PIH1D3 staining in nasal epithelial cells from patient 6270. TGN46 is labelled in green (Alexa Fluor 488), PIH1D3 is in red (Alexa Fluor 594), nucleous are in blue (DAPI).

Western Blot

Western Blot to detect PIH1D3 in cells from nasal brushing and in Air Liquid Interface (ALI) cells didn't give any clear result yet.

Nasal brushings, Light Microscopy and High Speed Video Method

Nasal brushing was properly performed on patient 6270 (GVA30) and samples were successfully prepared.

HSVM videos are available on request.

Electron Microscopy and Tomography

The EM shows that in patient 6270, who belongs to family GVA030, the CAAT deletion in exon6 of PIH1D3 results in a complete loss of ODAs and IDAs in almost all cilia (figure 60). Only in rare cases, cilia retain only one versus nine outer dynein arms (figure 61).

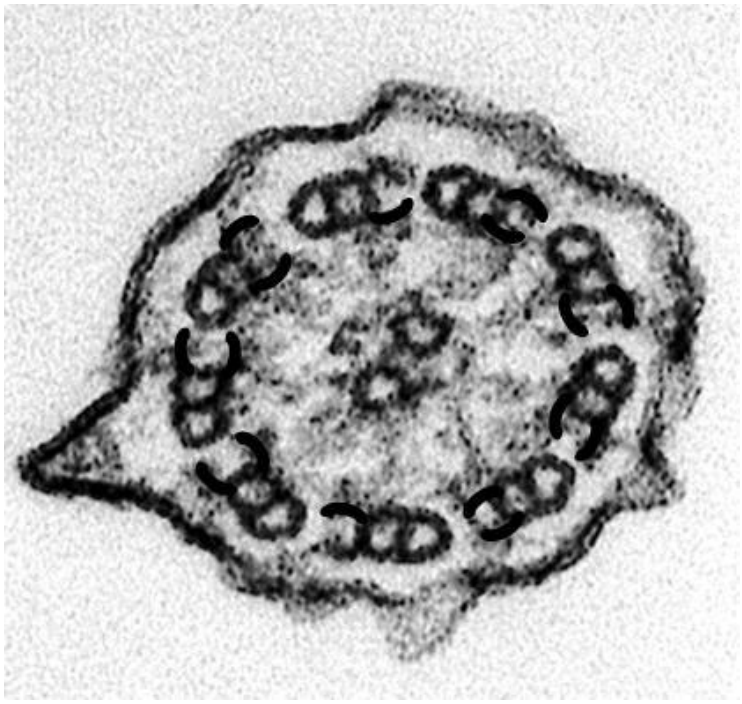
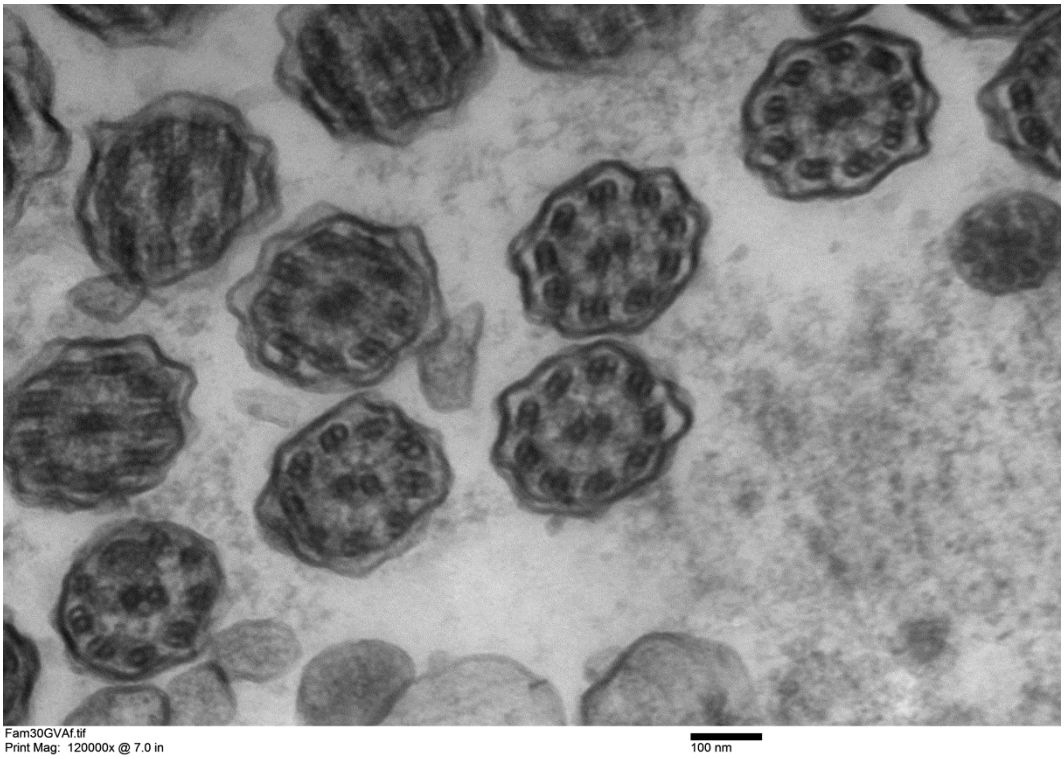


Figure 59. How a wild type 9+2 motile cilium cross-section appears.

A



B

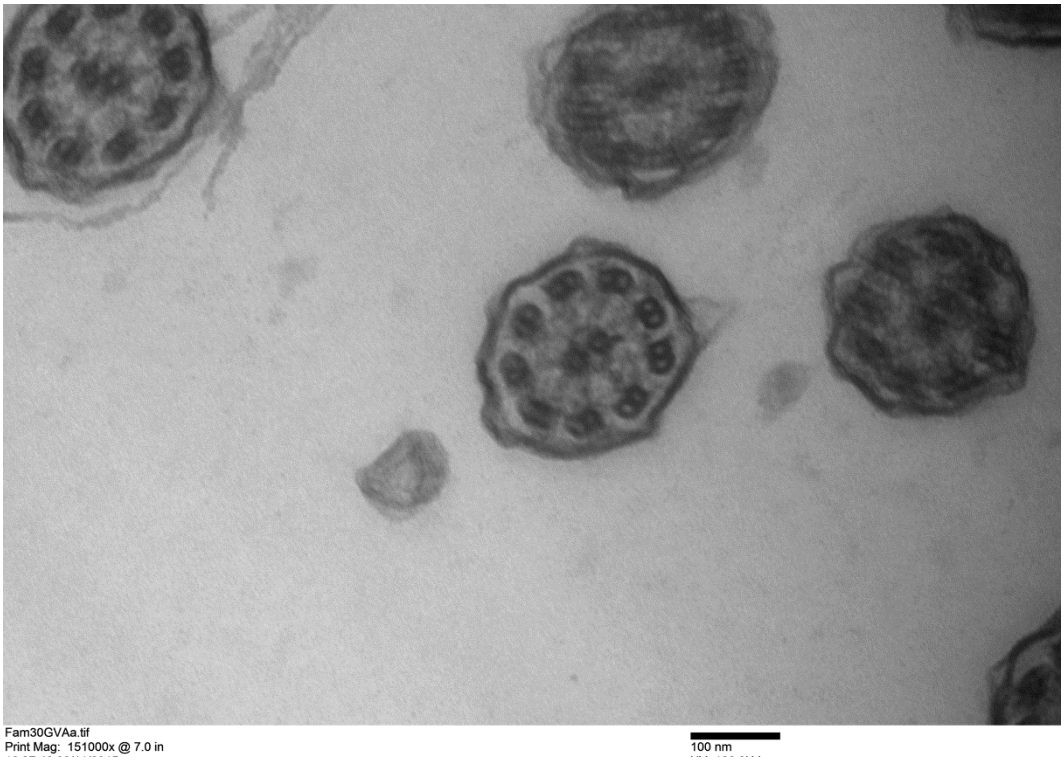


Figure 60. 9+2 Cilia cross section in patient 6270 at lower (A) and higher zoom (B). ODAs and IDAs are completely loss.

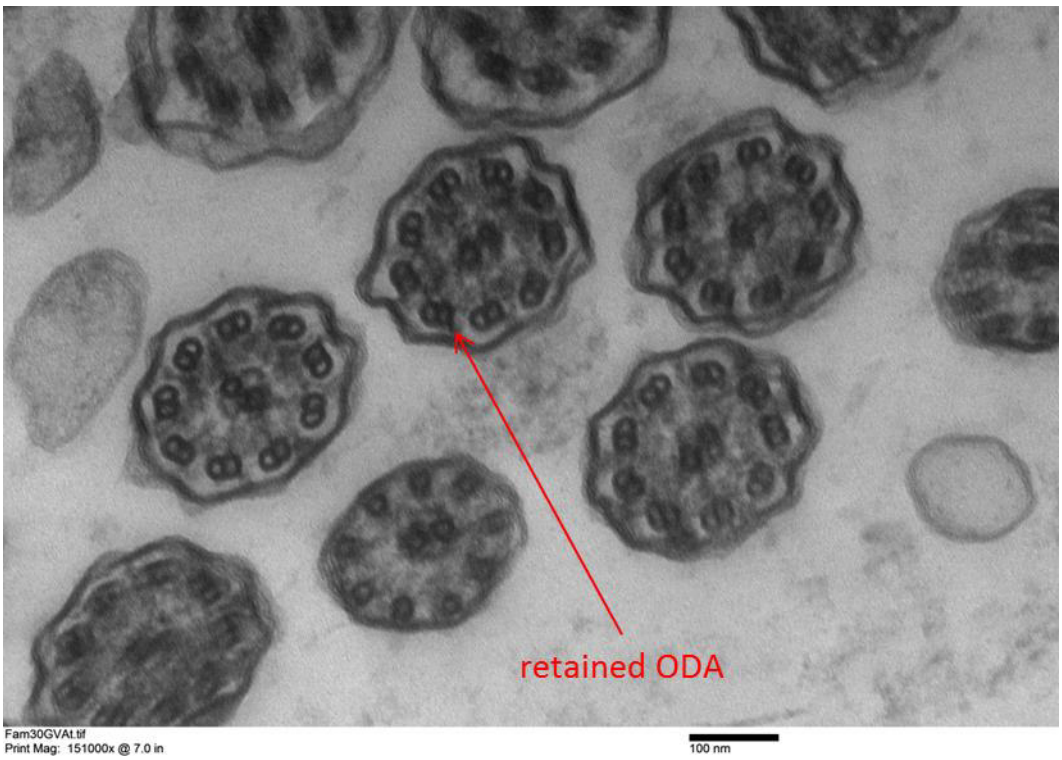


Figure 61. 9+2 Cilia cross-section in patient 6270. In only few cilia 1/9 ODA is retained.

Electron Tomography is on going in this moment. The results will be hopefully available in April 2016.

Family GVA065: possible individuation of a PCD phenocopy: LMTK2

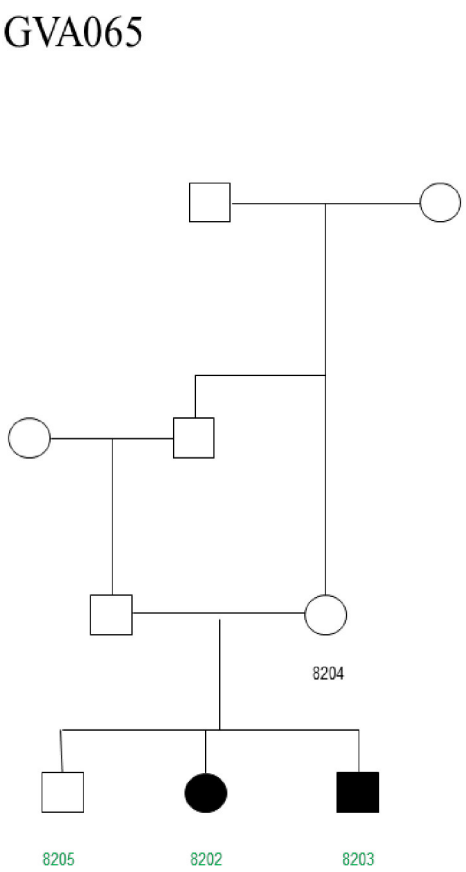


Figure 62. GVA065 genealogical tree.

Family GVA065 has Belgium-Israelite ethnic origin.

Individuals 8202 (sister) and 8203 (brother) are diagnosed as PCD while the mother 8204 and the brother 8205 are considered as healthy.

The 3 children were born from unaffected father and mother. The parents were consanguineous, descending from a common ancestor.

Whole Exome Sequencing was performed on the 3 children DNAs: 2 (8202, 8203) PCD affected and their sibling (8205).

The presence of new mutations or already reported ones in known PCD causative genes was excluded.

The Variant master individuated a good candidate homozygous nonsynonymous mutation in LMTK2 [NM_014916, Gene ID: 22853] where the substitution in exon 4 of the Guanine 379 by an Adenine is predicted to cause an AA switch from Glicine 127 to Arginine [LMTK2:NM_014916:exon4:c.379G>A:p.Gly127Arg].

This mutation in LMTK2, with Annovar Annotate good scores, was selected and investigated.

The observed SNV was not reported nor in dbSNP NCBI database neither in the Exome Variant Server (EVS). The Glycine is a polar (hydrophilic, non-charged) amino acid, meanwhile the Arginine is a basic (positively charged) one. Therefore this different behaviour can impair the

protein 3D folding and action.

PolyPhen2 indicates this SNP as a probably damaging SNV with the maximum score of 1.0 and the GERP Conservation score is 5.82. This means that the mutation is located in a very high conserved domain.

The Sanger sequencing confirmed the mutation to be homozygous in affected (8202, 8203) and heterozygous in health brother (8205) and in the mother (8204). The DNA from father was not available.

LMTK2 (Lemur Tyrosine Kinase 2) gene doesn't code for a protein involved in cilia structure and assembly or in the IFT. But it results to interacts with CFTR (figure 63), known to be mutated in Cystic Fibrosis which can be easily confused with PCD. LMTK2 enzyme, which belongs to protein kinase superfamily, is membrane-anchored kinase that phosphorylates PPP1C, phosphorylase b and CFTR [209].

Interestingly LMTK2 mutations have been also correlated to azoospermia [210, 211].

These findings suggest that LMTK2 mutated individuals from family GVA065 were probably diagnosed as PCD because of the presence of clinical evidences which mimated the PCD phenotype.

Functional studies on LMTK2 and its eventual inclusion in a ciliome panel can help to understand this protein involvement in Cystic Fibrosis and in the development of PCD-like symptoms as well as to create a gene panel for differential diagnosis.

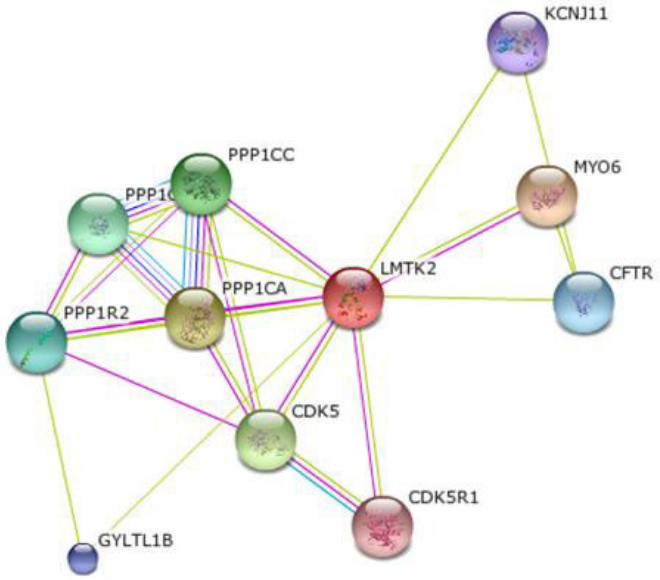


Figure 63. LMTK2 protein-protein interactions. Figures from Sting prediction databases. (http://string-db.org/newstring_cgi/sow_network_section.pl?taskId=MiMrVGSU5_3g&allnodes=1).

Non solved cases

32% of patients does not show any interesting variant, even if in one family (GVA001) the linked region was pointed out with this analysis.
Data interpretation is still in progress for those families in which no candidate gene was found by WES or by haloplex panel analysis.

Two examples (one unsolved family and the family in which the possible locus of the causative gene was pointed out) are reported.

Family GVA001: individuation of a possible linkage interval

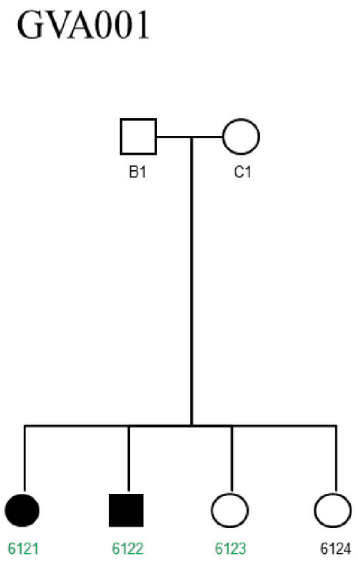


Figure 64. GVA001 genealogical tree.

Family GVA001 was referred for strong suspicion of PCD. This family, from Switzerland, has Turkish ethnic origin. The 4 children (2 males, 2 females) were born from unaffected individuals. Two of them are affected and reported recurrent infections of upper air ways and recurrent bronchopneumonia, but they didn't show situs inversus. Both light and electronic microscopy was performed.
In individual 6121 the cilia beating resulted to be uncoordinated in fresh tissues obtained by biopsy. Furthermore patient cells ciliogenesis resulted in 4% of the central pair absence and untypical cilia movement. Individual 6122 biopsy showed the complete absence of the central pair of microtubule, meanwhile cells from ciliogenesis resulted to be structurally normal but immotile.
DNA was obtained from all the affected and non-affected members of this family including mother and father (6119, 6120).

In order to identify the genetic cause of a possible Primary Ciliary Dyskinesia (PCD), WES was performed on 3 out of 4 brothers/sisters (6121, 6122, 6123): two affected and one healthy.
The presence of new or already reported mutations in known PCD causative genes was excluded.
The Variant Master Software was used to analyze family 001 WES data.
The recessive filter was applied.
A good candidate homozygous nonsynonymous mutation was found in PMFBP1 [NM_031293] where the substitution in exon 5 of the Timine 512 by a Guanine is predicted to cause an AA switch from Isoleucine 171 to Serine [PMFBP1:NM_031293:exon5:c.512T>G:p.Ile171Ser].
This mutation in PMFBP1, with Annovar good scores, was selected and investigated (Gene ID: 83449). The coverage and quality scores resulted appropriate. The allele frequency in the ESP6500 database resulted to be of 0.000077 and the frequency referred to the 1000 Genomes Project of 0.01% (quite high for a rare genetic disease in homozygosis). The observed SNV (rs117953773) in dbSNP ncbi database and in the Exome Variant Server (EVS) resulted not already validated and its minor allele frequency (Citosine) MAF is C=0.0068/34.
Interestingly, as reported in EVS, the C allele was never observed in homozygosis (All Genotype #: CC=0/CA=1/AA=6497).
The Isoleucine is a non-polar (hydrophobic) amino acid, meanwhile the Serine is a polar (hydrophilic, non-charged) one. Therefore this opposite behaviour might impair the protein 3D folding and action.
PolyPhen2 indicates this SNP as a probably damaging SNV with the maximum score of 1.0 and the GERP Conservation score is 6.17. This means that the mutation is located in a very high conserved domain.
The Polyamide modulated factor 1 binding protein [PMFBP1] gene is located on chromosome 16q22.2 and it is transcribed in two alternative mRNAs (NM_001160213.1; NM_031293.2) that are translated in two different protein isoforms (NP_001153685.1; NP_112583.2).
The encoded protein is cytoplasmic and it presents several coiled coil domain. On the basis of UNIProtKB and Swiss-Prot Databases it may affect fertility by playing a role in sperm morphology, especially the sperm tail. It seems also to be involved in the general organization of cytoskeleton.

The Sanger output confirmed the mutation to be homozygous in affected children (6121, 6122) and heterozygous in healthy brothers (6123, 6124).

Because of gene's good scores and functional information, PMFBP1 was proposed as candidate gene for PCD.

We searched mutations in PMFBP1 gene in all sequenced individuals. One healthy family member was found to present the same homozygous mutation than family GVA001 patients. Moreover the ExAC browser, meanwhile developed and soon rised as the best reference tool, reported PMFBP1:NM_031293:exon5:c.512T>G:p.Ile171Ser as a frequent mutation (236 mutated alleles in 121402 counted) which was observed twice also in homozygosity.

Therefore the family GVA001 analysis was restarted. Splicing mutations were also selected, although the difficulty to eventually validate them.
PMFPB1 resulted again the only good candidate.
Linkage studies executed in the past on the same families gave positive results on chr16. This leads to the hypothesis that the candidate mutation might fall near PMFBP1 genomic locus or anyway in chr16. HYDIN, a PCD known genes localizes to chr16. All the changes in HYDIN and PMFBP1 were studied and nothing of interest was found. A big bias in this analysis might be the length of the enourmous HYDIN gene, the potentially incomplete coverage of WES and the difficulty to detect splicing or intronic regulatory mutations.
In order to better define the interval in which the mutated gene might be, all the neighboring variants in homozygosity in the patients were selected and the minimal interval assessed.
At this point, the genome sequencing or the targeted chromosome 16 sequencing might be proposed to find out the molecular defect in this family.

Example of an unsolved family: Family GVA074

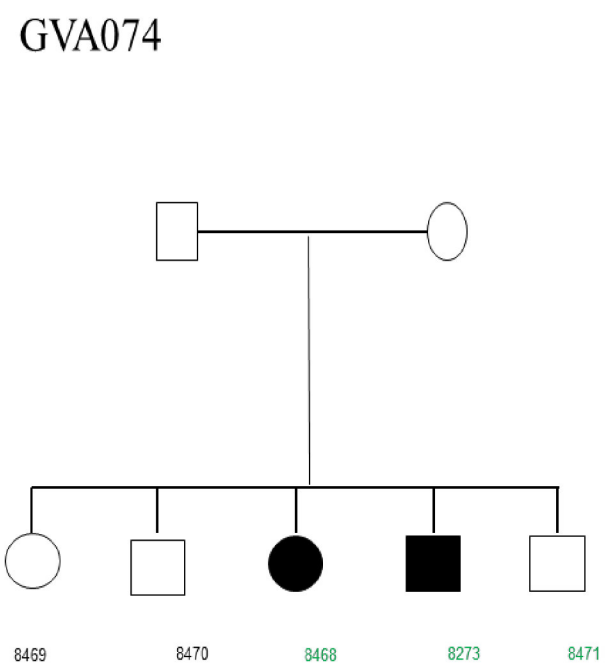


Figure 65. GVA074 genealogical tree.

Family GVA074, from France, has Algerian ethnic origin.

The 5 children (3 males, 2 females) were born from unaffected individuals. Two of them (1 male, 1 female) should be PCD affected. Only one of them presents situs inversus (8273), but they didn't undergo further specific phenotype analyses.

DNA was collected from all the affected and non-affected members of this family, including mother and father. The DNA of the affected individual 8468 underwent WES, while haloplex analysis was performed on affected individual 8273 and unaffected 8471.

The presence of new or already reported mutations in known PCD causative genes was excluded. Moreover Variant Master failed to individuate stong candidate mutations. By manual filtering two composed heterozygous mutation was proposed in SLC26A10 (Gene ID: 65012) (NM_133489; 1: exon9:c.1206G>A:p.Trp402*, stopgain SNV: rs113207856; 2: exon10:c.1247T>G: p.Leu416Arg, nonsynonimous SNV: rs111924104) and in GPR98 [now ADGRV1 (adhesion G protein-coupled receptor V1), Gene ID: 84059] (NM_032119; 1: exon14:c.2596C>T:p.Arg866Trp, nonsynonimous SNV: rs200389929; 2: exon28-:c.6133G>A:p.Gly2045Arg, nonsynonimous SNV: rs41308846).

All of them presented good coverage, quality, frequency and PolyPhen2 scores, but finally the Sanger sequencing did not confirm the real presence of these mutations in affected individual DNAs.

Because of the absence of other exonic nonsynonymous well scored candidate mutations and the heterogeneity of phenotypes, the presence of a splice-disrupting mutation is supposed.

The complete list of our result is shown in Table 6.

Family	number of sequenced patients	known gene		CFTR	nothing	candidate		Family	number of sequenced patients	known gene		CFTR	nothing	candidate
Whole Exome Sequencing								**Haloplex**						
DM_14	1	CCDC39	x					GVA101	1	CCDC40	x			
GVA001	3					x		GVA004	1					x
GVA003	2					x		GVA022	1	DNAH5	x			
GVA005	6					x		GVA002	3	DNAH5	x			
GVA007	3	CCDC40	x					GVA005	4					x
GVA008	1					x		GVA041	1	DNAAF3	x			
GVA010	1	ZMYND10	x					GVA042	1					x
GVA011	1	HYDIN	x					GVA043	1	ARMC4	x			
GVA012	1	DNAH5	x					GVA045	1	DNAAF3	x			
GVA013	1				x			GVA046	1				x	
GVA016	1				x			GVA047	1	CCDC40	x			
GVA024	3	DNAH5	x					GVA048	1	CCDC40	x			
GVA029	1	CCDC40	x					GVA051	1				x	
GVA030	3					x		GVA006	1				x	
GVA031	1	HYDIN	x					GVA052	1				x	
GVA050	1					x		GVA054	1	DNAH5	x			
GVA056	1				x			GVA055	1				x	
GVA059	1	CCDC39	x					GVA058	1	CCDC40	x			
GVA062	1	DNAH5	x					GVA063	1					x
GVA065	3					x		GVA014	1				x	
GVA072	1	HYDIN	x					GVA068	1				x	
GVA074	1				x			GVA069	1					x
GVA075	1					x		GVA071	1	ZMYND10	x			
GVA077	1				x			GVA073	1					x
GVA078	1	SPAG1	x					GVA076	1	HYDIN	x			
GVA079	1	DNAAF1	x					GVA081	1					x
GVA080	1	LRRC6	x					GVA084	1					x
GVA083	1	HYDIN	x					GVA086	1				x	
GVA088	1	DNAH5	x					GVA094	1	DNAH5	x			
GVA090	2			x				GVA036	1				x	
GVA091	2	HYDIN	x					GVA037	1	DNAH5	x			
GVA093	2				x			GVA040	1				x	
GVA097	1				x			GVA099	1			x		
GVA098	1				x			GVA100	1	DNAH5	x			
LB002	1				x			GVA015	1	DNAI1	x		x	
LB005	1	RSPH9	x					GVA028	2				x	
LB024	1					x		GVA033	1				x	
LB032	1				x			GVA067	1				x	
LB074	1				x			GVA023	2			x		

Family	number of sequenced patients	known gene		CFTR	nothing	candidate
Whole Exome Sequencing						
LB081	1	RSPH4A	x			
LB088	1				x	
LB090	1				x	
LB206	1					x
LB230	1				x	
sq11-51	1					x
sq12-06	1				x	

Family	number of sequenced patients	known gene		CFTR	nothing	candidate
Haloplex						
LB022	1				x	
LB029	1	CCDC39	x			
LB051	1	LRRC6	x			
LB019	1	DNAAF1	x			
LB036	1	CCDC39	x			
LB040	1	DNAH5	1			
LB035	1	DNAH5	x			
LB020	1					x
LB021	1				x	
LB041	1	CCDC40	x			
LB045	1				x	
LB050	1					x
LB053	1	LRRC6	x			
LB058	1				x	
LB062	1				x	
LB071	1	DYX1C1	x			
LB077	1					x
LB082	1	DNAH5	x			
LB084	1	DNAAF3	x			
LB103	1					x
LB107	1					x
LB110	1	ARMC4	x			
LB113	1	DNAH5	x			
LB121	1					x
LB145	1				x	
LB147	1				x	
LB150	1	CCDC40	x			
LB160	1	CCDC39	x			
LB167	1				x	
LB168	1				x	
LB169	1	DNAH5	x			
LB170	1				x	
LB171	1				x	
LB172	1				x	
LB174	1			x		
LB199	1				x	
LB201	1				x	

Family	number of sequenced patients	known gene		CFTR	nothing	candidate		Family	number of sequenced patients	known gene		CFTR	nothing	candidate
Whole Exome Sequencing								**Haloplex**						
								LB202	1					x
								LB203	1	CCDC39	x			
								LB204	1	DNAH11	x			
								LB205	1	CCDC40	x			
								LB208	1	DNAH11	x			
								LB232	1	DNAH11	x			
								LB236	1				x	
								LB237	1	DNAH5	x			
								LB239	1				x	
								LB240	1				x	
								LB241	1				x	
								LB255	1				x	
								sq1050	1	DNAAF3	x			

Table 6. **Summary of all results.** Each family ID (columns 1 and 9), the number of individuals sequenced (columns 2 and 10) and the results for each of them (known genes, nothing or candidates:columns 3, 4, 5, 6,7, 11, 12, 13, 14, 15) are shown for Whole Exome Sequencing (columns 1-7) and Haloplex (column 9-15).

Discussion

NGS utility in PCD diagnosis and new candidate genes research

PCD is the perfect model on which NGS method should be applied for diagnostic purpose and to search for new causative genes because of the high genetic heterogeneity. In fact, while Sanger sequencing is perfect if applied to a monogenetic disorder, in diseases with up to several hundreds of (candidate) genes involved, it is not effective [174].
The choice of using WES and a customer designed Haloplex panel allowed the resolution of 65% (known genes+ candidates) of Geneva patient cohort.
Although NGS results to be cost effective and more powerful than Sanger for studying heterogeneous diseaseas, it still presents limitations [212]. The whole exome sequencing excludes all non-coding regions of the genome and accounts only for mutations in coding-regions. Three steps in searching for the causative mutation in each family might be proposed 1) fast screening with a small panel of genes, 2) exome analysis, 3) genome analysis of not solved families. The three steps might be experimental (three different sequencing experiments) or bioinformatics (entire asset of data produced and analysis of just one part of it). To date the cost of genome sequencing is still too high to permit the bioinformatics approach. The targeted bioinformatics approach, however, might lower the risk of unsolicited findings, such as mutations in other genes not related to PCD and involved in other disorders, such as cancer.
Today the average cost of a diagnostic exome analysis is currently about €1,000 for enrichment and sequencing, while a customer designed panel like the one used in this project, is even less expensive. A 640 genes Haloplex panel costs ~€24800 for 96 patients, which means €260 for each individual. On the other hand Sanger sequencing has an average cost of €10 per exon. Heavy chain dynein genes (involved in PCD) are splitted in 60 to 80 exons. Therefore the analysis of just one gene will be almost as expensive as the WES and is a good option only to study the segregation of specific mutations in family members.

Moreover, it is true that the size of the 'medical genome' (the genome that can yield information, which currently may be used in clinical practice) and the genomic portion included in the ENCODE project are expanding. But only the coding sequence of the protein-coding genes, their splice sites, and the known regulatory elements of these genes are now the medical genetics target.
However pathogenic variants exist also in miRNAs, lncRNAs, snRNA, in all the repetitive fraction of the genome and in that DNA part not yet functionally assessed. Therefore, in the next future it will be a better option to analyse the patients with a well described phenotype and an extent familiar pedigree by Whole Genome Sequencing.

Overall prevalence of known genes in the studied cohort

Using WES analysis and a custom targeted panel of ciliary genes, novel mutations were found in the 34 already known genes in several patients.

In 40% of patients likely pathogenic variants in known genes were confirmed. The most frequent mutated gene in the selected cohort is DNAH5 (12.42%), followed by CCDC40 (6.83%), HYDIN (4.35%) and CCDC39 (3.73%). Mutation in other PCD genes accounted each for less than 3%: DNAAF3 (2.48%), DNAH11 (1.86%), LRRC6 (1.86%), ZMYND10 (1.24%), ARMC4 (1.24%) DNAAF1 (1.24%), DYX1C1 (0.62%), DNAI1 (0.62%), SPAG1(0.62%), RSPH9 (0.62%), RSPH4A (0.62%).

These data are different from a recent estimation made by an US group that assessed the known PCD genes prevalence around 60% [150]. This great discordance might be due to some founder effect in the American population.

This findings show that a small panel of known or very probable candidate genes might be useful and less expensive to screen out the 40% of the cases and submit the remaining 60% to the more expensive WES analysis.

Candidate genes

For 17% of patients we identified a strong novel candidate gene.

PIH1D3 was individuated as candidate gene in family GVA30.
This protein has a PIH domain, found also in Pf13/Ktu and MOT48: two genes, involved in dynein preassembly and stability [65, 68].
TWI1 (whose homologous in human seem to be PIH1D3) is a *C. reinhardtii* third conserved PIH protein, a candidate to act at the same level as MOT48 and Pf13/Ktu. This protein is a putative homologue of zebrafih TWISTER protein, shown to be responsible for polycystic kidney and other body defects.

The PIH1 domain was first identifid in the *Saccharomyces cerevisiae* protein PIH1 (also known as Nop17p), that interacts with the molecular chaperone HSP90 to enhance and maintain the assembly of the BOX C/D small nucleolar RNP and the prerRNA-processing complex.

In the same way the PIH1 domain–containing protein PF13 interacts with HSP70. Coimmunoprecipitation analysis in wt mice testis revealed that Pih1d3 also interacts with the molecular chaperons Hsp70 or Hsp90 [65].

The presence of more than one PIH- protein in some species led to hypothize that each of them is specific for a particular part of cilia ultrastructure assembly. PF13/KTU within ODA7 is thought to mainly account for outer arm dynein (ODA HCs folding and stability, HC–IC complex assembly) and MOT48 for inner arm dyneins, but their functions are probably superimposed and not strictly specific [65, 68, 69].

A murine knock out model for pih1d3 was created [64].
In mice there are two genes for PIH1D3, one on mouse chromosome X (E230019M04), probably the real ortholog of human PIH1D3, and an expressed pseudogene (Pih1d3) on chromosome 1, probably derived from the chromosome X gene.
The isoform knocked out in the mouse is the expressed pseudogene on chromosome 1, not present in human and therefore a recent evolutionary addition in mice. For clarity purposes, the chromosome X isoform is referred as Pih1d3x and the transcribed pseudogene on chromosome 1 as Pih1d3tp.

Pih1d3tp is expressed specifically in spermatogenic cells, while Pih3d3x is detectable in lung, brain, oviduct, and testis where motile cilia are present.

Therefore the homozygote null mice for the Pih1d3tp gene did not show abnormalities related to ciliary defects (such as situs inversus or respiratory infection), however males knockout mice manifest infertility. Pih1d3tp(-/-) sperm were immotile with an apparently normal head and a thinner, shorter and fragile flagellum. The flagellar axoneme lacks outer dynein arms (ODAs) and often inner dynein arms (IDAs) and shows a disturbed 9+2 microtubule organization. On the contrary, the motility and ultrastructure of cilia of the respiratory tract are normal. Moreover, Pih1d3tp is not detected in the node at day 8, when the symmetry is broken, and this is consequent with the knockouts never showing situs inversus.
Pih1d3tp is localized to the cytoplasm of spermatogenic cells and it is not present in elongating or elongated spermatids, therefore its action is needed before flagellum formation.

The PIH1 domain containing gene Twister (~54% of similarity to the human gene) was first identified in zebrafish associated to the formation of kidney cysts, as in polycystic kidney disease (PKD) in humans [80]. It has to be noted that kidney cilia in zebrafish are 9+2 motile cilia with ODAs and IDAs, while in mammals they are 9+0 immotile cilia [213].
New data on wild-type and mutated pih1d3 expression are expected from the zebrafish model in a collaborative effort for PIH1D3 validation [80]
Pih1d3 is a cytoplasmic protein. Because of the axonemal dynein complexes are preassembled in the cytoplasm before their transport to cilia, Pih1d3 was hypothesized to contribute to cytoplasmic preassembly of dynein complexes in spermatogenic cells by stabilizing and promoting complex formation by ODA and IDA proteins. The preassembly mechanism and proteins interaction at this level is still unclear and functional studies on involved proteins are quite hard.
The precise mechanism of ODA and IDA preassembly is unclear, as well as the exact composition of ODAs and IDAs. It is possible that they may differ between sperm flagella and motile cilia of the trachea. To date, five proteins have been identified that are required for cytoplasmic preassembly of dynein complexes: DNAAF2 (PF13 or Ktu or C14orf104 [65]), DNAAF1 (ODA7 or LRRC50 [66, 67]), MOT48 (PIH1D1 [68]), DNAAF3 (PF22, [69]) and PIH1D3tp ([64]).
The preassembly of dynein HCs require at least two steps [68]. The first step is required for HC stability and involves the folding of the globular dynein head domain. The second step is the formation of the HC–IC complex.
DNAAF2 and DNAAF1 (together with HSP70) are required for folding and stability of the three ODA heavy chains in the cytoplasm, DNAAF3 is required for preassembly of ODAs and IDAs, and MOT48 for preassembly of IDAs.
Pih1d3tp forms a complex with Dnaic2 (homologous to ODA IC2) as well as with Hsp70 and Hsp90, suggesting that it may function as a co-chaperone to help maintain the stability of dynein HCs and ICs in both ODA and IDA. However, Pih1d3tp was not found to be associated with Dnaaf1, Dnaaf2, and Dnaaf3, suggesting that Pih1d3tp functions at a different step during the assembly of dynein complex [64].

Being such a good candidate gene, PIH1D3 was Sanger sequenced in 12 additional families with a potentially X-linked segregation. Intresting pathological mutations were pointed out in two patients supporting the role of PIH1D3 in PCD. A collaborative effort with UCL, discovered two additional patients in the UK cohort.
These findings pushed to re-evaluate the phenotype of the twister zebrafish mutant and to proceed with functional studies in order to understand the real PIH1D3 function.

By Immunofluorescence analysis PIH1D3 resulted to localize in the cytoplasm. This finding was confirmed many times by staining wild type brush epithelial nasal cells for PIH1D3 together with tubulin (axonemal marker), DNALI1(IDAs marker), DNAI2 and DNAH5 (ODAs markers). The evidence of a cytoplasmic localization of PIH1D3 fits with the hypothesis of its involvement in cilia assembly.

PIH1D3 staining appeared to be very close to the nucleous. Therefore it was co-immunostained with GM130 and TGN46 which mark cis-Golgi and trans-Golgi respectively. Interestingly PIH1D3 resulted to exactly colocalize with the trans-Golgi. These data have to be carefully focused to choose the best for further studies.
All the stainings were performed in parallel in wild type and in patient 6270 nasal epithelial cells. PIH1D3 was always found to be totally absent in patient 6270. Moreover the patient cells lack completely DNAI2, DNALI1 and DNAH5.

The immunofluorescence results perfectly match to the EM observations. In fact, TEM pictures showed the complete loss of both ODAs and IDAs in almost all cilia.
These finding, although the exact step at which PIH1D3 is involved in cilia preassembly is still unknown, lead to define PIH1D3 as a novel PCD gene.

LMTK2 (Lemur Tyrosine Kinase 2) enzyme phosphorylates CFTR and its mutations have been linked to azoospermia. LMTK2 mutation found in family GVA065 were probably diagnosed as PCD because of the presence of clinical evidences which mimated the PCD phenotype. Therefore functional studies on LMTK2 and its eventual inclusion in a ciliome panel can help to understand this protein involvement in Cystic Fibrosis and in the development of PCD-like symptoms as well as to create a gene panel for differential diagnosis.

Other five PCD candidates are being considered to pass the functional validation.

Unsolved families

In 32% of patients no candidate genes were pointed out by this analysis.
This percentage cam be explained by the fact that current exome sequencing protocols are far from perfect and only ~90% of the genes is reliably covered.
Moreover, in this analysis we disregarded synonymous variants, while they can hide the introduction of cryptic splice sites.
Furthermore genetics defines the individual genotype, but it can't be precise in referring to phenotype, because of epigenetics. Even in the same family, siblings can have various and different symptoms. The presence of undiagnosed mild phenotypes is also possible. Therefore familiars of PCD patient are sometimes considered healthy, whereas they are asymptomatic. This event can lead to erroneous segregation analysis and to the failure of Variant Master or similar softwares use. Therefore these unsolved families have to be reanalyzed from the clinical point of view. The heterogeneity of PCD symptoms often results in a difficultous diagnosis process. It seems now to be essential the standardization of PCD management and diagnosis criteria across countries. The England system constitutes a good model. In fact the comparison between phenotypical and genomic data results to be less easy and effective if patient samples are collected everywhere by disparate protocols and after diagnosis which proceed on different evaluations of the clinical evidences.
Even in patients belonging to the same family the phenotypes can be highly heterogeneous. So far it is not unusual to confuse PCD with orher air-ways related diseases such as Cystic Fibrosis. Therefore NGS can help to detect and fill eventual gaps of the clinical diagnosis.

On the other hand, NGS bioinformatics analysis is fastly changing and these results has to be reanalyzed over and over. During only the last year ExAc became the most important human variant database, overgoing Exome Variant Server and NCBI SNV data. Furthermore during this study, the Human Genome Dec. 2013 (GRCh38/hg38) Assembly substituted the Human Feb. 2009 (GRCh37/hg19) Assembly in UCSC Genome Browser. This obliged and will oblige to rename and re-localize all the found mutations.

The families in which nothing was found will be analysed again using softwares and pipeline recently developed or upgraded.

The function of many mutated genes individuated by WES is unknown. This is a difficulty in the case of singletons. It is impossible to infer anything or to hypotise the phatogenity of interesting mutations in genes unknown and/or not clearly related to cilia structure and function if they are present only in one individual. On the contrary, the presence of a large pedigree and the aquisition

of other family members should help by permitting a segregation analysis and the use of predictive softwares such us Variant Master.

To date, nothing can be done to further analyse haloplexed patients who didn't show interesting mutations in the 640 targeted genes. But those who have sibling data available and a strong PCD phenotype can be again sequenced by WES.

Although the presence of many limitations in NGS methods and in their analysis effectiveness, the search for candidate genes appears now much less labor intensive with various Unix commands and tools than manual filtering within the 10'000 lines of an excel file. Whole genome sequencing, which allows to escape the drawbacks of DNA selection and will become reasonably feasible and affordable, will provide a further hope in finding remaining genes that escaped sofar to hunting in this and other "negative" families. It appears that it is also possible to analyze families with recessive disease with one affected only, if data from non-affecteds are available. Therefore the analysis can be extended to families where DNA sample of a second affected is missing or in which only one affected is present.

Perspective of future research

1_The 32% of the analysed patients did not show any interesting variant, even if in one family (GVA001) the linked region was pointed out with this analysis. The families in which nothing was found will be analysed again using softwares and pipeline recently developed or upgraded. Moreover we are triying to develop a method to bionformatically individuate variants that might create new splicing sites.

2_Patient cohort will be extended.
Many other individual samples are at disposal for Whole Exome Sequencing. New additional family member will be recruited and additional DNA samples will be analysed to search for additional disease-causing mutations in the unsolved families.

3_PCD diagnosis reassessment
3% of individuals were found mutated in CFTR and one family resulted to be mutated in LMTK2 which seems associated with CFTR and Cystic Fibrosis Therefore a complete revision of the clinical data of unsolved patients is ongoing by using recently developed and more effective diagnostic tools.

4_ Potential novel genes will be investigated.
In few families candidate pathogenetic variants were found in genes encoding for dyneins gene and therefore strictly associated with PCD. The segregation in these families was assessed and confirmed.

To date other PCD patient cohort are searched for additional disease-causing variants in these genes. The future studies will aim to individuate the functional and structural role of the protein encoded by these candidate genes. The used research tools will be immunofluorescence on nasal ciliated cells, high speed video records by light microscopy and pictures collection by confocal microscopy, conventional electron microscopy (TEM) and electron tomography (ET) starting from patient nasal brushing. IF will be performed both in patients and in ciliated cells from healthy individuals. The use of antibodies against the candidate protein will permit to check the localization of their cellular expression and whether this is altered in patients. Electron Tomography (ET) uses 2D projection images, collected by 'photographing' the targeted object from many different points of view, to profile its 3D image. The 3D reconstruction will allow the comparation of the detailed pictures of cilia ultrastructure and macromolecular complexes obtained from patients with expected healthy cilia structure. Moreover the PIH1D3 study sets the bases for fruitful collaborations with a zebrafish mutant expert group that will be involved also in the analysis of the other candidate genes.
Describing a novel gene and mutations involved in PCD will increase the diagnostic power that is available for patients, and shed light on the underlying cell biological defects in PCD.

Conclusion

This is the largest cohort of patients to date studied with high throughput sequencing technologies.
The next-generation exome sequencing and the Haloplex panel allowed the individuation of mutations in the PCD known genes in a total of 40% of the patients. In another 25% of patients new strong candidates were found. They will pass validation. In ~3% of individuals CFTR was found mutated suggesting a cystic fibrosis which was wrongly diagnosed as PCD.
For one family we pointed out the chromosomal region hypotised to be linked to the disease.
In 32% of patients no candidate gene was pointed out by this analysis. This might be due to the not perfect coverage of the selection method, to the involvement of other genes that we did not include in the Haloplex panel, to a too selective bioinformatics analysis or to a not perfect clinical evaluation of the patients and family members.
A good achievement was the discovery of the involvement of PIH1D3 in PCD, which is now submitted for publication in a collaborative paper.

Based on this study, the more effective diagnostic approach consists in sequential steps:
1-bioinformatics analysis of the known genes (40% of the patients solved);
2-Genetics analysis of ciliome genes panel (25% genes found);
3-WES analysis with Variant Master (30% patients solved).

The 32% of patient cannot still be explained by the known PCD genes: therefore a large window for gene exploration remains.

References

1. Badano, J.L., et al., *The ciliopathies: an emerging class of human genetic disorders*. Annu Rev Genomics Hum Genet, 2006. **7**: p. 125-48.
2. Habbig, S. and M.C. Liebau, *Ciliopathies - from rare inherited cystic kidney diseases to basic cellular function.* Mol Cell Pediatr, 2015. **2**(1): p. 8.
3. Drummond, I.A., *Cilia functions in development.* Curr Opin Cell Biol, 2012. **24**(1): p. 24-30.
4. Fliegauf, M., T. Benzing, and H. Omran, *When cilia go bad: cilia defects and ciliopathies.* Nat Rev Mol Cell Biol, 2007. **8**(11): p. 880-93.
5. Brown, J.M. and G.B. Witman, *Cilia and Diseases*, in *Bioscience*. 2014. p. 1126-1137.
6. Kobayashi, D. and H. Takeda, *Ciliary motility: the components and cytoplasmic preassembly mechanisms of the axonemal dyneins*, in *Differentiation*. 2012. p. S23-9.
7. Leigh, M.W., et al., *Clinical and genetic aspects of primary ciliary dyskinesia/Kartagener syndrome.* Genet Med, 2009. **11**(7): p. 473-87.
8. Hogg, C., *Primary ciliary dyskinesia: when to suspect the diagnosis and how to confirm it.* Paediatr Respir Rev, 2009. **10**(2): p. 44-50.
9. Gudis, D.A. and N.A. Cohen, *Cilia dysfunction*, in *Otolaryngol Clin North Am*. 2010. p. 461-72, vii.
10. Rosenbaum, J.L. and G.B. Witman, *Intraflagellar transport.* Nat Rev Mol Cell Biol, 2002. **3**(11): p. 813-25.
11. Oh, E.C. and N. Katsanis, *Cilia in vertebrate development and disease.* Development, 2012. **139**(3): p. 443-8.
12. Wheatley, D.N., *Landmarks in the first hundred years of primary (9+0) cilium research.* Cell Biol Int, 2005. **29**(5): p. 333-9.
13. Yoshimura, K., T. Kawate, and S. Takeda, *Signaling through the primary cilium affects glial cell survival under a stressed environment.* Glia, 2011. **59**(2): p. 333-44.
14. Yoshimura, K. and S. Takeda, *Hedgehog signaling regulates myelination in the peripheral nervous system through primary cilia.* Differentiation, 2012. **83**(2): p. S78-85.
15. Shi, Z.D. and J.M. Tarbell, *Fluid flow mechanotransduction in vascular smooth muscle cells and fibroblasts.* Ann Biomed Eng, 2011. **39**(6): p. 1608-19.
16. Christensen, S.T. and C.M. Ott, *Cell signaling. A ciliary signaling switch.* Science, 2007. **317**(5836): p. 330-1.
17. Christensen, S.T., et al., *Sensory cilia and integration of signal transduction in human health and disease.* Traffic, 2007. **8**(2): p. 97-109.
18. Hawkins, T., et al., *Mechanics of microtubules.* J Biomech, 2010. **43**(1): p. 23-30.
19. Fujiu, K., et al., *Mechanoreception in motile flagella of Chlamydomonas.* Nat Cell Biol, 2011. **13**(5): p. 630-2.
20. Porter, M.E. and W.S. Sale, *The 9 + 2 axoneme anchors multiple inner arm dyneins and a network of kinases and phosphatases that control motility.* J Cell Biol, 2000. **151**(5): p. F37-42.
21. Nicastro, D., et al., *The molecular architecture of axonemes revealed by cryoelectron tomography.* Science, 2006. **313**(5789): p. 944-8.
22. Chasey, D., *The three-dimensional arrangement of radial spokes in the flagella of Chlamydomonas reinhardii.* Exp Cell Res, 1974. **84**(1): p. 374-80.
23. Dentler, W.L. and W.P. Cunningham, *Structure and organization of radial spokes in cilia of Tetrahymena pyriformis.* J Morphol, 1977. **153**(1): p. 143-51.
24. Dymek, E.E. and E.F. Smith, *A conserved CaM- and radial spoke associated*

complex mediates regulation of flagellar dynein activity. J Cell Biol, 2007. **179**(3): p. 515-26.
25. Goodenough, U.W. and J.E. Heuser, *Substructure of inner dynein arms, radial spokes, and the central pair/projection complex of cilia and flagella.* J Cell Biol, 1985. **100**(6): p. 2008-18.
26. Barber, C.F., et al., *Three-dimensional structure of the radial spokes reveals heterogeneity and interactions with dyneins in Chlamydomonas flagella.* Mol Biol Cell, 2012. **23**(1): p. 111-20.
27. Piperno, G. and D.J. Luck, *An actin-like protein is a component of axonemes from Chlamydomonas flagella.* J Biol Chem, 1979. **254**(7): p. 2187-90.
28. Piperno, G., K. Mead, and W. Shestak, *The inner dynein arms I2 interact with a "dynein regulatory complex" in Chlamydomonas flagella.* J Cell Biol, 1992. **118**(6): p. 1455-63.
29. Huang, B., et al., *Purification and characterization of a basal body-associated Ca2+-binding protein.* J Cell Biol, 1988. **107**(1): p. 121-31.
30. Oda, T., H. Yanagisawa, and M. Kikkawa, *Detailed structural and biochemical characterization of the nexin-dynein regulatory complex.* Mol Biol Cell, 2015. **26**(2): p. 294-304.
31. Huang, B., Z. Ramanis, and D.J. Luck, *Suppressor mutations in Chlamydomonas reveal a regulatory mechanism for Flagellar function.* Cell, 1982. **28**(1): p. 115-24.
32. Mitchell, D.R., *Speculations on the evolution of 9+2 organelles and the role of central pair microtubules.* Biol Cell, 2004. **96**(9): p. 691-6.
33. Smith, E.F. and P. Yang, *The radial spokes and central apparatus: mechano-chemical transducers that regulate flagellar motility.* Cell Motil Cytoskeleton, 2004. **57**(1): p. 8-17.
34. Sturgess, J.M. and J. Chao, *Ultrastructural features of a human genetic defect of cilia.* Prog Clin Biol Res, 1982. **80**: p. 7-12.
35. Neugebauer, D.C., et al., *'9 + 0' axoneme in spermatozoa and some nasal cilia of a patient with totally immotile spermatozoa associated with thickened sheath and short midpiece.* Hum Reprod, 1990. **5**(8): p. 981-6.
36. Nonaka, S., et al., *Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein.* Cell, 1998. **95**(6): p. 829-37.
37. Marszalek, J.R., et al., *Situs inversus and embryonic ciliary morphogenesis defects in mouse mutants lacking the KIF3A subunit of kinesin-II.* Proc Natl Acad Sci U S A, 1999. **96**(9): p. 5043-8.
38. Nonaka, S., et al., *Determination of left-right patterning of the mouse embryo by artificial nodal flow.* Nature, 2002. **418**(6893): p. 96-9.
39. Gerdes, J.M., E.E. Davis, and N. Katsanis, *The vertebrate primary cilium in development, homeostasis, and disease*, in *Cell*. 2009. p. 32-45.
40. Farnum, C.E., R.M. Williams, and E. Donnelly, *Analyzing primary cilia by multiphoton microscopy.* Methods Cell Biol, 2009. **94**: p. 117-35.
41. Perkins, L.A., et al., *Mutant sensory cilia in the nematode Caenorhabditis elegans.* Dev Biol, 1986. **117**(2): p. 456-87.
42. Afzelius, B.A., *A human syndrome caused by immotile cilia.* Science, 1976. **193**(4250): p. 317-9.
43. Pennarun, G., et al., *The human dynein intermediate chain 2 gene (DNAI2): cloning, mapping, expression pattern, and evaluation as a candidate for primary ciliary dyskinesia.* Hum Genet, 2000. **107**(6): p. 642-9.
44. Guichard, C., et al., *Axonemal dynein intermediate-chain gene (DNAI1) mutations result in situs inversus and primary ciliary dyskinesia (Kartagener syndrome).* Am J Hum Genet, 2001. **68**(4): p. 1030-5.

45. Bartoloni, L., et al., *Mutations in the DNAH11 (axonemal heavy chain dynein type 11) gene cause one form of situs inversus totalis and most likely primary ciliary dyskinesia.* Proc Natl Acad Sci U S A, 2002. **99**(16): p. 10282-6.
46. Olbrich, H., et al., *Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry.* Nat Genet, 2002. **30**(2): p. 143-4.
47. Olbrich, H., et al., *Recessive HYDIN mutations cause primary ciliary dyskinesia without randomization of left-right body asymmetry.* Am J Hum Genet, 2012. **91**(4): p. 672-84.
48. Bisgrove, B.W., et al., *Polaris and Polycystin-2 in dorsal forerunner cells and Kupffer's vesicle are required for specification of the zebrafish left-right axis.* Dev Biol, 2005. **287**(2): p. 274-88.
49. Yoshiba, S., et al., *Cilia at the node of mouse embryos sense fluid flow for left-right determination via Pkd2.* Science, 2012. **338**(6104): p. 226-31.
50. Bataille, S., et al., *Association of PKD2 (polycystin 2) mutations with left-right laterality defects.* Am J Kidney Dis, 2011. **58**(3): p. 456-60.
51. Onoe, T., et al., *Situs inversus and cystic kidney disease: Two adult patients with this Heterogeneous syndrome.* Am J Case Rep, 2013. **14**: p. 20-5.
52. Oka, M., T. Mochizuki, and S. Kobayashi, *A novel mutation of the PKD2 gene in a Japanese patient with autosomal dominant polycystic kidney disease and complete situs inversus.* Am J Kidney Dis, 2014. **64**(4): p. 660.
53. Kumar, S., B. Nanjappa, and Y.R. Barapatre, *Autosomal dominant polycystic kidney disease with situs inversus.* Urology, 2012. **80**(2): p. e23-4.
54. Johnson, K.A. and J.L. Rosenbaum, *Polarity of flagellar assembly in Chlamydomonas.* J Cell Biol, 1992. **119**(6): p. 1605-11.
55. Kozminski, K.G., et al., *A motility in the eukaryotic flagellum unrelated to flagellar beating.* Proc Natl Acad Sci U S A, 1993. **90**(12): p. 5519-23.
56. Piperno, G. and K. Mead, *Transport of a novel complex in the cytoplasmic matrix of Chlamydomonas flagella.* Proc Natl Acad Sci U S A, 1997. **94**(9): p. 4457-62.
57. Cole, D.G., *The intraflagellar transport machinery of Chlamydomonas reinhardtii.* Traffic, 2003. **4**(7): p. 435-42.
58. Baker, S.A., et al., *IFT20 links kinesin II with a mammalian intraflagellar transport complex that is conserved in motile flagella and sensory cilia.* J Biol Chem, 2003. **278**(36): p. 34211-8.
59. Pedersen, L.B. and J.L. Rosenbaum, *Intraflagellar transport (IFT) role in ciliary assembly, resorption and signalling.* Curr Top Dev Biol, 2008. **85**: p. 23-61.
60. Blacque, O.E., S. Cevik, and O.I. Kaplan, *Intraflagellar transport: from molecular characterisation to mechanism.* Front Biosci, 2008. **13**: p. 2633-52.
61. Blacque, O.E., et al., *Functional genomics of the cilium, a sensory organelle.* Curr Biol, 2005. **15**(10): p. 935-41.
62. Blacque, O.E., et al., *Loss of C. elegans BBS-7 and BBS-8 protein function results in cilia defects and compromised intraflagellar transport.* Genes Dev, 2004. **18**(13): p. 1630-42.
63. Plotnikova, O.V., E.N. Pugacheva, and E.A. Golemis, *Primary cilia and the cell cycle.* Methods Cell Biol, 2009. **94**: p. 137-60.
64. Dong, F., et al., *Pih1d3 is required for cytoplasmic preassembly of axonemal dynein in mouse sperm.* J Cell Biol, 2014. **204**(2): p. 203-13.
65. Omran, H., et al., *Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins.* Nature, 2008. **456**(7222): p. 611-6.
66. Duquesnoy, P., et al., *Loss-of-function mutations in the human ortholog of Chlamydomonas reinhardtii ODA7 disrupt dynein arm assembly and cause primary ciliary dyskinesia.* Am J Hum Genet, 2009. **85**(6): p. 890-6.
67. Loges, N.T., et al., *Deletions and point mutations of LRRC50 cause primary ciliary*

dyskinesia due to dynein arm defects. Am J Hum Genet, 2009. **85**(6): p. 883-9.
68. Yamamoto, R., M. Hirono, and R. Kamiya, *Discrete PIH proteins function in the cytoplasmic preassembly of different subsets of axonemal dyneins.* J Cell Biol, 2010. **190**(1): p. 65-71.
69. Mitchison, H.M., et al., *Mutations in axonemal dynein assembly factor DNAAF3 cause primary ciliary dyskinesia.* Nat Genet, 2012. **44**(4): p. 381-9, S1-2.
70. Kakihara, Y. and W.A. Houry, *The R2TP complex: discovery and functions.* Biochim Biophys Acta, 2012. **1823**(1): p. 101-7.
71. Golinska, K., *Regulation of ciliary pattern in Dileptus (Ciliata). I. Sensory cilia and their conversion into locomotor cilia.* J Embryol Exp Morphol, 1982. **68**: p. 99-114.
72. Bloodgood, R.A., *Sensory reception is an attribute of both primary cilia and motile cilia.* J Cell Sci, 2010. **123**(Pt 4): p. 505-9.
73. Lehman, J.M., et al., *An essential role for dermal primary cilia in hair follicle morphogenesis.* J Invest Dermatol, 2009. **129**(2): p. 438-48.
74. Gluenz, E., et al., *Beyond 9+0: noncanonical axoneme structures characterize sensory cilia from protists to humans.* FASEB J, 2010. **24**(9): p. 3117-21.
75. Rohatgi, R. and W.J. Snell, *The ciliary membrane.* Curr Opin Cell Biol, 2010. **22**(4): p. 541-6.
76. Satir, P. and S.T. Christensen, *Overview of structure and function of mammalian cilia.* Annu Rev Physiol, 2007. **69**: p. 377-400.
77. Satir, P., L.B. Pedersen, and S.T. Christensen, *The primary cilium at a glance.* J Cell Sci, 2010. **123**(Pt 4): p. 499-503.
78. Logan, C.V., Z. Abdel-Hamed, and C.A. Johnson, *Molecular genetics and pathogenic mechanisms for the severe ciliopathies: insights into neurodevelopment and pathogenesis of neural tube defects.* Mol Neurobiol, 2011. **43**(1): p. 12-26.
79. Takeda, S., et al., *Left-right asymmetry and kinesin superfamily protein KIF3A: new insights in determination of laterality and mesoderm induction by kif3A-/- mice analysis.* J Cell Biol, 1999. **145**(4): p. 825-36.
80. Sun, Z., et al., *A genetic screen in zebrafish identifies cilia genes as a principal cause of cystic kidney.* Development, 2004. **131**(16): p. 4085-93.
81. Ibanez-Tallon, I., S. Gorokhova, and N. Heintz, *Loss of function of axonemal dynein Mdnah5 causes primary ciliary dyskinesia and hydrocephalus.* Hum Mol Genet, 2002. **11**(6): p. 715-21.
82. Davenport, J.R. and B.K. Yoder, *An incredible decade for the primary cilium: a look at a once-forgotten organelle.* Am J Physiol Renal Physiol, 2005. **289**(6): p. F1159-69.
83. Novarino, G., N. Akizu, and J.G. Gleeson, *Modeling human disease in humans: the ciliopathies.* Cell, 2011. **147**(1): p. 70-9.
84. Davis, E.E. and N. Katsanis, *The ciliopathies: a transitional model into systems biology of human genetic disease.* Curr Opin Genet Dev, 2012. **22**(3): p. 290-303.
85. Cardenas-Rodriguez, M. and J.L. Badano, *Ciliary biology: understanding the cellular and genetic basis of human ciliopathies.* Am J Med Genet C Semin Med Genet, 2009. **151C**(4): p. 263-80.
86. Murcia, N.S., et al., *The Oak Ridge Polycystic Kidney (orpk) disease gene is required for left-right axis determination.* Development, 2000. **127**(11): p. 2347-55.
87. Lager, D.J., et al., *The pck rat: a new model that resembles human autosomal dominant polycystic kidney and liver disease.* Kidney Int, 2001. **59**(1): p. 126-36.
88. Masyuk, T.V., et al., *Defects in cholangiocyte fibrocystin expression and ciliary structure in the PCK rat.* Gastroenterology, 2003. **125**(5): p. 1303-10.
89. Nishimura, D.Y., et al., *Bbs2-null mice have neurosensory deficits, a defect in social dominance, and retinopathy associated with mislocalization of rhodopsin.*

Proc Natl Acad Sci U S A, 2004. **101**(47): p. 16588-93.
90. Zhang, Q., et al., *Disruption of IFT results in both exocrine and endocrine abnormalities in the pancreas of Tg737(orpk) mutant mice.* Lab Invest, 2005. **85**(1): p. 45-64.
91. Nauli, S.M., et al., *Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells.* Nat Genet, 2003. **33**(2): p. 129-37.
92. Hildebrandt, F. and E. Otto, *Cilia and centrosomes: a unifying pathogenic concept for cystic kidney disease?* Nat Rev Genet, 2005. **6**(12): p. 928-40.
93. Hildebrandt, F., T. Benzing, and N. Katsanis, *Ciliopathies.* N Engl J Med, 2011. **364**(16): p. 1533-43.
94. Lehman, J.M., et al., *The Oak Ridge Polycystic Kidney mouse: modeling ciliopathies of mice and men.* Dev Dyn, 2008. **237**(8): p. 1960-71.
95. Menezes, L.F. and G.G. Germino, *Polycystic kidney disease, cilia, and planar polarity.* Methods Cell Biol, 2009. **94**: p. 273-97.
96. Deltas, C. and G. Papagregoriou, *Cystic diseases of the kidney: molecular biology and genetics.* Arch Pathol Lab Med, 2010. **134**(4): p. 569-82.
97. Winyard, P. and D. Jenkins, *Putative roles of cilia in polycystic kidney disease.* Biochim Biophys Acta, 2011. **1812**(10): p. 1256-62.
98. Sullivan-Brown, J., et al., *Zebrafish mutations affecting cilia motility share similar cystic phenotypes and suggest a mechanism of cyst formation that differs from pkd2 morphants.* Dev Biol, 2008. **314**(2): p. 261-75.
99. Zaghloul, N.A. and N. Katsanis, *Zebrafish assays of ciliopathies.* Methods Cell Biol, 2011. **105**: p. 257-72.
100. Zaghloul, N.A. and N. Katsanis, *Mechanistic insights into Bardet-Biedl syndrome, a model ciliopathy.* J Clin Invest, 2009. **119**(3): p. 428-37.
101. Lie, H. and T. Ferkol, *Primary ciliary dyskinesia: recent advances in pathogenesis, diagnosis and treatment.* Drugs, 2007. **67**(13): p. 1883-92.
102. Sutherland, M.J. and S.M. Ware, *Disorders of left-right asymmetry: heterotaxy and situs inversus.* Am J Med Genet C Semin Med Genet, 2009. **151C**(4): p. 307-17.
103. Amirav, I., et al., *A reach-out system for video microscopy analysis of ciliary motions aiding PCD diagnosis.* BMC Res Notes, 2015. **8**: p. 71.
104. Leigh, M.W., M.A. Zariwala, and M.R. Knowles, *Primary ciliary dyskinesia: improving the diagnostic approach.* Curr Opin Pediatr, 2009. **21**(3): p. 320-5.
105. Kuehni, C.E., et al., *Factors influencing age at diagnosis of primary ciliary dyskinesia in European children.* Eur Respir J, 2010. **36**(6): p. 1248-58.
106. Lucas, J.S., et al., *Diagnosis and management of primary ciliary dyskinesia.* Arch Dis Child, 2014. **99**(9): p. 850-6.
107. Strippoli, M.P., et al., *Management of primary ciliary dyskinesia in European children: recommendations and clinical practice.* Eur Respir J, 2012. **39**(6): p. 1482-91.
108. O'Callaghan, C., et al., *Diagnosing primary ciliary dyskinesia.* Thorax, 2007. **62**(8): p. 656-7.
109. Collins, S.A., et al., *Nasal nitric oxide screening for primary ciliary dyskinesia: systematic review and meta-analysis.* Eur Respir J, 2014. **44**(6): p. 1589-99.
110. Kouis, P., S.I. Papatheodorou, and P.K. Yiallouros, *Diagnostic accuracy of nasal nitric oxide for establishing diagnosis of primary ciliary dyskinesia: a meta-analysis.* BMC Pulm Med, 2015. **15**: p. 153.
111. Canciani, M., et al., *The saccharin method for testing mucociliary function in patients suspected of having primary ciliary dyskinesia.* Pediatr Pulmonol, 1988. **5**(4): p. 210-4.
112. Barbato, A., et al., *Primary ciliary dyskinesia: a consensus statement on diagnostic and treatment approaches in children.* Eur Respir J, 2009. **34**(6): p. 1264-76.

113. Frommer, A., et al., *Immunofluorescence Analysis and Diagnosis of Primary Ciliary Dyskinesia with Radial Spoke Defects.* Am J Respir Cell Mol Biol, 2015. **53**(4): p. 563-73.
114. McIntosh, R., D. Nicastro, and D. Mastronarde, *New views of cells in 3D: an introduction to electron tomography.* Trends Cell Biol, 2005. **15**(1): p. 43-51.
115. Baumeister, W., R. Grimm, and J. Walz, *Electron tomography of molecules and cells.* Trends Cell Biol, 1999. **9**(2): p. 81-5.
116. Koster, A.J., et al., *Perspectives of molecular and cellular electron tomography.* J Struct Biol, 1997. **120**(3): p. 276-308.
117. Noone, P.G., et al., *Discordant organ laterality in monozygotic twins with primary ciliary dyskinesia.* Am J Med Genet, 1999. **82**(2): p. 155-60.
118. Blouin, J.L., et al., *Primary ciliary dyskinesia: a genome-wide linkage analysis reveals extensive locus heterogeneity.* Eur J Hum Genet, 2000. **8**(2): p. 109-18.
119. Gherman, A., E.E. Davis, and N. Katsanis, *The ciliary proteome database: an integrated community resource for the genetic and functional dissection of cilia.* Nat Genet, 2006. **38**(9): p. 961-2.
120. Ostrowski, L.E., et al., *A proteomic analysis of human cilia: identification of novel components.* Mol Cell Proteomics, 2002. **1**(6): p. 451-65.
121. Kobayashi, D. and H. Takeda, *Ciliary motility: the components and cytoplasmic preassembly mechanisms of the axonemal dyneins.* Differentiation, 2012. **83**(2): p. S23-9.
122. Omran, H., et al., *Homozygosity mapping of a gene locus for primary ciliary dyskinesia on chromosome 5p and identification of the heavy dynein chain DNAH5 as a candidate gene.* Am J Respir Cell Mol Biol, 2000. **23**(5): p. 696-702.
123. Pennarun, G., et al., *Loss-of-function mutations in a human gene related to Chlamydomonas reinhardtii dynein IC78 result in primary ciliary dyskinesia.* Am J Hum Genet, 1999. **65**(6): p. 1508-19.
124. Horvath, J., et al., *Identification and analysis of axonemal dynein light chain 1 in primary ciliary dyskinesia patients.* Am J Respir Cell Mol Biol, 2005. **33**(1): p. 41-7.
125. Sadek, C.M., et al., *Sptrx-2, a fusion protein composed of one thioredoxin and three tandemly repeated NDP-kinase domains is expressed in human testis germ cells.* Genes Cells, 2001. **6**(12): p. 1077-90.
126. Imtiaz, F., et al., *Variation in DNAH1 may contribute to primary ciliary dyskinesia.* BMC Med Genet, 2015. **16**: p. 14.
127. Onoufriadis, A., et al., *Splice-site mutations in the axonemal outer dynein arm docking complex gene CCDC114 cause primary ciliary dyskinesia.* Am J Hum Genet, 2013. **92**(1): p. 88-98.
128. Hjeij, R., et al., *ARMC4 mutations cause primary ciliary dyskinesia with randomization of left/right body asymmetry.* Am J Hum Genet, 2013. **93**(2): p. 357-67.
129. Panizzi, J.R., et al., *CCDC103 mutations cause primary ciliary dyskinesia by disrupting assembly of ciliary dynein arms.* Nat Genet, 2012. **44**(6): p. 714-9.
130. Tarkar, A., et al., *DYX1C1 is required for axonemal dynein assembly and ciliary motility.* Nat Genet, 2013. **45**(9): p. 995-1003.
131. Knowles, M.R., et al., *Mutations in SPAG1 cause primary ciliary dyskinesia associated with defective outer and inner dynein arms.* Am J Hum Genet, 2013. **93**(4): p. 711-20.
132. Austin-Tse, C., et al., *Zebrafish Ciliopathy Screen Plus Human Mutational Analysis Identifies C21orf59 and CCDC65 Defects as Causing Primary Ciliary Dyskinesia.* Am J Hum Genet, 2013. **93**(4): p. 672-86.
133. Kott, E., et al., *Loss-of-function mutations in LRRC6, a gene essential for proper*

axonemal assembly of inner and outer dynein arms, cause primary ciliary dyskinesia. Am J Hum Genet, 2012. **91**(5): p. 958-64.
134. Moore, D.J., et al., *Mutations in ZMYND10, a gene essential for proper axonemal assembly of inner and outer dynein arms in humans and flies, cause primary ciliary dyskinesia.* Am J Hum Genet, 2013. **93**(2): p. 346-56.
135. Horani, A., et al., *CCDC65 mutation causes primary ciliary dyskinesia with normal ultrastructure and hyperkinetic cilia.* PLoS One, 2013. **8**(8): p. e72299.
136. Hjeij, R., et al., *CCDC151 mutations cause primary ciliary dyskinesia by disruption of the outer dynein arm docking complex formation.* Am J Hum Genet, 2014. **95**(3): p. 257-74.
137. Merveille, A.C., et al., *CCDC39 is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs.* Nat Genet, 2011. **43**(1): p. 72-8.
138. Becker-Heck, A., et al., *The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis formation.* Nat Genet, 2011. **43**(1): p. 79-84.
139. Wirschell, M., et al., *The nexin-dynein regulatory complex subunit DRC1 is essential for motile cilia function in algae and humans.* Nat Genet, 2013. **45**(3): p. 262-8.
140. Olbrich, H., et al., *Loss-of-Function GAS8 Mutations Cause Primary Ciliary Dyskinesia and Disrupt the Nexin-Dynein Regulatory Complex.* Am J Hum Genet, 2015. **97**(4): p. 546-54.
141. Castleman, V.H., et al., *Mutations in radial spoke head protein genes RSPH9 and RSPH4A cause primary ciliary dyskinesia with central-microtubular-pair abnormalities.* Am J Hum Genet, 2009. **84**(2): p. 197-209.
142. Kott, E., et al., *Loss-of-function mutations in RSPH1 cause primary ciliary dyskinesia with central-complex and radial-spoke defects.* Am J Hum Genet, 2013. **93**(3): p. 561-70.
143. Jeanson, L., et al., *RSPH3 Mutations Cause Primary Ciliary Dyskinesia with Central-Complex Defects and a Near Absence of Radial Spokes.* Am J Hum Genet, 2015. **97**(1): p. 153-62.
144. Wallmeier, J., et al., *Mutations in CCNO result in congenital mucociliary clearance disorder with reduced generation of multiple motile cilia.* Nat Genet, 2014. **46**(6): p. 646-51.
145. Boon, M., et al., *MCIDAS mutations result in a mucociliary clearance disorder with reduced generation of multiple motile cilia.* Nat Commun, 2014. **5**: p. 4418.
146. Moore, A., et al., *RPGR is mutated in patients with a complex X linked phenotype combining primary ciliary dyskinesia and retinitis pigmentosa.* J Med Genet, 2006. **43**(4): p. 326-33.
147. Budny, B., et al., *A novel X-linked recessive mental retardation syndrome comprising macrocephaly and ciliary dysfunction is allelic to oral-facial-digital type I syndrome.* Hum Genet, 2006. **120**(2): p. 171-8.
148. Knowles, M.R., et al., *Primary ciliary dyskinesia. Recent advances in diagnostics, genetics, and characterization of clinical disease.* Am J Respir Crit Care Med, 2013. **188**(8): p. 913-22.
149. Kurkowiak, M., E. Zietkiewicz, and M. Witt, *Recent advances in primary ciliary dyskinesia genetics.* J Med Genet, 2015. **52**(1): p. 1-9.
150. Marshall, C.R., et al., *Whole-Exome Sequencing and Targeted Copy Number Analysis in Primary Ciliary Dyskinesia.* G3 (Bethesda), 2015. **5**(8): p. 1775-81.
151. Failly, M., et al., *DNAI1 mutations explain only 2% of primary ciliary dykinesia.* Respiration, 2008. **76**(2): p. 198-204.
152. Zariwala, M., et al., *Germline mutations in an intermediate chain dynein cause*

primary ciliary dyskinesia. Am J Respir Cell Mol Biol, 2001. **25**(5): p. 577-83.
153. Zariwala, M.A., et al., *Mutations of DNAI1 in primary ciliary dyskinesia: evidence of founder effect in a common mutation.* Am J Respir Crit Care Med, 2006. **174**(8): p. 858-66.
154. Zietkiewicz, E., et al., *Population specificity of the DNAI1 gene mutation spectrum in primary ciliary dyskinesia (PCD).* Respir Res, 2010. **11**: p. 174.
155. Failly, M., et al., *Mutations in DNAH5 account for only 15% of a non-preselected cohort of patients with primary ciliary dyskinesia.* J Med Genet, 2009. **46**(4): p. 281-6.
156. Loges, N.T., et al., *DNAI2 mutations cause primary ciliary dyskinesia with defects in the outer dynein arm.* Am J Hum Genet, 2008. **83**(5): p. 547-58.
157. Duriez, B., et al., *A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia.* Proc Natl Acad Sci U S A, 2007. **104**(9): p. 3336-41.
158. Pifferi, M., et al., *New DNAH11 mutations in primary ciliary dyskinesia with normal axonemal ultrastructure.* Eur Respir J, 2010. **35**(6): p. 1413-6.
159. Lucas, J.S., et al., *Static respiratory cilia associated with mutations in Dnahc11/DNAH11: a mouse model of PCD.* Hum Mutat, 2012. **33**(3): p. 495-503.
160. Mazor, M., et al., *Primary ciliary dyskinesia caused by homozygous mutation in DNAL1, encoding dynein light chain 1.* Am J Hum Genet, 2011. **88**(5): p. 599-607.
161. Neesen, J., et al., *Identification of the human ortholog of the t-complex-encoded protein TCTE3 and evaluation as a candidate gene for primary ciliary dyskinesia.* Cytogenet Genome Res, 2002. **98**(1): p. 38-44.
162. Ben Khelifa, M., et al., *Mutations in DNAH1, which encodes an inner arm heavy chain dynein, lead to male infertility from multiple morphological abnormalities of the sperm flagella.* Am J Hum Genet, 2014. **94**(1): p. 95-104.
163. Zietkiewicz, E., et al., *Mutations in radial spoke head genes and ultrastructural cilia defects in East-European cohort of primary ciliary dyskinesia patients.* PLoS One, 2012. **7**(3): p. e33667.
164. Daniels, M.L., et al., *Founder mutation in RSPH4A identified in patients of Hispanic descent with primary ciliary dyskinesia.* Hum Mutat, 2013. **34**(10): p. 1352-6.
165. Alsaadi, M.M., et al., *From a single whole exome read to notions of clinical screening: primary ciliary dyskinesia and RSPH9 p.Lys268del in the Arabian Peninsula.* Ann Hum Genet, 2012. **76**(3): p. 211-20.
166. Doggett, N.A., et al., *A 360-kb interchromosomal duplication of the human HYDIN locus.* Genomics, 2006. **88**(6): p. 762-71.
167. Antony, D., et al., *Mutations in CCDC39 and CCDC40 are the major cause of primary ciliary dyskinesia with axonemal disorganization and absent inner dynein arms.* Hum Mutat, 2013. **34**(3): p. 462-72.
168. Horani, A., et al., *Whole-exome capture and sequencing identifies HEATR2 mutation as a cause of primary ciliary dyskinesia.* Am J Hum Genet, 2012. **91**(4): p. 685-93.
169. Diggle, C.P., et al., *HEATR2 plays a conserved role in assembly of the ciliary motile apparatus.* PLoS Genet, 2014. **10**(9): p. e1004577.
170. Zariwala, M.A., et al., *ZMYND10 is mutated in primary ciliary dyskinesia and interacts with LRRC6.* Am J Hum Genet, 2013. **93**(2): p. 336-45.
171. Alsaadi, M.M., et al., *Nonsense mutation in coiled-coil domain containing 151 gene (CCDC151) causes primary ciliary dyskinesia.* Hum Mutat, 2014. **35**(12): p. 1446-8.
172. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.

173. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors. 1977.* Biotechnology, 1992. **24**: p. 104-8.
174. Neveling, K., et al., *A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases.* Hum Mutat, 2013. **34**(12): p. 1721-6.
175. Cooper, G.M. and J. Shendure, *Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data.* Nat Rev Genet, 2011. **12**(9): p. 628-40.
176. Stone, E.A. and A. Sidow, *Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity.* Genome Res, 2005. **15**(7): p. 978-86.
177. Ng, P.C. and S. Henikoff, *Predicting the effects of amino acid substitutions on protein function.* Annu Rev Genomics Hum Genet, 2006. **7**: p. 61-80.
178. Ng, P.C. and S. Henikoff, *Predicting deleterious amino acid substitutions.* Genome Res, 2001. **11**(5): p. 863-74.
179. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.
180. Cooper, G.M., et al., *Distribution and intensity of constraint in mammalian genomic sequence.* Genome Res, 2005. **15**(7): p. 901-13.
181. Cooper, G.M., et al., *Single-nucleotide evolutionary constraint scores highlight disease-causing mutations.* Nat Methods, 2010. **7**(4): p. 250-1.
182. Yip, Y.L., et al., *The Swiss-Prot variant page and the ModSNP database: a resource for sequence and structure information on human protein variants.* Hum Mutat, 2004. **23**(5): p. 464-70.
183. Stenson, P.D., et al., *The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics.* Hum Genomics, 2009. **4**(2): p. 69-72.
184. Stenson, P.D., et al., *Human Gene Mutation Database (HGMD): 2003 update.* Hum Mutat, 2003. **21**(6): p. 577-81.
185. Hindorff, L.A., et al., *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.* Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9362-7.
186. Onoufriadis, A., et al., *Combined exome and whole-genome sequencing identifies mutations in ARMC4 as a cause of primary ciliary dyskinesia with defects in the outer dynein arm.* J Med Genet, 2014. **51**(1): p. 61-7.
187. Onoufriadis, A., et al., *Targeted NGS gene panel identifies mutations in RSPH1 causing primary ciliary dyskinesia and a common mechanism for ciliary central pair agenesis due to radial spoke defects.* Hum Mol Genet, 2014. **23**(13): p. 3362-74.
188. Watson, C.M., et al., *Robust diagnostic genetic testing using solution capture enrichment and a novel variant-filtering interface.* Hum Mutat, 2014. **35**(4): p. 434-41.
189. Knowles, M.R., et al., *Mutations in RSPH1 cause primary ciliary dyskinesia with a unique clinical and ciliary phenotype.* Am J Respir Crit Care Med, 2014. **189**(6): p. 707-17.
190. Sui, W., et al., *CCDC40 mutation as a cause of primary ciliary dyskinesia: a case report and review of literature.* Clin Respir J, 2015.
191. Garcia-Gonzalo, F.R., et al., *A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition.* Nat Genet, 2011. **43**(8): p. 776-84.
192. Kircher, M., U. Stenzel, and J. Kelso, *Improved base calling for the Illumina Genome Analyzer using machine learning strategies.* Genome Biol, 2009. **10**(8): p. R83.
193. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-*

Wheeler transform. Bioinformatics, 2010. **26**(5): p. 589-95.
194. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
195. Kozarewa, I. and D.J. Turner, *Amplification-free library preparation for paired-end Illumina sequencing.* Methods Mol Biol, 2011. **733**: p. 257-66.
196. Ye, K., et al., *Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads.* Bioinformatics, 2009. **25**(21): p. 2865-71.
197. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.
198. Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.* Nat Protoc, 2009. **4**(7): p. 1073-81.
199. Schwarz, J.M., et al., *MutationTaster2: mutation prediction for the deep-sequencing age.* Nat Methods, 2014. **11**(4): p. 361-2.
200. Welter, D., et al., *The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1001-6.
201. MacDonald, J.R., et al., *The Database of Genomic Variants: a curated collection of structural variation in the human genome.* Nucleic Acids Res, 2014. **42**(Database issue): p. D986-92.
202. Li, H., *A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data.* Bioinformatics, 2011. **27**(21): p. 2987-93.
203. Chou, Q., et al., *Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications.* Nucleic Acids Res, 1992. **20**(7): p. 1717-23.
204. Chilvers, M.A. and C. O'Callaghan, *Analysis of ciliary beat pattern and beat frequency using digital high speed imaging: comparison with the photomultiplier and photodiode methods.* Thorax, 2000. **55**(4): p. 314-7.
205. Rutland, J., et al., *Nasal brushing for the study of ciliary ultrastructure.* J Clin Pathol, 1982. **35**(3): p. 357-9.
206. Owen, C.H. and W.J. Landis, *Alignment of electron tomographic series by correlation without the use of gold particles.* Ultramicroscopy, 1996. **63**(1): p. 27-38.
207. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
208. Narayan, D., et al., *Unusual inheritance of primary ciliary dyskinesia (Kartagener's syndrome).* J Med Genet, 1994. **31**(6): p. 493-6.
209. Luz, S., et al., *LMTK2-mediated phosphorylation regulates CFTR endocytosis in human airway epithelial cells.* J Biol Chem, 2014. **289**(21): p. 15080-93.
210. Sakugawa, N., et al., *LMTK2 and PARP-2 gene polymorphism and azoospermia secondary to meiotic arrest.* J Assist Reprod Genet, 2009. **26**(9-10): p. 545-52.
211. Kawa, S., et al., *Azoospermia in mice with targeted disruption of the Brek/Lmtk2 (brain-enriched kinase/lemur tyrosine kinase 2) gene.* Proc Natl Acad Sci U S A, 2006. **103**(51): p. 19344-9.
212. Makrythanasis, P. and S.E. Antonarakis, *Pathogenic variants in non-protein-coding sequences.* Clin Genet, 2013. **84**(5): p. 422-8.
213. Kramer-Zucker, A.G., et al., *Cilia-driven fluid flow in the zebrafish pronephros, brain and Kupffer's vesicle is required for normal organogenesis*, in *Development*. 2005. p. 1907-21.

Other citations

1.
https://webcache.googleusercontent.com/search?q=cache:Zlpij4SuRUIJ:https://tools.thermofisher.com/content/sfs/manuals/qubit_3_fluorometer_man.pdf+&cd=2&hl=fr&ct=clnk&gl=it

2.
Illumina Sureselect XT Target Enrichment
http://webcache.googleusercontent.com/search?q=cache:PVWgOWMUnDEJ:www.agilent.com/cs/library/usermanuals/Public/G7530-90000.pdf+&cd=1&hl=fr&ct=clnk&gl=it

3.
http://support.illumina.com/sequencing/sequencing_instruments/cbot.html

4.
http://www.crg.eu/en/content/processing-and-analysis-illumina-sequencing-data

5.
http://support.illumina.com/downloads/bcl2fastq_conversion_software_184.html

6.
https://webcache.googleusercontent.com/search?q=cache:M1UishfJqmsJ:https://www.agilent.com/cs/library/usermanuals/Public/G9900-90001.pdf+&cd=1&hl=fr&ct=clnk&gl=it

7.
http://www.putty.org/

8.
http://bio-bwa.sourceforge.net/

9.
http://samtools.sourceforge.net/

10.
http://gmt.genome.wustl.edu/packages/pindel/background.html

11.
http://www.openbioinformatics.org/annovar/

12.
http://compgen.cshl.edu/phast/background.php

13.
http://www.ebi.ac.uk/gwas

14.
http://dgv.tcag.ca/dgv/app/home

15.
http://www.hgmd.cf.ac.uk/ac/index.php

16.
http://www.orpha.net/

17.
Copyright © 1989, 1991, 1992, 1993, 1996–2005, 2007, 2009–2015 Free Software Foundation, Inc.

18.
http://samtools.sourceforge.net/tview.shtml

19.
http://primer3.ut.ee/

20.
https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html

21.
https://genome.ucsc.edu/cgi-bin/hgBlat?command=start

22.
http://webcache.googleusercontent.com/search?q=cache:W14Ul6kQv4kJ:mvz.berkeley.edu/egl/inserts/Big_Dye_v3.1_Protocol_Manual.pdf+&cd=1&hl=fr&ct=clnk&gl=it

23.
http://staden.sourceforge.net/

24.
https://www.neb.com/products/p7712-color-prestained-protein-standard-broad-range-11-245-kda

25.
https://www.thermofisher.com/order/catalog/product/32106

26.
Kremer JR, Mastronarde DN, McIntosh JR. 1996, Computer visualization of three-dimensional image data using IMOD. J Struct Biol 116:71–76.

27.
http://evs.gs.washington.edu/EVS/

28.
http://exac.broadinstitute.org/

Video

Video are available on request.

Appendix

Customer designed Haloplex –ciliome gene list

The Variant Master Software

AccuPrime ™ Taq DNA Polymerase System

SYBR® Safe DNA Gel Stain, cat N: S33102

NuSieve™ 3:1 Agarose

MicroCLEAN

EtHO washing

HaloPlex Design Report-640 Ciliome genes

File Summary
File Type: HaloPlex Design Report
Created By: HaloPlex Standard Design Wizard
User: jean-louis.blouin@unige.ch
Workgroup: UNIGE-BLOUIN
Folder: UNIGE-BLOUIN
Timestamp: 12-Feb-2015

Design Summary
Design Name: PCD_Haloplex_Finalized_13.2.2015
Design ID: 07419-1423760936
Design Category: HaloPlex
Species: H. sapiens (H. sapiens, hg19, GRCh37, February 2009)
Platform: Illumina
Read Length: 100 bp

Target Summary
640 Target IDs resolved to 663 targets comprising 10040 regions.
0 Target IDs were not found.
Region Size: 1.824 Mbp

Amplicon Summary
Total Amplicons: 128774
Total Target Bases Analyzable: 1.79 Mbp
Total Sequenceable Design Size: 4.69 Mbp
Target Coverage: 98.05 %
Recommended Minimum Sequencing per Sample: 938.60 Mbp
Pricing: Illumina Tier 2 (Target Region Size = 0.501 - 2.5 Mb; up to 200K probes).
P/N: G9911B/C

Target Parameters
Databases: RefSeq, Ensembl, CCDS, Gencode, VEGA, SNP, CytoBand
Region: Coding Exons
Region Extension: 10 bases from 3' end and 10 bases from 5' end.
Allow Synonyms: Yes

Target and Probe Details
TargetID: The identifier entered in the Targets list.
Interval: The genomic interval of the target.
Regions: The number of regions within this target.
Size: The total size (in base pairs) of the regions.
Database(s): The databases in which this target was found.
#High Coverage: Number of regions where analyzable amplicon overlap >= 90%.
#Low Coverage: Number of regions where analyzable amplicon overlap < 90%

TargetID	Interval	Regions	Size	Databases	Coverage
ABC17	chr6:32813346-32821603	11	2647	Gencode, RefSeq, VEGA	99.70
ABCB2	chr6:32813346-32821603	11	2647	Gencode, RefSeq, VEGA	99.70
ACE	chr17:61554446-61574931	27	4688	Gencode, RefSeq, VEGA	98.55
ACTL9	chr19:8807791-8809061	1	1271	Gencode, RefSeq	100.0
ACTN4	chr19:39138376-39220317	24	3498	Gencode, RefSeq, VEGA	100.00
ACTR2	chr2:65455034-65495878	10	1469	Gencode, RefSeq, VEGA	99.66

ACTR6	chr12:100594620-100617703	11	1411	Gencode, RefSeq	100.00
AGT	chr1:230838877-230846606	4	1538	Gencode, RefSeq, VEGA	97.01
AGT	chr2:241808273-241818248	11	1399	Gencode, RefSeq, VEGA	99.64
AGTR1	chr3:148447957-148459912	3	1285	Gencode, RefSeq, VEGA	99.38
AHI1	chr6:135606754-135830229	30	4324	Gencode, RefSeq, VEGA	98.27
AIPL1	chr17:6328770-6338434	6	1275	Gencode, RefSeq, VEGA	100.00
AK1	chr9:130630277-130636906	7	780	Gencode, RefSeq, VEGA	100.00
AK7	chr14:96858482-96954690	18	2532	Gencode, RefSeq	98.14
ALMS1	chr2:73612987-73836749	24	12999	Gencode, RefSeq, VEGA	98.11
ANKAR	chr2:190541207-190611363	22	4778	Gencode, RefSeq, VEGA	99.14
APP	chr21:27253971-27542948	20	2821	Gencode, RefSeq, VEGA	97.84
APT1	chr6:32813346-32821603	11	2647	Gencode, RefSeq, VEGA	99.70
APT1	chr8:54960615-55014393	10	920	Gencode, RefSeq, VEGA	100.00
APT1	chr10:90749216-90774217	10	1354	Gencode, RefSeq, VEGA	99.41
AQP2	chr12:50344604-50349401	4	896	Gencode, RefSeq	100.00
ARHGAP24	chr4:86491685-86921885	11	2601	Gencode, RefSeq, VEGA	100.00
ARL13	chrX:100228669-100245703	8	1183	Gencode, RefSeq, VEGA	100.00
ARL13A	chrX:100228669-100245703	8	1183	Gencode, RefSeq, VEGA	100.00
ARL13B	chr3:93699258-93772117	11	1644	Gencode, RefSeq, VEGA	98.84
ARL2	chr11:64781670-64789337	5	655	Gencode, RefSeq	100.00
ARL2BP	chr16:57279270-57286189	6	612	Gencode, RefSeq, VEGA	95.10
ARL6	chr3:97486942-97516903	7	701	Gencode, RefSeq, VEGA	90.58
ARL6	chr17:41477091-41477716	1	626	Gencode, RefSeq	100.00
ARMC1	chr8:66516619-66539643	6	969	Gencode, RefSeq, VEGA	100.00
ARMC10	chr7:102715710-102739010	11	1433	Gencode, RefSeq, VEGA	95.60
ARMC12	chr6:35704876-35716657	6	1224	RefSeq	100.00
ARMC2	chr6:109175461-109294727	18	2995	Gencode, RefSeq, VEGA	99.33
ARMC3	chr10:23220916-23326418	19	3025	Gencode, RefSeq, VEGA	99.93
ARMC4	chr10:28089751-28284081	22	3827	Gencode, RefSeq, VEGA	96.97
ARMC5	chr16:31469708-31478220	9	3626	Gencode, RefSeq	99.94
ARMC6	chr19:19145009-19168611	8	1830	Gencode, RefSeq	99.62
ARMC7	chr17:73106374-73125143	3	657	Gencode, RefSeq	100.00
ARMC8	chr3:137906387-138014744	25	2641	Gencode, RefSeq, VEGA	99.96
ARMC9	chr2:232070942-232236256	27	3179	Gencode, RefSeq, VEGA	99.81
ARMCX1	chrX:100807904-100809285	1	1382	Gencode, RefSeq, VEGA	99.42
ARMCX2	chrX:100910666-100912584	1	1919	Gencode, RefSeq, VEGA	100.00
ARMCX3	chrX:100879960-100881119	1	1160	Gencode, RefSeq, VEGA	99.57
ARMCX4	chrX:100742595-100753186	4	7310	Gencode, RefSeq, VEGA	99.74
ARMCX5	chrX:101857060-101858756	1	1697	Gencode, RefSeq, VEGA	99.41
ARMCX5-GPRASP2	chrX:101969788-101972324	1	2537	RefSeq	99.01
ARMCX6	chrX:100870698-100871620	1	923	Gencode, RefSeq, VEGA	99.35
ARMS2	chr10:124214234-124216459	2	364	Gencode, RefSeq	97.25
ARP5	chr22:39410358-39414402	4	653	Gencode, RefSeq, VEGA	100.00
ARP5	chr20:37377112-37400469	9	2004	Gencode, RefSeq, VEGA	100.00
ARP5	chr19:10203255-10207249	5	1513	Gencode, RefSeq	100.00
ARPC1A	chr7:98930967-98984345	13	1549	Gencode, RefSeq, VEGA	98.77
ARPC1B	chr7:98983328-98992122	10	1360	Gencode, RefSeq, VEGA	100.00
ARPC3	chr12:110872721-110888075	8	768	Gencode, RefSeq, VEGA	97.14
ARPC4	chr3:9834796-9870746	10	1039	Gencode, RefSeq, VEGA	100.00
ARPC5L	chr9:127631560-127639229	4	542	Gencode, RefSeq, VEGA	100.00
ARVCF	chr22:19958741-19978327	18	3270	Gencode, RefSeq, VEGA	99.69
ATP6B1	chr2:71163075-71192261	14	2016	Gencode, RefSeq, VEGA	100.00
ATP6N1B	chr7:138391359-138456002	20	2923	Gencode, RefSeq, VEGA	100.00
AVPR2	chrX:153170590-153172280	3	1284	Gencode, RefSeq, VEGA	100.00
AXDND1	chr1:179335635-179523664	27	3731	Gencode, RefSeq	99.06

BBS1	chr11:66278121-66299518	18	2179	Gencode, RefSeq	100.00
BBS10	chr12:76739583-76742148	2	2212	Gencode, RefSeq, VEGA	97.11
BBS12	chr4:123663038-123665190	1	2153	Gencode, RefSeq, VEGA	99.63
BBS2	chr16:56518663-56553784	17	2506	Gencode, RefSeq	100.00
BBS4	chr15:72978559-73029938	18	2008	Gencode, RefSeq, VEGA	96.71
BBS5	chr2:170336054-170382216	17	2077	Gencode, RefSeq, VEGA	97.64
BBS7	chr4:122747005-122791478	19	2533	Gencode, RefSeq, VEGA	100.00
BBS9	chr7:33185855-33644848	28	3427	Gencode, RefSeq, VEGA	100.00
BECN1	chr17:40962768-40975905	11	1573	Gencode, RefSeq	100.00
BSND	chr1:55464850-55474311	4	1043	Gencode, RefSeq, VEGA	99.42
C10ORF118	chr10:115884892-115933915	17	3359	Gencode, RefSeq, VEGA	99.82
C12ORF68	chr12:48577896-48578500	1	605	Gencode, RefSeq	100.00
C17ORF66	chr17:34182057-34195756	15	2013	Gencode, RefSeq, VEGA	100.00
C17ORF76	chr17:16346892-16395314	4	1115	Gencode, RefSeq, VEGA	100.00
C1ORF65	chr1:223566808-223568699	1	1892	Gencode, RefSeq, VEGA	99.52
C21ORF59	chr21:33951122-33984563	10	1581	Gencode, RefSeq, VEGA	99.37
C22ORF36	chr22:24981844-24989016	6	1137	Gencode, RefSeq, VEGA	97.80
CAPS	chr3:62385071-62860714	38	5204	Gencode, RefSeq, VEGA	97.27
CAPS	chr19:5913934-5915343	5	1069	Gencode, RefSeq	100.00
CC2D2A	chr4:15477547-15603058	38	5843	Gencode, RefSeq, VEGA	99.18
CCDC101	chr16:28592381-28603037	9	1062	Gencode, RefSeq, VEGA	97.74
CCDC102A	chr16:57546643-57563099	8	1813	Gencode, RefSeq, VEGA	100.00
CCDC102B	chr18:66503991-66721384	9	1777	Gencode, RefSeq, VEGA	97.19
CCDC103	chr17:42978357-42980195	4	834	Gencode, RefSeq, VEGA	100.00
CCDC105	chr19:15121628-15133941	7	1640	Gencode, RefSeq	99.27
CCDC106	chr19:56160403-56164122	5	943	Gencode, RefSeq	100.00
CCDC108	chr2:219867649-219906244	35	6929	Gencode, RefSeq, VEGA	99.08
CCDC109B	chr4:110481484-110608758	10	1282	Gencode, RefSeq, VEGA	99.92
CCDC110	chr4:186366646-186392847	11	3001	Gencode, RefSeq, VEGA	97.80
CCDC112	chr5:114603563-114632250	10	1790	Gencode, RefSeq, VEGA	96.98
CCDC114	chr19:48800223-48822038	13	2273	Gencode, RefSeq, VEGA	98.33
CCDC115	chr2:131096683-131099812	6	919	Gencode, RefSeq, VEGA	100.00
CCDC116	chr22:21987263-21991369	4	2120	Gencode, RefSeq, VEGA	99.29
CCDC117	chr22:29168828-29182324	5	944	Gencode, RefSeq, VEGA	96.93
CCDC12	chr3:46963537-47018271	7	680	Gencode, RefSeq, VEGA	98.68
CCDC120	chrX:48919563-48926204	10	2313	Gencode, RefSeq, VEGA	100.00
CCDC121	chr2:27849820-27851640	2	1369	Gencode, RefSeq, VEGA	97.52
CCDC122	chr13:44411406-44443522	6	1125	Gencode, RefSeq, VEGA	83.20
CCDC124	chr19:18047220-18054534	4	752	Gencode, RefSeq	100.00
CCDC125	chr5:68578546-68616377	11	1756	Gencode, RefSeq, VEGA	100.00
CCDC126	chr7:23650925-23682744	2	463	Gencode, RefSeq, VEGA	100.00
CCDC127	chr5:205402-216974	2	823	Gencode, RefSeq, VEGA	96.96
CCDC129	chr7:31569373-31698016	17	3689	Gencode, RefSeq, VEGA	99.00
CCDC13	chr3:42750462-42799847	16	2570	Gencode, RefSeq, VEGA	99.61
CCDC130	chr19:13862613-13873892	8	1440	Gencode, RefSeq	100.00
CCDC132	chr7:92861771-92987758	30	3793	Gencode, RefSeq, VEGA	91.27
CCDC134	chr22:42204885-42221837	6	810	Gencode, RefSeq, VEGA	100.00
CCDC135	chr16:57731852-57765180	17	2965	Gencode, RefSeq, VEGA	100.00
CCDC136	chr7:128431538-128461973	20	4145	Gencode, RefSeq, VEGA	99.93
CCDC137	chr17:79633787-79639744	6	990	Gencode, RefSeq	100.00
CCDC138	chr2:109403269-109492719	16	2367	Gencode, RefSeq, VEGA	95.94
CCDC14	chr3:123616348-123680174	15	3281	Gencode, RefSeq, VEGA	99.70
CCDC140	chr2:223168612-223169123	1	512	Gencode, RefSeq, VEGA	100.00
CCDC141	chr2:179698890-179914678	24	5161	Gencode, RefSeq, VEGA	99.94
CCDC142	chr2:74699912-74709974	10	2455	Gencode, RefSeq, VEGA	100.00

CCDC146	chr7:76796976-76924193	18	3232	RefSeq, VEGA	99.47
CCDC149	chr4:24810001-24914511	13	1921	Gencode, RefSeq, VEGA	100.00
CCDC15	chr11:124824619-124910617	15	3189	Gencode, RefSeq	93.67
CCDC150	chr2:197504481-197597296	31	4312	Gencode, RefSeq, VEGA	100.00
CCDC151	chr19:11531493-11546602	14	2150	Gencode, RefSeq	100.00
CCDC152	chr5:42759214-42799893	8	925	Gencode, RefSeq, VEGA	100.00
CCDC153	chr11:119060999-119066207	6	753	Gencode, RefSeq	100.00
CCDC154	chr16:1484399-1494333	17	2371	Gencode, RefSeq, VEGA	99.96
CCDC155	chr19:49894131-49920777	19	2069	Gencode, RefSeq	99.13
CCDC157	chr22:30761980-30772744	10	2463	Gencode, RefSeq, VEGA	99.80
CCDC158	chr4:77234313-77324370	24	4014	Gencode, RefSeq, VEGA	99.58
CCDC159	chr19:11457273-11465563	14	1542	Gencode, RefSeq	92.41
CCDC160	chrX:133378821-133379818	1	998	Gencode, RefSeq, VEGA	100.00
CCDC163P	chr1:45960721-45965291	6	750	Gencode	100.00
CCDC164	chr2:26624848-26679395	17	2563	RefSeq	100.00
CCDC166	chr8:144788854-144790289	2	1360	RefSeq	100.00
CCDC167	chr6:37450952-37467649	4	374	RefSeq	90.37
CCDC168	chr13:103381791-103411292	4	21326	RefSeq	98.61
CCDC169	chr13:36801328-36871866	9	1006	RefSeq	100.00
CCDC169-SOHLH2	chr13:36743161-36857753	15	1809	RefSeq	99.67
CCDC17	chr1:46085937-46089591	13	2129	Gencode, RefSeq, VEGA	100.00
CCDC170	chr6:151815254-151939292	11	2368	RefSeq	98.82
CCDC171	chr9:15564077-15971844	25	4481	RefSeq	98.59
CCDC172	chr10:118084514-118138880	8	937	RefSeq	99.47
CCDC173	chr2:170502341-170550847	9	1839	RefSeq	100.00
CCDC174	chr3:14693334-14712711	11	1624	RefSeq	100.00
CCDC175	chr14:59971947-60043503	20	2782	RefSeq	98.96
CCDC176	chr14:74486172-74531412	12	1830	RefSeq	99.29
CCDC177	chr14:70038206-70040349	1	2144	RefSeq	100.00
CCDC178	chr18:30517965-30992062	21	3024	RefSeq	99.34
CCDC179	chr11:22869046-22881977	4	287	RefSeq	100.00
CCDC18	chr1:93646078-93744044	29	5348	Gencode, RefSeq, VEGA	99.01
CCDC180	chr9:100070028-100139186	37	5846	RefSeq	99.47
CCDC181	chr1:169364275-169394175	5	1627	RefSeq	99.88
CCDC22	chrX:49092087-49106977	18	2295	Gencode, RefSeq, VEGA	100.00
CCDC23	chr1:43273075-43282225	2	241	Gencode, RefSeq, VEGA	100.00
CCDC24	chr1:44457541-44461842	8	1084	Gencode, RefSeq, VEGA	100.00
CCDC25	chr8:27593723-27630067	10	832	Gencode, RefSeq, VEGA	99.64
CCDC27	chr1:3669036-3688097	12	2211	Gencode, RefSeq, VEGA	100.00
CCDC28A	chr6:139094802-139113950	6	945	Gencode, RefSeq, VEGA	92.59
CCDC28B	chr1:32667527-32670859	5	881	Gencode, RefSeq, VEGA	100.00
CCDC3	chr10:12940406-13043580	3	873	Gencode, RefSeq, VEGA	100.00
CCDC30	chr1:42986876-43119709	20	2926	Gencode, RefSeq, VEGA	95.35
CCDC33	chr15:74509603-74628808	25	3893	Gencode, RefSeq	99.85
CCDC34	chr11:27360358-27384751	6	1326	Gencode, RefSeq	98.11
CCDC36	chr3:49249204-49294725	9	2197	Gencode, RefSeq, VEGA	97.63
CCDC37	chr3:126114834-126155257	16	2288	Gencode, RefSeq, VEGA	99.69
CCDC38	chr12:96260844-96330297	15	1992	Gencode, RefSeq	100.00
CCDC39	chr3:180331624-180466079	24	3613	Gencode, RefSeq, VEGA	99.36
CCDC40	chr17:78010452-78073584	23	4448	Gencode, RefSeq, VEGA	98.36
CCDC42	chr17:8633438-8647937	7	1091	Gencode, RefSeq	100.00
CCDC42B	chr12:113587653-113595494	7	1067	Gencode, RefSeq	99.06
CCDC43	chr17:42756214-42767131	5	775	Gencode, RefSeq	100.00
CCDC47	chr17:61824231-61843545	12	1692	Gencode, RefSeq	98.76
CCDC50	chr3:191047454-191109559	12	1689	Gencode, RefSeq, VEGA	100.00

CCDC51	chr3:48473808-48476548	3	1296	Gencode, RefSeq, VEGA	100.00
CCDC53	chr12:102406876-102455750	7	725	Gencode, RefSeq	100.00
CCDC54	chr3:107096425-107097431	1	1007	Gencode, RefSeq, VEGA	99.80
CCDC57	chr17:80059548-80159830	21	4167	Gencode, RefSeq, VEGA	99.18
CCDC58	chr3:122078706-122102081	6	621	Gencode, RefSeq, VEGA	100.00
CCDC59	chr12:82746920-82752165	4	806	Gencode, RefSeq	100.00
CCDC6	chr10:61552665-61666192	9	1605	Gencode, RefSeq, VEGA	99.75
CCDC60	chr12:119772972-119978530	15	2189	Gencode, RefSeq	100.00
CCDC61	chr19:46498329-46521655	15	2010	Gencode, RefSeq	99.75
CCDC62	chr12:123259208-123311026	13	2363	Gencode, RefSeq	98.69
CCDC63	chr12:111290749-111345290	11	1912	Gencode, RefSeq	100.00
CCDC64	chr12:120427663-120532137	12	2227	Gencode, RefSeq	98.97
CCDC64B	chr16:3078097-3085507	9	1707	Gencode, RefSeq	100.00
CCDC65	chr12:49298110-49325270	9	1700	Gencode, RefSeq	100.00
CCDC66	chr3:56591261-56655656	24	3537	Gencode, RefSeq, VEGA	99.86
CCDC67	chr11:93065416-93170895	15	2287	Gencode, RefSeq	99.61
CCDC68	chr18:52571579-52610032	10	1208	Gencode, RefSeq, VEGA	95.94
CCDC69	chr5:150562602-150603541	10	1575	Gencode, RefSeq, VEGA	99.62
CCDC7	chr10:32740561-32863422	17	1912	Gencode, RefSeq, VEGA	97.91
CCDC70	chr13:52439505-52440226	1	722	Gencode, RefSeq, VEGA	81.72
CCDC71	chr3:49200228-49201651	1	1424	Gencode, RefSeq, VEGA	99.37
CCDC71L	chr7:106300625-106301352	1	728	RefSeq	100.00
CCDC73	chr11:32624347-32781799	17	3628	Gencode, RefSeq	99.70
CCDC74A	chr2:132285534-132290981	8	2028	Gencode, RefSeq	99.85
CCDC74B	chr2:130897118-130902579	8	2364	Gencode, RefSeq, VEGA	99.87
CCDC77	chr12:518540-551096	11	1687	Gencode, RefSeq, VEGA	100.00
CCDC78	chr16:772760-776856	14	2358	Gencode, RefSeq, VEGA	99.24
CCDC79	chr16:66788869-66830711	17	2544	Gencode, RefSeq 99.02	
CCDC8	chr19:46914441-46916077	1	1637	Gencode, RefSeq, VEGA	99.27
CCDC80	chr3:112324254-112358762	7	3148	Gencode, RefSeq, VEGA	100.00
CCDC81	chr11:86086196-86133767	15	2259	Gencode, RefSeq	99.65
CCDC82	chr11:96086800-96117921	7	1775	Gencode, RefSeq	99.27
CCDC83	chr11:85576157-85630563	11	1555	Gencode, RefSeq	97.43
CCDC84	chr11:118868898-118886318	11	1219	Gencode, RefSeq	98.69
CCDC85A	chr2:56411750-56611500	6	1782	Gencode, RefSeq, VEGA	98.26
CCDC85B	chr11:65658245-65658873	1	629	Gencode, RefSeq	100.00
CCDC85C	chr14:99981573-100070306	6	1380	Gencode, RefSeq	97.39
CCDC86	chr11:60609588-60617808	4	1163	Gencode, RefSeq	100.00
CCDC87	chr11:66357927-66360496	1	2570	Gencode, RefSeq	100.00
CCDC88A	chr2:55518837-55646225	32	6789	Gencode, RefSeq, VEGA	99.12
CCDC88B	chr11:64107729-64124970	28	5457	Gencode, RefSeq, VEGA	99.84
CCDC88C	chr14:91738959-91884044	34	6890	Gencode, RefSeq	98.56
CCDC89	chr11:85396039-85397183	1	1145	Gencode, RefSeq	100.00
CCDC9	chr19:47761648-47774945	11	1816	Gencode, RefSeq	100.00
CCDC90B	chr11:82972943-82997225	10	1076	Gencode, RefSeq	97.21
CCDC91	chr12:28410139-28702116	13	1605	Gencode, RefSeq	100.00
CCDC92	chr12:124421595-124428862	4	1076	Gencode, RefSeq	99.07
CCDC93	chr2:118677909-118771581	25	2461	Gencode, RefSeq, VEGA	97.60
CCDC94	chr19:4247134-4268703	8	1132	Gencode, RefSeq	100.00
CCDC96	chr4:7042988-7044675	1	1688	Gencode, RefSeq, VEGA	96.98
CCDC97	chr19:41816206-41828630	5	1132	Gencode, RefSeq	100.00
CCDS13803	chr22:23401630-23482617	6	1167	CCDS	99.14
CCNA1	chr13:37006749-37016812	9	1578	Gencode, RefSeq, VEGA	99.43
CCNA2	chr4:122738783-122744793	8	1459	Gencode, RefSeq, VEGA	98.90
CCNB1	chr5:68463080-68473468	9	1659	Gencode, RefSeq, VEGA	100.00

CCNB1IP1	chr14:20779699-20784692	3	894	Gencode, RefSeq, VEGA	100.00
CCNB2	chr15:59397458-59417086	9	1377	Gencode, RefSeq, VEGA	100.00
CCNB3	chrX:50028154-50094712	11	4408	Gencode, RefSeq, VEGA	99.39
CCNC	chr6:99991421-100016413	15	1442	Gencode, RefSeq, VEGA	100.00
CCND1	chr11:69456072-69466060	5	988	Gencode, RefSeq	100.00
CCND2	chr12:4383197-4409185	5	970	Gencode, RefSeq	98.25
CCND3	chr6:41903579-41909397	6	1151	Gencode, RefSeq, VEGA	100.00
CCNDBP1	chr15:43477687-43487049	11	1303	Gencode, RefSeq, VEGA	95.55
CCNE1	chr19:30303453-30314694	11	1471	Gencode, RefSeq	99.59
CCNE2	chr8:95893850-95906459	12	1554	Gencode, RefSeq, VEGA	100.00
CCNF	chr16:2479473-2507031	17	2701	Gencode, RefSeq, VEGA	99.81
CCNG1	chr5:162866253-162870719	6	1038	Gencode, RefSeq, VEGA	100.00
CCNG2	chr4:78079676-78106272	9	1343	Gencode, RefSeq, VEGA	98.59
CCNH	chr5:86690253-86708621	9	1209	Gencode, RefSeq, VEGA	100.00
CCNI	chr4:77969362-77996706	7	1346	Gencode, RefSeq, VEGA	97.10
CCNI2	chr5:132083178-132088672	6	1230	Gencode, RefSeq, VEGA	100.00
CCNJ	chr10:97804141-97818008	5	1252	Gencode, RefSeq, VEGA	98.96
CCNJL	chr5:159680375-159766538	10	1982	Gencode, RefSeq, VEGA	95.36
CCNK	chr14:99959005-99977129	11	2060	Gencode, RefSeq	99.90
CCNL1	chr3:127317300-156877893	31	5279	Gencode, RefSeq, VEGA	98.48
CCNL2	chr1:1322601-1334696	16	2126	Gencode, RefSeq, VEGA	89.98
CCNO	chr5:54527193-54529361	3	1128	Gencode, RefSeq, VEGA	100.00
CCNT1	chr12:49086806-49110468	9	2361	Gencode, RefSeq	98.81
CCNT2	chr2:135676415-135712228	11	2534	Gencode, RefSeq, VEGA	96.37
CCNY	chr10:35625972-35858108	10	1226	Gencode, RefSeq, VEGA	99.76
CCNYL1	chr2:208576611-208626415	12	1451	Gencode, RefSeq, VEGA	100.00
CCNYL2	chr10:42905521-42949843	10	1286	Gencode, VEGA	100.00
CDH2	chr18:25532107-25756996	18	3179	Gencode, RefSeq, VEGA	99.37
CDH23	chr10:73199579-73575045	72	12298	Gencode, RefSeq, VEGA	99.84
CDK1	chr10:62539914-62553743	7	1050	Gencode, RefSeq, VEGA	98.38
CENPF	chr1:214787088-214837147	22	9818	Gencode, RefSeq, VEGA	99.84
CEP164	chr11:117209293-117282894	33	5880	Gencode, RefSeq	99.59
CEP290	chr12:88442951-88535094	53	8509	Gencode, RefSeq	98.87
CEP41	chr7:130038722-130080817	12	1430	RefSeq	100.00
CETN2	chrX:151996375-151999264	5	619	Gencode, RefSeq, VEGA	100.00
CFC1	chr2:131350421-131356870	6	811	Gencode, RefSeq, VEGA	97.29
CFTR	chr7:117120139-117307172	27	5013	Gencode, RefSeq, VEGA	98.32
CLCNKB	chr1:16370978-16383790	20	2711	Gencode, RefSeq, VEGA	99.19
CLUAP1	chr16:3551058-3586281	13	1559	Gencode, RefSeq	100.00
COL4A1	chr13:110802700-110959384	55	6299	Gencode, RefSeq, VEGA	98.41
COL4A3	chr2:228029433-228176670	59	6482	Gencode, RefSeq, VEGA	97.98
COL4A4	chr2:227872031-228012209	48	6042	Gencode, RefSeq, VEGA	99.22
COL4A5	chrX:107683346-107939618	55	6235	Gencode, RefSeq, VEGA	97.83
CRB1	chr1:197237533-197447019	15	4784	Gencode, RefSeq, VEGA	99.81
CRX	chr19:48337691-48364779	5	1881	Gencode, RefSeq	100.00
CTNS	chr17:3543491-3564038	12	1606	Gencode, RefSeq, VEGA	100.00
CYP4F8	chr19:15726418-15740178	12	1802	Gencode, RefSeq	99.56
DAW1	chr2:228736396-228788694	13	1508	RefSeq	99.93
DCTN1	chr2:74588616-74617764	34	4733	Gencode, RefSeq, VEGA	99.98
DISC1	chr1:231762604-232172587	20	3963	Gencode, RefSeq, VEGA	94.55
DKK3	chr11:11986001-12030139	7	1235	Gencode, RefSeq	100.00
DLG1	chr3:196771484-197024085	29	3741	Gencode, RefSeq, VEGA	99.36
DNAAF1	chr16:84179036-84211457	12	2418	RefSeq	98.88
DNAAF2	chr14:50092250-50101877	3	2574	RefSeq	98.95
DNAAF3	chr19:55670420-55678026	12	2146	RefSeq	100.00

DNAH1	chr3:52356449-52434472	77	14421	Gencode, RefSeq, VEGA	99.92
DNAH10	chr12:124247057-124420038	78	14976	Gencode, RefSeq, VEGA	99.48
DNAH10OS	chr12:124418629-124419140	1	512	Gencode	100.00
DNAH11	chr7:21582854-21940882	84	15338	Gencode, RefSeq, VEGA	96.91
DNAH12	chr3:57327799-57528607	60	10732	Gencode, RefSeq, VEGA	99.22
DNAH14	chr1:225083954-225586981	93	16876	Gencode, RefSeq, VEGA	98.13
DNAH17	chr17:76419977-76571149	81	15269	Gencode, RefSeq, VEGA	99.59
DNAH2	chr17:7623043-7737067	88	15831	Gencode, RefSeq	99.28
DNAH3	chr16:20944466-21170772	63	14086	Gencode, RefSeq, VEGA	99.39
DNAH5	chr5:13692083-13944557	79	15455	Gencode, RefSeq, VEGA	99.46
DNAH6	chr2:84744941-85046542	76	14060	Gencode, RefSeq, VEGA	99.06
DNAH7	chr2:196602635-196935740	67	13583	Gencode, RefSeq, VEGA	99.42
DNAH8	chr6:38690576-38998178	92	16006	Gencode, RefSeq, VEGA	98.25
DNAH9	chr17:11501806-11872854	71	14977	Gencode, RefSeq, VEGA	99.43
DNAI1	chr9:34458994-34520764	21	2778	Gencode, RefSeq, VEGA	99.42
DNAI2	chr17:72277947-72310365	12	2058	Gencode, RefSeq	99.71
DNAJA2	chr16:46990931-47007493	9	1419	Gencode, RefSeq, VEGA	100.00
DNAJB6	chr7:157151257-157208802	10	1260	Gencode, RefSeq, VEGA	100.00
DNAJC25	chr9:114393678-114429230	6	1401	Gencode, RefSeq, VEGA	98.14
DNAL1	chr14:74111733-74162655	8	733	Gencode, RefSeq	99.32
DNAL4	chr22:39175444-39178746	3	432	Gencode, RefSeq, VEGA	100.00
DNALI1	chr1:38022520-38030672	6	963	Gencode, RefSeq, VEGA	99.17
DNHD1	chr11:6519436-6593226	40	15332	Gencode, RefSeq, VEGA	99.74
DPCD	chr10:103347199-103369238	8	1522	Gencode, RefSeq	100.00
DRC1	chr1:152382060-152384719	2	1528	Gencode, RefSeq, VEGA	100.00
DRC1	chr2:26624848-26679395	17	2563	RefSeq	100.00
DYDC1	chr10:82095902-82112367	6	705	Gencode, RefSeq, VEGA	100.00
DYDC2	chr10:82104529-82126717	4	656	Gencode, RefSeq, VEGA	99.39
DYNC1H1	chr14:102431019-102516910	78	15501	Gencode, RefSeq	99.45
DYNC1I1	chr7:95434032-95739393	18	2440	Gencode, RefSeq, VEGA	98.36
DYNC1I2	chr2:172546656-172604409	18	2342	Gencode, RefSeq, VEGA	95.77
DYNC1LI1	chr3:32568281-32612272	15	2062	Gencode, RefSeq, VEGA	98.30
DYNC1LI2	chr16:66757268-66785504	14	1816	Gencode, RefSeq, VEGA	100.00
DYNC2H1	chr11:102980294-103349991	91	14777	Gencode, RefSeq	98.52
DYNC2LI1	chr2:44001268-44036916	16	1822	Gencode, RefSeq, VEGA	87.05
DYNLL1	chr12:120934215-120936023	2	310	Gencode, RefSeq	100.00
DYNLL2	chr17:56164442-56166650	2	310	Gencode, RefSeq	100.00
DYNLRB1	chr20:33104239-33128438	7	889	Gencode, RefSeq, VEGA	100.00
DYNLRB2	chr16:80574964-80584465	5	477	Gencode, RefSeq, VEGA	100.00
DYNLT1	chr6:159057849-159065750	5	668	Gencode, RefSeq, VEGA	98.80
DYNLT3	chrX:37696457-37706773	7	685	Gencode, RefSeq, VEGA	82.34
DYSF	chr2:71681119-71913632	58	7650	Gencode, RefSeq, VEGA	99.10
DYX1C1	chr15:55648408-55790537	18	4195	Gencode, RefSeq, VEGA	98.78
EFCAB2	chr1:245133415-245285260	10	1188	Gencode, RefSeq, VEGA	97.31
EFHC1	chr6:52285199-52387766	13	2264	Gencode, RefSeq, VEGA	98.06
EFHC2	chrX:44008031-44202928	15	2639	Gencode, RefSeq, VEGA	97.50
EML1	chr14:100259804-100406459	23	2965	Gencode, RefSeq	98.15
ENKUR	chr10:25273289-25350087	7	952	Gencode, RefSeq, VEGA	97.69
EVC	chr4:5713098-5812774	22	3464	Gencode, RefSeq, VEGA	97.40
EVC2	chr4:5544848-5710250	23	4469	Gencode, RefSeq, VEGA	98.14
EYA1	chr8:72111565-72268731	19	2281	Gencode, RefSeq, VEGA	98.95
FAM161A	chr2:62053580-62081186	9	2443	Gencode, RefSeq, VEGA	95.91
FKBP5	chr6:35543602-35610611	10	1632	Gencode, RefSeq, VEGA	99.45
FOXJ1	chr17:74133424-74136486	2	1306	Gencode, RefSeq	100.00
FTO	chr16:53738087-54145837	11	1860	Gencode, RefSeq, VEGA	97.96

GHR	chr5:42424681-42719536	13	2258	Gencode, RefSeq, VEGA	100.00
GLI2	chr2:121554887-121748261	13	5021	Gencode, RefSeq, VEGA	99.84
GLI3	chr7:42003918-42262862	15	5278	Gencode, RefSeq, VEGA	99.19
GLIS2	chr16:4382272-4387535	6	1695	Gencode, RefSeq, VEGA	99.23
GPR98	chr5:89825289-90459727	94	20981	Gencode, RefSeq, VEGA	99.06
GSK3B	chr3:119545625-119812291	12	1542	Gencode, RefSeq, VEGA	98.44
GUCY2D	chr17:7906356-7919858	18	3672	Gencode, RefSeq, VEGA	99.73
HDAC6	chrX:48660330-48683388	30	4511	Gencode, RefSeq, VEGA	100.00
HEATR1	chr1:236714192-236767405	44	7319	Gencode, RefSeq, VEGA	98.29
HEATR2	chr7:766348-829200	16	3065	Gencode, RefSeq, VEGA	98.66
HEATR3	chr16:50100033-50138982	15	2343	Gencode, RefSeq, VEGA	99.32
HEATR4	chr14:73945301-73989866	16	3401	Gencode, RefSeq	100.00
HEATR5A	chr14:31762501-31872191	36	6921	Gencode, RefSeq	99.83
HEATR5B	chr2:37195601-37310567	37	7111	Gencode, RefSeq, VEGA	98.75
HEATR6	chr17:58120914-58156285	20	3948	Gencode, RefSeq	100.00
HNF1B	chr17:36047273-36104885	9	1947	Gencode, RefSeq, VEGA	99.44
HSPA14	chr10:14880392-14913615	18	2753	Gencode, RefSeq, VEGA	99.27
HSPA1A	chr6:31783524-31785469	1	1946	Gencode, RefSeq, VEGA	99.38
HSPB11	chr1:54387314-54405765	5	576	Gencode, RefSeq, VEGA	100.00
HTT	chr4:3076543-3241796	67	10773	Gencode, RefSeq, VEGA	99.03
HTT	chr17:28525464-28549849	14	2299	Gencode, RefSeq, VEGA	99.70
HYDIN	chr16:70841473-71264600	88	17452	Gencode, RefSeq, VEGA	90.79
HYDIN2	chr16:70841473-71264600	88	17452	Gencode, RefSeq, VEGA	90.79
IFT122	chr3:129159164-129239118	35	4816	Gencode, RefSeq, VEGA	97.99
IFT140	chr16:1560935-1657277	30	5005	Gencode, RefSeq, VEGA	99.80
IFT172	chr2:27667289-27712530	50	6384	Gencode, RefSeq, VEGA	99.17
IFT20	chr17:26655646-26659257	6	815	Gencode, RefSeq	98.77
IFT27	chr22:37154345-37171761	7	701	Gencode, RefSeq	99.14
IFT43	chr14:76452120-76549930	10	1016	RefSeq	97.15
IFT46	chr11:118415621-118430554	11	1288	Gencode, RefSeq	100.00
IFT52	chr20:42223329-42275633	13	1574	Gencode, RefSeq, VEGA	96.19
IFT57	chr3:107881314-107941179	11	1608	Gencode, RefSeq, VEGA	100.00
IFT74	chr9:26961956-27062744	23	2472	Gencode, RefSeq, VEGA	99.72
IFT80	chr3:159976303-160102398	21	2856	Gencode, RefSeq, VEGA	99.30
IFT81	chr12:110565176-110656041	19	2519	Gencode, RefSeq	100.00
IFT88	chr13:21142106-21265324	27	3166	Gencode, RefSeq, VEGA	98.61
IL15RA	chr10:5991243-6019570	10	1357	Gencode, RefSeq, VEGA	100.00
IMPDH1	chr7:128033051-128049965	17	2178	Gencode, RefSeq, VEGA	100.00
INF2	chr14:105167693-105185170	22	4223	Gencode, RefSeq, VEGA	99.48
INPP5E	chr9:139305111-139333881	23	4257	Gencode, RefSeq, VEGA	99.58
INVS	chr9:102866794-103062966	17	3735	Gencode, RefSeq, VEGA	100.00
IQCA1	chr2:237233321-237415920	22	3107	Gencode, RefSeq, VEGA	95.82
IQCB1	chr3:121489182-121547817	13	2057	Gencode, RefSeq, VEGA	97.76
IQCD	chr12:113633370-113645981	6	1851	Gencode, RefSeq	90.38
IQCE	chr7:2598806-2649806	25	3027	Gencode, RefSeq, VEGA	97.26
IQCG	chr3:197616441-197672508	10	1577	Gencode, RefSeq, VEGA	99.37
KAL1	chrX:8501026-8700087	14	2323	Gencode, RefSeq, VEGA	100.00
KAP3	chr1:169890807-170043618	21	2884	Gencode, RefSeq, VEGA	100.00
KCNJ1	chr11:128709010-128736457	3	1266	Gencode, RefSeq, VEGA	96.37
KIF11	chr10:94353123-94413563	22	3611	Gencode, RefSeq, VEGA	99.56
KIF18A	chr11:28042737-28119504	16	3017	Gencode, RefSeq	96.19
KIF19	chr17:72322479-72351461	20	3457	Gencode, RefSeq, VEGA	99.83
KIF20A	chr5:137515360-137523112	18	3033	Gencode, RefSeq, VEGA	99.51
KIF22	chr16:29802071-29816641	15	2325	Gencode, RefSeq, VEGA	100.00
KIF25	chr6:168430256-168445686	9	1345	Gencode, RefSeq, VEGA	98.22

KIF26A	chr14:104605050-104646137	15	5949	Gencode, RefSeq, VEGA	100.00
KIF2B	chr17:51900385-51902426	1	2042	Gencode, RefSeq	99.80
KIF3A	chr5:132032316-132073121	21	2638	Gencode, RefSeq, VEGA	99.51
KIF3B	chr20:30897571-30919132	8	2404	Gencode, RefSeq, VEGA	100.00
KIF3C	chr2:26151837-26204796	10	2647	Gencode, RefSeq, VEGA	98.49
KIF4B	chr5:154393410-154397134	1	3725	Gencode, RefSeq, VEGA	99.92
KIF5A	chr12:57944045-57976972	28	3659	Gencode, RefSeq	100.00
KIF7	chr15:90171640-90196171	18	4495	Gencode, RefSeq, VEGA	99.73
KIF9	chr3:47270132-47322488	22	2910	Gencode, RefSeq, VEGA	100.00
KIFC1	chr6:33359753-33377477	12	2385	Gencode, RefSeq, VEGA	99.25
KLC2	chr11:66026056-66034919	16	2637	Gencode, RefSeq, VEGA	100.00
KLC2	chr19:45848748-45854625	12	1797	Gencode, RefSeq, VEGA	99.72
KLC3	chr19:45848748-45854625	12	1797	Gencode, RefSeq, VEGA	99.72
KLC4	chr6:43027938-43042420	16	2443	Gencode, RefSeq, VEGA	100.00
LAMB2	chr19:2430899-2456941	12	2103	Gencode, RefSeq, VEGA	100.00
LAMB2	chr3:49158649-49170310	32	6037	Gencode, RefSeq, VEGA	100.00
LAMB2	chr1:182992842-183111935	28	5390	Gencode, RefSeq, VEGA	99.80
LCA5	chr6:80196711-80228621	7	2234	Gencode, RefSeq, VEGA	99.46
LCA9	chr1:10032122-10042769	5	984	Gencode, RefSeq, VEGA	100.00
LMTK2	chr7:97736480-97834814	14	4792	Gencode, RefSeq, VEGA	100.00
LMTK3	chr19:48989020-49016456	16	4790	Gencode, RefSeq	99.12
LRAT	chr4:155665469-155670298	2	733	Gencode, RefSeq, VEGA	98.36
LRRC1	chr6:53660045-53787601	15	1894	Gencode, RefSeq, VEGA	99.00
LRRC10	chr12:70003775-70004628	1	854	Gencode, RefSeq	100.00
LRRC10B	chr11:61276461-61277359	1	899	Gencode, RefSeq	100.00
LRRC14	chr8:145745100-145746872	3	1542	Gencode, RefSeq	100.00
LRRC14B	chr5:191644-195478	2	1585	Gencode, RefSeq, VEGA	98.99
LRRC15	chr3:194080017-194084109	2	1804	Gencode, RefSeq, VEGA	100.00
LRRC16A	chr6:25280014-25619821	38	5016	Gencode, RefSeq, VEGA	100.00
LRRC16B	chr14:24521353-24538623	40	4919	Gencode, RefSeq	99.86
LRRC17	chr7:102574351-102585064	4	1420	Gencode, RefSeq, VEGA	100.00
LRRC18	chr10:50118272-50122210	2	830	Gencode, RefSeq, VEGA	98.80
LRRC19	chr9:26995509-26999702	4	1193	Gencode, RefSeq, VEGA	97.65
LRRC2	chr3:46560494-46593091	8	1276	Gencode, RefSeq, VEGA	98.35
LRRC20	chr10:72061100-72136299	4	635	Gencode, RefSeq, VEGA	98.43
LRRC23	chr12:7014788-7023245	7	1425	Gencode, RefSeq, VEGA	99.58
LRRC24	chr8:145747849-145750368	4	1622	Gencode, RefSeq	99.26
LRRC25	chr19:18502787-18507783	2	958	Gencode, RefSeq	100.00
LRRC26	chr9:140063296-140064405	2	1045	Gencode, RefSeq, VEGA	100.00
LRRC27	chr10:134147019-134188756	14	2184	Gencode, RefSeq, VEGA	97.85
LRRC28	chr15:99796153-99926317	9	1284	Gencode, RefSeq, VEGA	99.61
LRRC29	chr16:67241498-67244101	4	1091	Gencode, RefSeq, VEGA	100.00
LRRC3	chr21:45876518-45877311	1	794	Gencode, RefSeq, VEGA	100.00
LRRC30	chr18:7231127-7232052	1	926	Gencode, RefSeq	100.00
LRRC31	chr3:169557760-169587605	9	1857	Gencode, RefSeq, VEGA	98.28
LRRC32	chr11:76370638-76377008	2	2029	Gencode, RefSeq, VEGA	100.00
LRRC34	chr3:169511413-169530346	11	1637	Gencode, RefSeq, VEGA	97.43
LRRC36	chr16:67360756-67419002	14	2573	Gencode, RefSeq	97.40
LRRC37A	chr17:44372490-62893385	26	10533	Gencode, RefSeq, VEGA	31.44
LRRC37A2	chr17:44590069-44632955	14	5383	Gencode, RefSeq	35.04
LRRC37A3	chr17:62850707-62893385	12	5150	Gencode, RefSeq, VEGA	35.84
LRRC37B	chr17:30348075-30380357	12	3165	Gencode, RefSeq, VEGA	85.50
LRRC38	chr1:13802304-13840098	2	925	Gencode, RefSeq, VEGA	100.00
LRRC39	chr1:100614588-100634082	9	1256	Gencode, RefSeq, VEGA	95.46
LRRC3B	chr3:26751154-26751953	1	800	Gencode, RefSeq, VEGA	100.00

LRRC3C	chr17:38097767-38100997	2	868	RefSeq	100.00
LRRC4	chr7:127668722-127670703	1	1982	Gencode, RefSeq, VEGA	100.00
LRRC40	chr1:70611473-70671233	15	2109	Gencode, RefSeq, VEGA	98.58
LRRC41	chr1:46744367-46769004	11	2829	Gencode, RefSeq, VEGA	99.65
LRRC42	chr1:54417663-54433622	7	1427	Gencode, RefSeq, VEGA	100.00
LRRC43	chr12:122667682-122687999	13	2399	Gencode, RefSeq	100.00
LRRC45	chr17:79981510-79988691	17	2353	Gencode, RefSeq	100.00
LRRC46	chr17:45909346-45914496	8	1126	Gencode, RefSeq	100.00
LRRC47	chr1:3697642-3713050	7	1892	Gencode, RefSeq, VEGA	90.70
LRRC48	chr17:17880903-17919995	13	1882	Gencode, RefSeq, VEGA	98.41
LRRC49	chr15:71165535-71341961	19	2644	Gencode, RefSeq	99.02
LRRC4B	chr19:51020818-51052105	2	2182	Gencode, RefSeq	100.00
LRRC4C	chr11:40135910-40137852	1	1943	Gencode, RefSeq	100.00
LRRC52	chr1:165513524-165533071	2	982	Gencode, RefSeq, VEGA	100.00
LRRC55	chr11:56949358-56954964	2	1066	Gencode, RefSeq, VEGA	100.00
LRRC56	chr11:540675-554286	11	1849	Gencode, RefSeq, VEGA	100.00
LRRC57	chr15:42836271-42840642	5	820	Gencode, RefSeq, VEGA	100.00
LRRC58	chr3:120050037-120068100	4	1196	Gencode, RefSeq, VEGA	99.41
LRRC59	chr17:48460339-48474688	7	1064	Gencode, RefSeq, VEGA	100.00
LRRC6	chr8:133584544-133687749	16	1885	Gencode, RefSeq, VEGA	99.68
LRRC61	chr7:150033941-150034740	1	800	Gencode, RefSeq, VEGA	100.00
LRRC63	chr13:46787160-46850948	11	2180	Gencode, RefSeq, VEGA	100.00
LRRC66	chr4:52860535-52883789	4	2723	Gencode, RefSeq, VEGA	100.00
LRRC69	chr8:92114880-92231238	9	1251	Gencode, RefSeq, VEGA	96.72
LRRC7	chr1:70034313-70587580	30	5342	Gencode, RefSeq, VEGA	98.88
LRRC70	chr5:61875256-61877144	1	1889	Gencode, RefSeq, VEGA	99.47
LRRC71	chr1:156890586-156902771	15	1980	Gencode, RefSeq	99.85
LRRC72	chr7:16566568-16621124	9	1044	RefSeq	99.81
LRRC73	chr6:43474966-43477533	6	1071	RefSeq	100.00
LRRC8A	chr9:131669434-131678660	2	2473	Gencode, RefSeq, VEGA	100.00
LRRC8B	chr1:90048200-90058612	2	2452	Gencode, RefSeq, VEGA	100.00
LRRC8C	chr1:90152023-90180551	2	2452	Gencode, RefSeq, VEGA	100.00
LRRC8D	chr1:90398618-90401214	1	2597	Gencode, RefSeq, VEGA	95.46
LRRC8E	chr19:7960479-7965808	2	2431	Gencode, RefSeq	100.00
LRRCC1	chr8:86019521-86057756	19	3527	Gencode, RefSeq, VEGA	98.21
LZTFL1	chr3:45867796-45957120	13	1505	Gencode, RefSeq, VEGA	100.00
MAPK14	chr6:35995925-36076234	14	1530	Gencode, RefSeq, VEGA	99.22
MAPT	chr17:44039694-44101547	16	2670	Gencode, RefSeq	98.58
MCIDAS	chr5:54516184-54522976	7	1298	RefSeq	98.15
MCKD1	chr1:155158601-155162644	10	2080	Gencode, RefSeq, VEGA	82.02
MKKS	chr20:10385885-10394172	4	1793	Gencode, RefSeq, VEGA	99.16
MKS1	chr17:56282902-56296882	21	2308	Gencode, RefSeq, VEGA	100.00
MLF1	chr3:158289080-158323001	10	1264	Gencode, RefSeq, VEGA	100.00
MNS1	chr15:56721289-56757180	10	1688	Gencode, RefSeq, VEGA	99.29
MRE11A	chr11:94153281-94225977	21	2583	Gencode, RefSeq	99.96
MTOR	chr1:11166652-11319476	60	9217	Gencode, RefSeq, VEGA	99.72
MYH9	chr22:36678285-36745291	42	6900	Gencode, RefSeq, VEGA	99.86
MYO1E	chr15:59429569-59664709	28	3887	Gencode, RefSeq	100.00
NCOA3	chr20:46250982-46282171	21	4725	Gencode, RefSeq, VEGA	99.20
NEK8	chr14:75551257-75593634	22	3451	Gencode, RefSeq	99.77
NEK8	chr17:27055822-27069015	14	2519	Gencode, RefSeq	100.00
NF1	chr17:29422318-29705959	60	9902	Gencode, RefSeq, VEGA	96.70
NGS17	chr6:33269527-33281828	9	1822	Gencode, RefSeq, VEGA	99.07
NME1	chr17:49231725-49239216	6	679	Gencode, RefSeq, VEGA	100.00
NME1-NME2	chr17:49231725-49248975	8	1039	Gencode, RefSeq, VEGA	100.00

NME2	chr17:49231725-49248975	8	1039	Gencode, RefSeq, VEGA	100.00
NME3	chr16:1820640-1821545	5	610	Gencode, RefSeq, VEGA	99.34
NME4	chr16:447213-456375	12	1176	Gencode, RefSeq, VEGA	98.89
NME5	chr5:137451360-137474479	5	739	Gencode, RefSeq, VEGA	100.00
NME6	chr3:48336117-48342812	8	815	Gencode, RefSeq, VEGA	100.00
NME7	chr1:169102013-169336958	12	1371	Gencode, RefSeq, VEGA	100.00
NME8	chr7:37889859-37936704	15	2067	RefSeq	98.74
NME9	chr3:137981372-138043749	9	972	RefSeq	100.00
NODAL	chr10:72192682-72207717	4	1152	Gencode, RefSeq, VEGA	100.00
NOS2	chr19:46417525-46417961	1	437	Gencode, RefSeq	100.00
NOS2	chr17:26084262-26125845	26	3982	Gencode, RefSeq, VEGA	97.14
NPHP1	chr2:110881358-110962555	22	2883	Gencode, RefSeq, VEGA	100.00
NPHP3	chr3:132363641-132441209	31	4824	Gencode, RefSeq, VEGA	97.76
NPHP4	chr1:5923315-6046359	31	5062	Gencode, RefSeq, VEGA	99.98
NPHS1	chr19:36317406-36342749	29	4306	Gencode, RefSeq	100.00
NPHS2	chr1:179520298-179545009	8	1312	Gencode, RefSeq, VEGA	99.47
NR3C1	chr5:142658919-142780414	9	2576	Gencode, RefSeq, VEGA	99.73
OFD1	chrX:13753181-13787237	25	3591	Gencode, RefSeq, VEGA	99.92
PAFAH1B1	chr17:2541573-2585106	10	1433	Gencode, RefSeq	96.72
PARF	chr9:139690840-139734984	30	4697	Gencode, RefSeq, VEGA	99.30
PAX2	chr10:102506008-102587450	12	1645	Gencode, RefSeq, VEGA	100.00
PCDH15	chr10:55566329-56424032	45	8504	Gencode, RefSeq, VEGA	97.22
PCDP1	chr2:120303698-120414056	23	2983	RefSeq	99.43
PCLN1	chr3:190105899-190127835	5	1018	Gencode, RefSeq, VEGA	99.12
PDE6D	chr2:232597652-232645834	6	633	Gencode, RefSeq, VEGA	100.00
PIH1D3	chrX:106456096-106486538	6	765	RefSeq	100.00
PKD1	chr16:2139718-2185700	47	13987	Gencode, RefSeq, VEGA	80.45
PKD2	chr4:88928876-88996856	15	3207	Gencode, RefSeq, VEGA	98.66
PKD2	chr19:47177770-47219637	18	2997	Gencode, RefSeq	99.53
PKHD1	chr6:51483869-51949741	69	13672	Gencode, RefSeq, VEGA	98.89
PLCE1	chr10:95790794-96087767	33	7884	Gencode, RefSeq, VEGA	99.85
PMFBP1	chr16:72146650-72205117	21	3651	Gencode, RefSeq	100.00
POLM	chr7:44112651-44122047	12	2232	Gencode, RefSeq, VEGA	100.00
PSF-1	chr6:32813346-32821603	11	2647	Gencode, RefSeq, VEGA	99.70
PSF1	chr20:25388447-25426637	8	873	Gencode, RefSeq, VEGA	92.44
PSF1	chr6:32813346-32821603	11	2647	Gencode, RefSeq, VEGA	99.70
RABL5	chr7:100958405-100965000	5	689	Gencode, RefSeq, VEGA	100.00
RD3	chr1:211652368-211654767	2	628	Gencode, RefSeq, VEGA	99.68
RDH12	chr14:68189350-68200575	7	1091	Gencode, RefSeq	100.00
REN	chr17:7256252-7256970	1	719	Gencode, RefSeq, VEGA	100.00
REN	chr1:204124134-204135431	11	1450	Gencode, RefSeq, VEGA	98.14
RFX3	chr9:3225032-3395598	19	3032	Gencode, RefSeq, VEGA	99.70
RIBC2	chr22:45809773-45828252	7	1274	Gencode, RefSeq, VEGA	99.69
RING4	chr6:32813346-32821603	11	2647	Gencode, RefSeq, VEGA	99.70
ROPN1L	chr5:10442270-10465069	5	793	Gencode, RefSeq, VEGA	100.00
RP1	chr18:32558472-32720343	9	1250	Gencode, RefSeq, VEGA	100.00
RP1	chr8:55533517-55682383	23	9465	Gencode, RefSeq	99.14
RP1	chr6:31939764-31981574	9	1334	Gencode, RefSeq, VEGA	100.00
RP2	chrX:46696526-46739214	5	1153	Gencode, RefSeq, VEGA	94.02
RP2	chr19:33182857-33202873	3	1188	Gencode, RefSeq	97.64
RPE65	chr1:68895449-68915598	14	1882	Gencode, RefSeq, VEGA	99.36
RPGR	chrX:38128869-38186630	20	5131	Gencode, RefSeq, VEGA	79.73
RPGRIP1	chr14:21756126-21819385	25	4462	Gencode, RefSeq	98.86
RPGRIP1L	chr16:53635978-53734645	26	4468	Gencode, RefSeq	98.72
RSPH1	chr21:43892918-43916306	9	1110	Gencode, RefSeq, VEGA	99.55

RSPH10B	chr7:5965989-6838184	38	5986	Gencode, RefSeq, VEGA	94.87
RSPH10B2	chr7:5965989-6838184	38	5986	Gencode, RefSeq	94.87
RSPH3	chr6:159398560-159421018	8	1843	Gencode, RefSeq, VEGA	95.93
RSPH4A	chr6:116937777-116953614	6	2271	Gencode, RefSeq	98.55
RSPH6A	chr19:46299117-46318444	6	2274	Gencode, RefSeq	99.69
RSPH9	chr6:43612826-43640197	7	1210	Gencode, RefSeq, VEGA	100.00
RTDR1	chr22:23401630-23482617	9	1433	Gencode, RefSeq, VEGA	99.30
RTTN	chr18:67671377-67872904	50	7716	Gencode, RefSeq	99.26
SCNN1A	chr12:6457029-6486442	14	2548	Gencode, RefSeq	95.68
SCNN1B	chr16:23315311-23392132	14	2338	Gencode, RefSeq, VEGA	100.00
SCNN1D	chr1:1216032-1227000	17	2829	Gencode, RefSeq, VEGA	100.00
SCNN1G	chr16:23197583-23226800	12	2190	Gencode, RefSeq, VEGA	99.13
SDCCAG8	chr1:243419466-243663097	19	2537	Gencode, RefSeq, VEGA	100.00
SESN1	chr6:109308737-109415286	11	1978	Gencode, RefSeq, VEGA	99.70
SIX1	chr14:61112991-61115917	2	895	Gencode, RefSeq, VEGA	100.00
SIX5	chr19:46268749-46272112	3	2280	Gencode, RefSeq	99.65
SLC12A1	chr15:48499907-48595092	29	4057	Gencode, RefSeq, VEGA	99.29
SLC12A3	chr16:56899138-56947327	26	3613	Gencode, RefSeq, VEGA	99.25
SLC26A10	chr12:58013994-58019538	15	2099	Gencode, RefSeq, VEGA	100.00
SLC3A1	chr2:44502665-44547788	12	2632	Gencode, RefSeq, VEGA	98.02
SLC4A1	chr17:42327816-42340244	19	3116	Gencode, RefSeq, VEGA	99.74
SLC4A4	chr4:72102284-72433620	25	4173	Gencode, RefSeq, VEGA	97.20
SLC7A9	chr19:33321516-33359450	12	1704	Gencode, RefSeq	99.94
SPA17	chr11:124545151-124564352	4	536	Gencode, RefSeq	100.00
SPAG1	chr8:101174499-101253260	18	3296	Gencode, RefSeq, VEGA	99.15
SPAG11A	chr8:7705559-7721067	7	836	Gencode, RefSeq, VEGA	100.00
SPAG11B	chr8:7305527-7707675	10	1238	Gencode, RefSeq, VEGA	100.00
SPAG16	chr2:214149198-215275049	21	2907	Gencode, RefSeq, VEGA	98.14
SPAG17	chr1:118506412-118727790	48	7643	Gencode, RefSeq, VEGA	99.25
SPAG4	chr20:34203916-34208954	12	1581	Gencode, RefSeq, VEGA	100.00
SPAG5	chr17:26904707-26925974	24	4062	Gencode, RefSeq	99.26
SPAG6	chr10:22634531-22705627	13	2125	Gencode, RefSeq, VEGA	98.45
SPAG7	chr17:4862819-4871109	7	824	Gencode, RefSeq	98.67
SPAG8	chr9:35808160-35812154	9	1899	Gencode, RefSeq, VEGA	100.00
SPAG9	chr17:49042074-49198027	39	5208	Gencode, RefSeq, VEGA	97.96
SPATA7	chr14:88852153-88904776	12	2040	Gencode, RefSeq	96.62
SPEF2	chr5:35618090-35814665	48	7966	Gencode, RefSeq, VEGA	98.54
TAP1	chr22:30793096-30818406	15	1796	Gencode, RefSeq, VEGA	100.00
TAP1	chr6:32813346-32821603	11	2647	Gencode, RefSeq, VEGA	99.70
TAP2	chr6:32782244-32806020	15	2817	Gencode, RefSeq, VEGA	99.22
TAP2	chr22:30845064-30867955	15	1766	Gencode, RefSeq, VEGA	100.00
TAPA	chr6:33269527-33281828	9	1822	Gencode, RefSeq, VEGA	99.07
TAPBP	chr6:33269527-33281828	9	1822	Gencode, RefSeq, VEGA	99.07
TAPBPL	chr12:6561405-6571325	7	1547	Gencode, RefSeq	100.00
TCTE3	chr6:170140271-170151564	4	677	Gencode, RefSeq, VEGA	100.00
TCTN1	chr12:111051978-111085707	16	2270	Gencode, RefSeq, VEGA	96.52
TCTN2	chr12:124155778-124192270	18	2454	Gencode, RefSeq	99.02
TEKT2	chr1:36550513-36553797	9	1473	Gencode, RefSeq, VEGA	100.00
TMEM216	chr11:61160094-61165758	5	547	Gencode, RefSeq	100.00
TMEM231	chr16:75573882-75590179	6	1230	Gencode, RefSeq	100.00
TMEM67	chr8:94767133-94828690	32	3974	Gencode, RefSeq, VEGA	93.73
TPN	chr6:33269527-33281828	9	1822	Gencode, RefSeq, VEGA	99.07
TPPP3	chr16:67424067-67425024	3	591	Gencode, RefSeq, VEGA	100.00
TPSN	chr6:33269527-33281828	9	1822	Gencode, RefSeq, VEGA	99.07
TRIM32	chr9:119460012-119461993	1	1982	Gencode, RefSeq, VEGA	99.90

TRPC6	chr11:101323676-101454244	13	3056	Gencode, RefSeq	98.92
TRPM6	chr9:77339519-77502782	43	7171	Gencode, RefSeq, VEGA	99.34
TSC1	chr9:135771612-135804269	21	3987	Gencode, RefSeq, VEGA	99.92
TSC2	chr16:2098261-2138621	42	6297	Gencode, RefSeq, VEGA	97.78
TTBK2	chr15:43037983-43170825	18	5377	Gencode, RefSeq	99.50
TTC21B	chr2:166731255-166810225	29	4532	Gencode, RefSeq, VEGA	99.10
TTC30A	chr2:178481422-178483439	1	2018	Gencode, RefSeq, VEGA	95.64
TTC30B	chr2:178415484-178417501	1	2018	Gencode, RefSeq, VEGA	98.36
TTC8	chr14:89291042-89343764	17	2168	Gencode, RefSeq	94.28
TUB	chr11:8040840-8123176	15	2080	Gencode, RefSeq	100.00
TUBA1A	chr12:49578783-49582772	5	1459	Gencode, RefSeq	95.20
TUBA1B	chr12:49521731-49525093	4	1436	Gencode, RefSeq	94.71
TUBA1C	chr12:49621727-49667126	6	1810	Gencode, RefSeq	97.68
TUBA3C	chr13:19747993-19755897	5	1453	Gencode, RefSeq, VEGA	93.32
TUBA3D	chr2:132233763-132240431	5	1453	Gencode, RefSeq, VEGA	99.11
TUBA3E	chr2:130949394-130955943	5	1453	Gencode, RefSeq, VEGA	73.98
TUBA4A	chr2:220115064-220118590	6	1548	Gencode, RefSeq, VEGA	100.00
TUBA8	chr22:18593622-18613913	5	1525	Gencode, RefSeq, VEGA	100.00
TUBAL3	chr10:5435470-5446765	4	1421	Gencode, RefSeq, VEGA	99.30
TUBB1	chr20:57594568-57599848	4	1436	Gencode, RefSeq, VEGA	100.00
TUBB1	chr6:30688274-30692184	5	1438	Gencode, RefSeq, VEGA	98.33
TUBB2A	chr6:3154087-3157707	4	1418	Gencode, RefSeq	94.36
TUBB2B	chr6:3224975-3227787	4	1418	Gencode, RefSeq, VEGA	98.03
TUBB3	chr16:89989800-90002222	4	1433	Gencode, RefSeq, VEGA	96.09
TUBB4A	chr19:6495165-6502233	4	1415	RefSeq	100.00
TUBB4B	chr9:140135803-140138018	4	1418	RefSeq	99.58
TUBB6	chr18:12308282-12326139	4	1421	Gencode, RefSeq, VEGA	99.79
TUBB8	chr10:92987-95188	4	1415	RefSeq	76.54
TUBD1	chr17:57937673-57968373	8	1535	Gencode, RefSeq	100.00
TUBE1	chr6:112392605-112408647	13	1752	Gencode, RefSeq, VEGA	100.00
TUBG2	chr17:40811512-40818828	11	1576	Gencode, RefSeq	99.37
TULP1	chr6:35466094-35480645	15	1944	Gencode, RefSeq, VEGA	100.00
TULP2	chr19:49384258-49401135	12	1836	Gencode, RefSeq, VEGA	100.00
TULP3	chr12:3000104-3049902	13	1820	Gencode, RefSeq	99.89
TULP4	chr6:158735039-158927736	14	4912	Gencode, RefSeq, VEGA	100.00
UMOD	chr16:20344626-20362069	11	2242	Gencode, RefSeq, VEGA	99.46
USH1G	chr17:72914158-72919178	3	1446	Gencode, RefSeq	100.00
USH2A	chr1:215799113-216595688	72	17141	Gencode, RefSeq, VEGA	99.53
VHL	chr3:10183522-10191659	3	702	Gencode, RefSeq, VEGA	93.59
WDR19	chr4:39184168-39280280	36	4774	Gencode, RefSeq, VEGA	99.85
WDR34	chr9:131396013-131419015	9	1934	Gencode, RefSeq, VEGA	100.00
WDR35	chr2:20113309-20189786	28	4142	Gencode, RefSeq, VEGA	99.59
WDR60	chr7:158649417-158738480	26	3766	Gencode, RefSeq, VEGA	99.73
WT1	chr11:32410594-32456901	11	1784	Gencode, RefSeq, VEGA	100.00
ZIC3	chrX:136648841-136659440	4	1634	Gencode, RefSeq, VEGA	99.94
ZMYND10	chr3:50378831-50383020	12	1653	Gencode, RefSeq, VEGA	100.00
ZMYND11	chr10:225943-298420	16	2260	Gencode, RefSeq, VEGA	99.87
ZMYND12	chr1:42896397-42921678	8	1258	Gencode, RefSeq, VEGA	99.36
ZMYND15	chr17:4643834-4649295	13	2513	Gencode, RefSeq, VEGA	100.00
ZMYND19	chr9:140476985-140484725	6	804	Gencode, RefSeq, VEGA	99.00
ZMYND8	chr20:45839396-45985424	25	4284	Gencode, RefSeq, VEGA	99.35
ZNF423	chr16:49525176-49856606	8	4015	Gencode, RefSeq	100.00

VariantMaster - v1.0

Federico Andrea Santoni

Federico.santoni@unige.ch

Department of Medical Genetics and Development
University of Geneva, Switzerland

Contents

1. Download and Installation

VariantMaster is a python package delivered as a binary executable program for Linux based systems. Source code could be provided upon request to the author.

DISCLAIMER

Neither the University of Geneva nor the author assume any responsibility whatsoever for its use by other parties and makes no guarantees, expressed or implied, about its quality, reliability, or any other characteristics in particular for the use of this tool in clinical contexts.

By downloading the software from this page, you agree to the specified terms.

Download:

http://sourceforge.net/projects/variantmaster/

To use:

```
> gunzip VariantMaster_1.0.gz
> VariantMaster_1.0 -h
Usage: VariantMaster -c <cfg_file> -t <tfam_file> -o <output_prefix> [-s <csv_folder>
|-v <vcf_folder>] -b <bam_folder> -H <hap_file [optional]> <options>

Options:
  -h, --help            show this help message and exit
  -v VCF, --vcf=VCF     Path to vcf folder or to a multisample vcf file
  -s CSV, --csv=CSV     Path to csv folder
  -b BAM, --bam=BAM     Path to bam folder
  -c CONFIGURATION, --configuration=CONFIGURATION
                        Configuration file
  -o OUTPUT, --output=OUTPUT
                        Output filename prefix
  -t TFAM, --tfam=TFAM  pathway to pedigree structure (tfam format)
  -H HAPLOTYPE, --haplotype=HAPLOTYPE
                        Path to the Haplotype file (tped format)
  -R, --report          Only the report please!
  -A, --annotate        Annotate only!
```

VariantMaster uses Annovar to annotate VCF files. Last Annovar release and accompanying databases can be downloaded from http://www.openbioinformatics.org/annovar/.

Quick start

- create a VCF folder with your VCF files (multisample or one per sample)
- create a configuration file (see “Write the configuration file”) specifying Annovar coordinates
- create a TFAM file (see “Structure of the pedigree(TFAM) file”)
- annotate VCFs using -A option ONCE

```
> VariantMaster -c cfg_file -t <tfam_file>  -v vcf_folder -b <bam_folder> -A
```

- analyze the CSV files created by VariantMaster inside the folder *masterCSV*

```
> VariantMaster -c cfg_file -t <tfam_file>  -s masterCSV -b <bam_folder> -o <output-
prefix>
```

2. Configure the input data

Creating the workspace

VariantMaster requires several input files depending on what is available for the analysis. To perform the pedigree analysis it needs:

1a. one folder (e.g. named "VCF") containing a file VCF (Variant Call Format) for at least one affected individual OR a multisample VCF;

OR

1b. one folder (e.g. "CSV") containing a file CSV for at least one affected individual (see 2.5 to adapt annotated text files to VariantMaster format);

3. one TFAM file (PLINK format - see 2.3);

4. one folder (e.g. BAM) containing the BAM files (Binary Aligned Mapped reads) and the BAI (BAM Index) files for all sequenced individual - it's not mandatory but considerably improves the accuracy of the algorithm;

5. one configuration file (see 2.2).

For matched Tumor-Germline analysis or for Unrelated individuals, VariantMaster requires:

1. one folder containing VCF or CSV files from samples
2. one TFAM file describing the pairs Germline - Tumor
3. one folder containing the BAM and BAI files from all samples
4. one configuration file

If all individuals have been sequenced, it is strongly advised to prepare a workspace with the following setup:

```
CSV/ (or VCF/)      BAM/   study.cfg    families.tfam
```

where CSV (or VCF) is prepared as:

```
> mkdir CSV
> ln -s  <path_to_csv_files>  CSV/
```

Analogously, BAM is created as:

```
> mkdir BAM
> ln -s  <path_to_bam_files>  BAM/
```

If only few individuals were sequenced but all haplotypes are available (see 2.4 to create haplotypes from genotyped data) then the workspace would be similar to:

```
CSV/ (or VCF/)      BAM/   HAP/   study.cfg families.tfam
```

where HAP is created with:

```
> mkdir HAP
> ln -s  <path_to_csv_files>  HAP/
```

It is worth to say again that the usage of CSV or VCF is mutually exclusive.

The optional switches -A and -R solely allow to annotate VCF files (the CSV folder is produced) and to recalculate the final report, respectively.

IMPORTANT:

All file names (CSV, VCF, BAM or HAP) associated to an individual must contain the string *_ID_* where *ID* is the identifier of the individual as reported in the TFAM file (2.3).

- in a multisample VCF file, the name of each sample must be the related ID

Write the Configuration file

The configuration file contains some general settings.

CARR_THR is the threshold probability to consider a sample as a carrier of a given SNV (0.95 is the default),

HOM_THR allelic percentage to consider a variant as homozygous (75% is the default),

reference full path of the genomic fasta (and respective .fai) reference file to be used for INDEL processing.

gene_reference full path of the gene reference in the UCSC format. It is used for the final report.

annovar path to annovar scripts.

annovar_db path to annovar databases.

Variables are defined by the syntax

```
Var := value
```

Each analysis module has its specific filter and it is introduced by the construct:

```
Filter_NamemoduleFilter
..
END_Filter
```

Following modules are available: `DenovoFilter, DominantFilter, XLinkedFilter, RecessiveFilter, TumorFilter, IndependentFilter.`

IMPORTANT
Specific filtering operations are executed through `select`, `conditional_select` and `remove`. These commands act over the fields specified in the header of the annotated CSV file.

`select Field Operator Value`

select for the variants having a value in the specified Field satisfying the condition defined by Operator (one among >,>=,<,<=, == ,!=,in) and Value.

example:

```
select  Func    in      exonic;splicing
```

select for variants having exonic, splicing or exonic;splicing in the field Func

---

`conditional_select Field Operator Value Field Operator Value`

execute a select defined by the second triplet only if the first triplet is satisfied

example:

```
conditional_select      INFO    in      PINDEL  QS      >=      600
```

select for PINDEL variants having quality score >=600

`remove  Field  -  -`

will remove all variants with whatever value in Field

example:

```
remove  segdup  -       -
```

Here is an example of a configuration file for a Denovo filter:

```
#General variables
CARR_THR := 0.9
HOM_THR := 0.6
reference := '/data/hg19/hg19.unmasked.fa'
annovar := '/usr/local/annovar'
annovar_db := '/usr/local/annovar_db'
#Filter definitions
Filter_DenovoFilter
conditional_select      INFO    in      PINDEL  QS      >=      600
conditional_select      INFO    in      INDEL   QS      >=      200
select  Func    in      exonic;splicing
select  QS      >=      50
```

```
select  ExonicFunc      !=      synonymous SNV
select  Zyg     ==      het
remove  segdup  -       -
select  1000g   <=      0.01
select  snp137_freq        <=      0.01
select  esp6500     <=      0.01
End_Filter
```

IMPORTANT
Commands, operators and values are TAB separated!!!

Structure of the Pedigree (TFAM) file

The TFAM format, as defined in PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/), is

```
Family_ID Individual_ID Father_ID Mother_ID Sex Phenotype
```

where sex is (0 - male, 1 - female) and Phenotype can be (1 - unaffected, 2 - affected)

For example, two nuclear families with unaffected parents and, respectively, one affected boy and two children, one affected girl and one unaffected boy may be represented as:

```
Fam001 Fath_001 0 0 0 1
Fam001 Moth_001 0 0 1 1
Fam001 Boy_001 Fath_001 Moth_001 0 2
Fam002 Fath_002 0 0 0 1
Fam002 Moth_002 0 0 1 1
Fam002 Boy_002 Fath_002 Moth_002 0 1
Fam002 Girl_002 Fath_002 Moth_002 1 2
```

IMPORTANT
All files (CSV, VCF, HAP or BAM) related to a specific individual ID must be named as *whatever_ID_somethingelse*. In other words:
- in a multisample VCF file, the name of each sample must be ID
- the ID reported in the TFAM file has to be repeated in filenames between underscores (_ID_)
- a CSV file with ID = *Fath001* should be named *XYZ_Fath001_KJH.csv* (the extension *.csv* is mandatory).

Creating haplotypes from genotyped data (* NOT TESTED)

VariantMaster accepts haplotypes in the form of TPED files. TPED format is:

The easiest way to generate haplotypes compatible with VariantMaster from genotyped data is to download Beagle (http://faculty.washington.edu/browning/beagle/beagle.jar) and the script

http://seaseq.unige.ch/~fsantoni/VariantMaster/tools/ped2beagle2ped.py.

First, we start from genotyping data in the transposed PLINK file format TPED *(chr, snpname, centimorgan, position, [genotypes]* see *http://pngu.mgh.harvard.edu/~purcell/plink/data.shtml*).

After having installed Beagle,

```
> ped2beagle2ped.py -h

Usage: ped2beagle2ped.py -c "beagle_cmd"
Options:
  -h, --help            show this help message and exit
  -c CMD, --cmd=CMD     java -Xmx2048m -jar path_to_beagle.jar
                        <trio/phased/unphased>=%s out=<output_prefix>
                        missing=?
  -t TPED, --tped=TPED  tped file
```

by applying the script to the file *Fam01.tped* (unphased genotypes)

```
> ped2beagle2ped.py -c "java -Xmx2048m -jar /usr/local/beagle.jar unphased=%s
out=example1 missing=? -t Fam01.tped"
```

we get as output *example1.Fam01.tped.beagle.phased.* Note that here beagle.jar is in */usr/local*.

```
1 rs10492937 3339339 C C C C C C C C C C C C
1 rs4648500 3340855 C C C C C C C C C C C C
1 rs2244013 3342530 G G G G G G G G G G G G
1 rs870171 3342804 C C C C C C C C C C C C
1 rs2493272 3349513 A A A A A A A A A A A A
1 rs1537406 3352227 G G G G G G G G G G G G
1 rs2236518 3352872 A A A A A A A A A A A A
```

The format (*chr, snpname,position, [genotypes]*) describes a VariantMaster compliant haplotype.

Convert any annotated variant file to be VariantMaster compliant

It is relatively easy to convert structured text files to be VariantMaster CSV compliant.

The only requirements are:

a. all files need to have the same header so they can be addressed by the filtering system;
b. first three columns must be **Chr, Start, End;**

c. additional **mandatory fields** are**: Zyg ={hom,het}; INFO** beginning with **INDEL** if the variant is an insertion/deletion and containing **DP4** as in the VCF format.
d. fields are comma separated (CSV).

For example, it is possible to use awk as follows

```
awk -F, -v OFS="," '{print $4,$5,$6,$0}' file
```

to rearrange the following file

```
Gene;Function;ExonicFunction;Chr;Start;End;Zyg;dbSNP;esp6500;INFO
FOX3;exonic;non synonymous;chr17;77090533;77090533;hom;rs17344067;0;DP4=1,1,34,56
…
```

IMPORTANT
Avoid spaces in field names

3. Running the program

Example: Nuclear family

The family **fam01** is composed by unaffected father F, mother M, daughter D and one affected son (S). all members have been been sequenced. The multisample VCF file with samples *_F[M,D,S]_* is available and, accordingly, BAM files are named *_ F[M,D,S]_fam01.bam*.
The annotated CSV files are generated by VariantMaster in the folder *./masterCSV* with names *seq_F[M,D,S]_fam01.csv.*
The header of all CSV files is:

```
Chr,Start,End,Func,ExonicFunc,QS,Zyg,segdup,1000g,dbSNP,snp_freq,esp,...
```

Therefore, the family structure is (*fam01.tfam*):

```
fam01 F 0 0 0 1
fam01 M 0 0 1 1
fam01 S F M 0 2
fam01 D F M 1 1
```

The analysis is executed on the denovo, recessive and X linked model. Accordingly, the configuration file is (fam01.cfg - REMEMBER - it is a **TAB delimited file**):

```
#General variables
CARR_THR := 0.95
HOM_THR := 0.75
reference := '/data/hg19/hg19.unmasked.fa'
#Filter definitions
Filter_DenovoFilter
```

```
conditional_select      INFO    in      PINDEL  QS      >=      600
conditional_select      INFO    in      INDEL   QS      >=      200
select  Func    in      exonic;splicing
select  QS      >=      50
select  ExonicFunc      !=      synonymous SNV
select  Zyg     ==      het
remove  segdup  -       -
select  1000g   <=      0.01
select  snp_freq        <=      0.01
select  esp     <=      0.01
End_Filter
Filter_RecessiveFilter
conditional_select      INFO    in      PINDEL  QS      >=      600
conditional_select      INFO    in      INDEL   QS      >=      200
select  Func    in      exonic;splicing
select  QS      >=      50
select  ExonicFunc      !=      synonymous SNV
remove  segdup  -       -
select  1000g   <=      0.01
select  snp_freq        <=      0.01
select  esp     <=      0.01
End_Filter
Filter_XlinkedFilter
conditional_select      INFO    in      PINDEL  QS      >=      600
conditional_select      INFO    in      INDEL   QS      >=      200
select  Func    in      exonic;splicing
select  QS      >=      50
select  ExonicFunc      !=      synonymous SNV
remove  segdup  -       -
select  1000g   <=      0.01
select  snp_freq        <=      0.01
select  esp     <=      0.01
End_Filter
```

Eventually the program runs as:

```
> VariantMaster -c  fam01.cfg  -t fam01.tfam  -o My_output -v VCFfile -b BAM
```

and the logfile should read as:

```
Mon Feb 18 15:20:25 2013 - N. of Families: 1
Mon Feb 18 15:20:25 2013 - Family: fam01
Mon Feb 18 15:20:25 2013 - Loaded Individual F - father 0 - mother 0 - sex M - status unaffected
Mon Feb 18 15:20:25 2013 - Loaded Individual M - father 0 - mother 0 - sex F - status unaffected
Mon Feb 18 15:20:25 2013 - Loaded Individual S - father F - mother M - sex M - status affected
Mon Feb 18 15:20:25 2013 - Loaded Individual D - father F - mother M - sex F - status unaffected
Mon Feb 18 15:20:25 2013 - Loading hg19 reference from /data/hg19/hg19.unmasked.fa
```

VariantMaster generates the folder *My_output* with a report for each requested models. Specifically, these are extensions of the original CSV files where only variants in affected individuals satisfying the respective inheritance model are reported. Columns describing the presence of the mutations in the other family member have been added. For further statistical analysis, VariantMaster produces a summary file in which it reports the mutations found in all the affected individuals, ordered by model and by gene (see §4).

IMPORTANT

VCF files need to be annotated only once. When the folder *masterCSV* is created, use option *-s masterCSV* for further analyses.

Example: Tumor-to-Germline comparison from VCF files

In this example, Tumor-Germline pairs have been taken and sequenced from 10 unrelated individuals presenting the same type of tumor. Unnnotated VCF files are available in the folder *./somepath/VCF* with names *seq_T[G]01_ex[1-10].csv* and, accordingly, BAM files in the format *bamfile_ [T/G][1-10]_ex[1-10].bam*.
Therefore, the structure is (*e.g.tfam*)

```
ex1    T1     0      0      0      2
ex1    G1     T1     0      0      1
ex2    T2     0      0      0      2
ex2    G2     T2     0      0      1
…
ex10   T10    0      0      0      2
ex10   G10    T10    0      0      1
```

where tumor samples have been tag with 2 (-affected) and germ cells sample with 1 (-unaffected).

It is possible to first annotate the variants with Annovar with the switch (-A) and create a new folder (*masterCSV)* with the annotations:

```
>  VariantMaster -c  ex.cfg  -t fam01.tfam  -v VCF -A
```

The configuration file may be written according to entries in the header of newly created csv files in the masterCSV folder (- REMEMBER - it is a **TAB delimited file**):

```
#General variables
CARR_THR := 0.95
HOM_THR := 0.75
reference := '/data/hg19/hg19.unmasked.fa'
gene_reference := '/data/hg19/refGene.txt'
#Filter definitions
Filter_TumorFilter
conditional_select      INFO    in      PINDEL  QS      >=      300
conditional_select      INFO    in      INDEL   QS      >=      100
select  Func    in      exonic;splicing
select  QS      >=      10
select  ExonicFunc      !=      synonymous SNV
remove  segdup  -       -
End_Filter
```

The analysis is eventually started with:

```
>  VariantMaster -c  ex.cfg  -t fam01.tfam  -o test -s masterCSV -b BAM
```

Log should then read as:

```
Tue Feb 19 15:36:30 2013 - N. of Families: 10
Tue Feb 19 15:36:30 2013 - Family: ex1
Tue Feb 19 15:36:30 2013 - Loaded Individual T1 - father 0 - mother 0 - sex M - status affected
Tue Feb 19 15:36:30 2013 - Loaded Individual G1 - father T1 - mother 0 - status unaffected
…
Tue Feb 19 15:36:30 2013 - Loading hg19 reference from /data/hg19/hg19.unmasked.fa
Tue Feb 19 15:37:21 2013 - Applying the Normal-Tumor design
Tue Feb 19 15:39:42 2013 - Inspecting test_Tumor_GAIN_LOSS
…
Tue Feb 19 15:39:42 2013 - VariantMaster - Cleaning workspace
Tue Feb 19 15:39:42 2013 - VariantMaster - Processing complete
```

Example: Unrelated individuals

VariantMaster analyzes pools of unrelated individuals sharing a similar phenotype.

In this example, 5 unrelated individuals have been sequenced and respective CSV files are available in the folder *./somepath/CSV* with names Ind*[1-10].csv*. Accordingly, BAM files are in the format *bamfile_ ind[1-10].bam*.

Therefore, the samples are represented as (*unrel.tfam*)

```
U      ind1   0      0      0      2
U      ind2   0      0      1      2
U      ind3   0      0      0      2
U      ind4   0      0      1      2
U      ind5   0      0      0      2
```

where all individuals (3 males and 2 females) have been tagged as affected.

The configuration file (*unrel.cfg* - REMEMBER - it is a **TAB delimited file**) is

```
#General variables
CARR_THR := 0.95
HOM_THR := 0.75
reference := '/data/hg19/hg19.unmasked.fa'
#Filter definitions
Filter_UnrelatedFilter
conditional_select      INFO    in      PINDEL  QS      >=      300
conditional_select      INFO    in      INDEL   QS      >=      100
select  Func    in      exonic;splicing
select  QS      >=      10
select  ExonicFunc      !=      synonymous SNV
remove  segdup  -       -
End_Filter
```

4. Interpreting the output

VariantMaster yields comma separated (CSV) text files, one per each requested model, which are structured similarly to the input annotated CSV files. It is worth noting that it reports all the annotations by keeping the same header from CSV files.

Moreover it creates some additional columns:

- *Filter* - [Pass/Rejected] the variants fulfill the conditions dictated by the requested model
- *Owner* - who is the variant coming from
- *is_in_<someID>* - The presence or absence of the variant in the unaffected individual *someID* identified by conditional variant calling with the calculated coverage and zygosity

Chr	Start	…	Owner	Filter	is_in_Father	is_in_Mother
chrX	12313	…	Proband	Pass	NotFound|Father|aa:0.0|cov:87	Mother|het|aa:0.62|p:0.99|cov:323

Some variants are not reported in the output and that is dependent on the mode of inheritance. Usually these are variants that can be "obviously" eliminated during specific analyses because they are present in the CSV files of the unaffected individuals (e.g. search for denovo variants in family trios).

VariantMaster produces a final report with all the variants that survived the filtering and some additional statistics. For each selected model, VariantMaster calculates the number of times a specific gene has been mutated considering all the affected ids from all provided families (column *NV*). If the appropriate reference is provided[1], the ratio number of variants/exon length (KB) in the column (column *NV/ExLen*) is reported. Moreover VariantMaster calculates the occurrence of each single variant in the given pool (*Occurrence)* to help the identification of eventual recurrent mutations.

Specific fields in final report are:

- *Gene* - Gene name
- *NV* - Total Number of variants found in *Gene*
- *MutInd* - Number of Individuals with mutated *Gene*
- *NV/ExLen* - Total Number of variants found in *Gene* normalized per exon length
- *Occurrence* - Occurrence of a specific variant in the pool

[1] gene_reference := /path to UCSC gene reference has to be added in the configuration file

invitrogen
by life technologies

AccuPrime™ Taq DNA Polymerase High Fidelity

Cat. nos.	Size	Conc. 5 U/µL
12346-086	200 rxns	Store at –30°C to –10°C
12346-094	1000 rxns	

Pub. Part no. 12346.pps MAN0001081 **Rev. Date** 22 November 2011

Description

AccuPrime™ *Taq* DNA Polymerase High Fidelity provides qualified reagents for the high-fidelity amplification of nucleic acid templates by PCR. It includes an enzyme mixture composed of recombinant *Taq* DNA polymerase, *Pyrococcus* species *GB-D* polymerase, and Platinum® *Taq* Antibody.[1,2] This enzyme blend results in a nine-fold increase in fidelity over *Taq* alone and is effective over a wide range of target sizes—up to 20 kb with some optimization. Like regular *Taq*, AccuPrime™ *Taq* DNA Polymerase High Fidelity has a nontemplate-dependent terminal transferase activity that adds a single deoxyadenosine (A) to the 3' ends of PCR products.

Pyrococcus species *GB-D* polymerase is a proofreading enzyme that possesses a 3′ → 5′ exonuclease activity.[3] Mixture of this enzyme with *Taq* DNA polymerase increases fidelity and allows amplification of simple and complex DNA templates over a large range of target sizes. The Platinum® antibody complexes with *Taq* DNA polymerase and inhibits activity at room temperature. Activity is restored after the initial denaturation step at 94ºC, providing an automatic "hot start" PCR.[4,5]

The thermostable AccuPrime™ protein enhances specific primer-template hybridization during every cycle of PCR. Antibody/AccuPrime™ protein-mediated amplification dramatically improves specificity and provides the most robust PCR for multiplexing and suboptimal primer sets. Two AccuPrime™ PCR buffer mixtures are provided for amplifying specific types of templates. Buffer I is optimized for plasmids, cDNA, and λ DNA. Buffer II is optimized for genomic DNA.

Storage Buffer

20 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 1 mM DTT, 50% (v/v) glycerol, stabilizers.

Unit Definition

One unit incorporates 10 nmol of deoxyribonucleotide into DNA in 30 minutes at 74°C.

Product Use: For research use only.
Not intended for any animal or human therapeutic or diagnostic use.

Contents

Component	Kit Size	
	200 Rxns	1000 Rxns
AccuPrime™ *Taq* DNA Polymerase High Fidelity	40 µL	200 µL
10X AccuPrime™ PCR Buffer I	1 mL	4 × 1.25 mL
10X AccuPrime™ PCR Buffer II	1 mL	4 × 1.25 mL
Mg SO_4 (50 mM)	1 mL	1 mL

10X AccuPrime™ PCR Buffer I and II

Buffer I and II differ in their concentration of thermostable AccuPrime™ protein. Their components are:

600 mM Tris-SO_4 (pH 8.9), 180 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 2 mM dGTP, 2 mM dATP, 2 mM dTTP, 2 mM dCTP, thermostable AccuPrime™ protein, 10% glycerol.

The supplied PCR Buffer is a 10X concentrate and should be diluted for use.

Protocol Guidelines and Recommendations

- Critical parameters and troubleshooting information are documented in reference 1. Assemble PCR reactions in a DNA-free environment. We recommend use of clean dedicated automatic pipettors and aerosol resistant barrier tips. **Always** keep the control DNA and other templates to be amplified isolated from the other components.
- If PCR efficiency is not optimal, repeat the reaction with different primer concentrations from 100 to 500 nM (final concentration), in 100-nM increments.
- $MgSO_4$ is included in the 10X AccuPrime™ PCR Buffer at a final concentration of 2 mM, which is sufficient for most targets. For some targets, more Mg^{2+} may be required; use the 50-mM $MgSO_4$ provided in the kit to prepare a titration from 2 mM to 4 mM (final concentration) in 0.25-mM increments.
- For longer genomic DNA targets (>15 kb), we recommend using 2–2.5 U of AccuPrime™ *Taq* DNA Polymerase High Fidelity and increasing the extension time as specified (1 min per kb).
- Do not denature for more than 30 seconds if target is larger than 12 kb.

Protocol

Use the following protocol as a starting point and guideline when preparing your reactions. Adjust the reaction size as needed. Use Buffer I for plasmids, cDNA, and λ DNA and Buffer II for genomic DNA (up to 20 kb) as indicated.

1. Add the following components to a DNase/RNase-free, thin-walled PCR tube. For multiple reactions, prepare a master mix of common components to minimize reagent loss and enable accurate pipeting.

	Template Type	
Component	**Plasmids/cDNA/λ DNA**	**Genomic DNA**
10X AccuPrime™ PCR Buffer I	5 µl	—
10X AccuPrime™ PCR Buffer II	—	5 µl
Sense primer (10 µM)	1 µl	1 µl
Anti-sense primer (10 µM)	1 µl	1 µl
Template DNA	0.1 pg–20 ng	10 pg–500 ng
AccuPrime™ *Taq* High Fidelity	0.2 µl*	0.2 µl*
Autoclaved, distilled water	to 50 µl	to 50 µl

*0.2 µl = 1.0 unit, which is sufficient for amplifying most targets. In some cases (e.g., longer targets), more enzyme may be required, up to 2.5 units.

2. Cap the tube, tap gently to mix, and centrifuge briefly to collect the contents.

3. Place the tube in the thermal cycler and run the following program.

   Initial denaturation: 94ºC for 15 seconds to 2 minutes (do not denature for more than 30 seconds if target is larger than 12 kb)

   25–35 cycles of:
   Denature: 94ºC for 15–30 seconds
   Anneal: 52–64°C for 15–30 seconds
   Extend: 68ºC for 1 minute per kb of PCR product

   After cycling, maintain the reaction at 4ºC. Samples can be stored at -20ºC until use.

4. Analyze the amplification products by agarose gel electrophoresis .We recommend using E-Gel® 1.2% gels and TrackIt™ 100 bp or 1kb Plus DNA ladders (see **Additional Products** on page 4).

References

1. Innis, M.A., Myambo, K.B., Gelfand, D.H. and Brow, M.A.D. (1988) Proc. Natl. Acad. Sci. USA 85, 9436.
2. Barnes, W.M. (1994) *Proc. Natl. Acad. Sci. USA* 91, 2216.
3. Tindall, K.R. and Kunkel, T.A. (1988) *Biochemistry 27*, 6008.
4. Chou, Q., Russel, M., Birch, D., Raymond, J., Bloch, W. (1992) *Nucl. Acids Res., 20*, 1717.
5. Sharkey, D.J., Scalice, E.R., Christy, K.G., Atwood, S.M., Daiss, J.L. (1994) *BioTechnology, 12*, 506.
6. Westfall, B., Sitaraman, K., Solus, J., Hughes, J., Rashtchian, A. (1997) *Focus® 19*, 46.

Additional Products

Product	Amount	Catalog no.
10 mM dNTP Mix, PCR Grade	100 µL	18427-013
10 mM dNTP Mix, PCR Grade	1 mL	18427-088
E-Gel® 1.2% Starter Pak	6 gels plus PowerBase™	G6000-01
E-Gel® 1.2% 18-Pak	18 gels	G5018-01
TrackIt™ 100 bp DNA Ladder	100 applications	10488-058
TrackIt™ 1kb Plus DNA Ladder	100 applications	10488-085

Product Qualification and SDS

The Certificate of Analysis provides detailed quality control and product qualification information for each product. Certificates of Analysis are available on our website. Go to **www.lifetechnologies.com/support** and search for the Certificate of Analysis by product lot number, which is printed on the box.

Safety Data Sheets (SDSs) are available at **www.lifetechnologies.com/sds**.

Limited Use Label License No. 358: Research Use Only

The purchase of this product conveys to the purchaser the limited, non-transferable right to use the purchased amount of the product only to perform internal research for the sole benefit of the purchaser. No right to resell this product or any of its components is conveyed expressly, by implication, or by estoppel. This product is for internal research purposes only and is not for use in commercial applications of any kind, including, without limitation, quality control and commercial services such as reporting the results of purchaser's activities for a fee or other form of consideration. For information on obtaining additional rights, please contact outlicensing@lifetech.com or Out Licensing, Life Technologies, 5791 Van Allen Way, Carlsbad, California 92008

Limited Use Label License

This product is sold under licensing arrangements with Stratagene. The purchase price of this product includes limited, nontransferable rights to use only this amount of the product to practice the claims in said patents solely for activities of the purchaser within the field of research. Further information on purchasing licenses under the above patents may be obtained by contacting the Director of Business Development, Stratagene, 11011 North Torrey Pines Road, La Jolla, California 92037.

For support visit www.lifetechnologies.com/support or email techsupport@lifetech.com
www.lifetechnologies.com

life technologies™

SYBR® Safe DNA Gel Stain

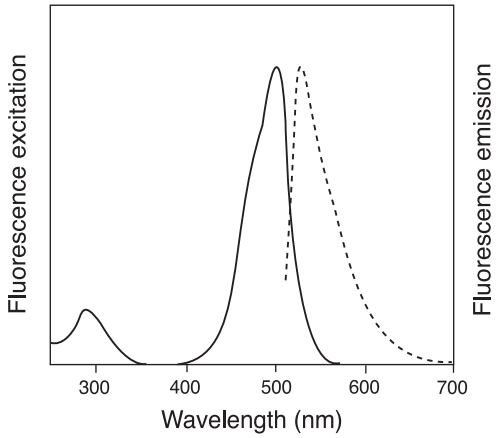
Table 1 Contents and storage

Material	Amount	Concentration	Storage	Stability
SYBR® Safe in TBE buffer	1 L* or 4 L†	0.5X, 45 mM Tris-borate, 1 mM EDTA, pH ~8.3	• 2–25°C • Protect from light	~6 months, when stored as directed
SYBR® Safe in TAE buffer	1 L* or 4 L†	1X, 40 mM Tris-acetate, 1 mM EDTA, pH ~8.3		
SYBR® Safe in DMSO	400 µL	10,000X		
SYBR® Safe DNA Gel Stain Starter Kit	1 L of SYBR® Safe gel stain, and one SYBR® Safe photographic filter	0.5X TBE		
Number of labelings: * Provides sufficient material to stain ~20 minigels. **†** Provides sufficient material to stain ~80 minigels. 4 L unit size is packaged in a cube-shaped container with a removable spigot for easy dispensing and storage.				
Approximate fluorescence excitation and emission maxima: 280, 502/530 nm, bound to DNA				

Introduction

SYBR® Safe DNA gel stain has been specifically developed for reduced mutagenicity, making it safer than ethidium bromide for staining DNA in agarose or acrylamide gels. SYBR® Safe stain comes either as a concentrate or as a ready-to-use solution that can be used like an ethidium bromide solution, and the detection sensitivity with SYBR® Safe stain is comparable to that obtained with ethidium bromide. DNA bands stained with SYBR® Safe DNA gel stain can be detected using a standard UV transilluminator, a visible-light transilluminator, or a laser-based scanner. The stain is also suitable for staining RNA in gels. Bound to nucleic acids, SYBR® Safe stain has fluorescence excitation maxima at 280 and 502 nm, and an emission maximum at 530 nm (Figure 1, page 2).

Figure 1 Normalized fluorescence excitation and emission spectra of SYBR® Safe DNA gel stain, determined in the presence of DNA



Before Starting

Storage, Handling, and Disposal

You may store the SYBR® Safe DNA gel stain at any temperature between 2 to 25°C. SYBR® Safe in DMSO freezes at low temperatures; therefore, the product must be completely thawed and mixed before using. Repeated freeze-thawing has minimal impact on product performance.

SYBR® Safe DNA gel stain showed no or very low mutagenic activity when tested by an independent, licensed testing laboratory, and this stain is not classified as hazardous waste under U.S. Federal regulations. The safety testing included 3 well-established mammalian cell–based tests (Table 2, page 3), a battery of well-established Ames-test bacterial strains (Figure 2, below), and extensive testing for environmental safety (Tables 3 and 4, page 3). Use care when using this reagent and dispose of the stain in compliance with all pertaining local regulations.

Figure 2 Summary of Ames test results for mutagenicity. Samples were pre-treated with a mammalian S9 fraction and then tested using the indicated Ames test strain. With strains TA97a, TA98, TA100, and TA102, a result of less than two-fold above background suggests that the compound is nonmutagenic; whereas, a result of greater than this value suggests that the compound is mutagenic. With strains TA1535, TA1537, and TA1538, a result of less than three-fold above background suggests that the compound is nonmutagenic; whereas, a result of greater than this value suggests that the compound is mutagenic. All tests were performed by Covance Laboratories, Inc., Vienna, VA, an independent testing laboratory.

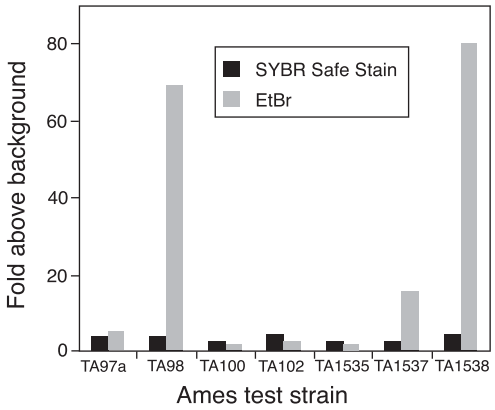


Table 2 Summary of mammalian cell–based tests for DNA genotoxicity

Test*	Cell Type	Test Result with S9 Activation†	Test Result without S9 Activation†
Transformation test [1]	Syrian hamster embryo (SHE) cells	Not applicable	Negative
Chromosomal aberration test [2]	Cultured human peripheral blood lymphocytes	Negative	Negative
Forward-mutation test [3,4]	L5178Y TK mouse lymphoma cells	Negative	Negative

***** All tests were performed by Covance Laboratories, Inc., Vienna, VA, an independent testing laboratory. **†** S9, a mammalian extract obtained from Aroclor™ 1254-induced rat liver.

1. Fundamental and Molecular Mechanisms of Mutagenesis 356:1 (1996); **2.** Evans, H.J., in *Chemical Mutagens, Principles and Methods for their Detection Vol 4,* A. Hollaender, Ed., Kluwer Academic/Plenum Publishers (1976) pp. 1–29; **3.** Mutation Res 72, 447 (1980); **4.** Mutation Res 59, 61 (1979).

Table 3 Summary of environmental safety test results

Test*	Method	Results
Aquatic toxicity	Fathead minnow CA Title 22 acute screening	Not classified as hazardous or toxic to aquatic life
Ignitability	EPA 1010	Not ignitable (>100°C)
Corrosivity	EPA 150.1	Not corrosive (pH = 8.25)
Corrosivity (by Corrositex)	DOT-E 10904	Category 2 noncorrosive
Reactivity	EPA 9010B/9030A	No reactivity detected

***** All tests were independently confirmed by AMEC Earth and Environmental San Diego Bioassay Laboratory, San Diego, CA.

Table 4 Summary of pollutant discharge test results

Test*	SYBR® Safe Stain†	0.5X TBE
pH (150.1)	8.45	8.48
Total cyanide (335.2)	None detected	None detected
Chemical oxygen demand (COD; 410.1)	7020	6840
Ammonia as nitrogen (350.1)	253	248
Total organic carbon (415.1)	2480	2360
Total phenolics (420.1)	None detected	None detected
Organochlorine pesticides and PCBs (608M)	None detected	None detected
Semi-volatile organic compounds (625)	None detected	None detected
Volatile organic compounds (624)	None detected	None detected
Metals (6010B, 7060A, 7421, 7470A, 7740, 7841)	None detected	None detected

***** Code of Federal Regulations Title 40, Part 136; **†** 1X SYBR® Safe stain (Lot X40023) in 0.5X TBE.

Experimental Protocols

Staining Nucleic Acids after Electrophoresis

1.1 Soak the gel in SYBR® Safe stain. If using SYBR® Safe gel stain concentrate, dilute 10,000X in TAE or TBE buffer (as appropriate) prior to use. Place the gel in a plastic container, such as a pipet-tip box lid or a household food-storage container. Do not use a glass container, as the dye in the staining solution may adsorb to the walls of the container, resulting in poor gel staining. Add sufficient SYBR® Safe DNA gel stain to cover the gel. A 50 mL volume is sufficient for staining most standard minigels. To stain larger gels, increase the volume of staining solution in proportion to the increased gel volume, and ensure that the gel is fully immersed during staining.

1.2 Incubate for 30 minutes. Protect the gel and staining solution from light by covering it with aluminum foil or by placing it in the dark. Continuously agitate the gel on an orbital shaker at 50 rpm. No destaining is required.

Precasting SYBR® Safe Stain in Agarose Gels

2.1 Prepare the agarose gel directly in SYBR® Safe DNA gel stain. SYBR® Safe stain is provided in buffer; simply substitute SYBR® Safe stain for the buffer when preparing the molten agarose. If using the 10,000X SYBR® Safe stain concentrate, dilute the concentrated stain 1:10,000 in agarose gel buffer (e.g., 1X TBE or 1X TAE) and add the buffer plus stain mixture to the powdered agarose. For example if you run TBE gels and require 30 mL of molten agarose for your tray, add 3 µL of 10,000X SYBR® Safe stain concentrate to 30 mL of 1X TBE, mix well, and add to the powdered agarose.

Note: The agarose/SYBR® Safe stain mixture may be heated in the microwave. As with precasting gels with ethidium bromide, the mobility of nucleic acid fragments in the gel may be somewhat slower when run in these gels compared to their mobility in the gel without stain.

2.2 Run the gel. Use a running buffer appropriate to the SYBR® Safe gel stain formulation. No post-staining or destaining is needed.

Viewing and Photographing the Gel

Stained gels can be viewed using a standard 300 nm transilluminator, a 254 nm epi- or transilluminator, or a blue-light transilluminator such as the Safe Imager™ 2.0 Blue-Light transilluminator from Molecular Probes (Cat. no. G6600). DNA stained with SYBR® Safe stain can also be visualized and analyzed using imaging systems equipped with an excitation source in the UV range or between 470–530 nm. Refer to Table 5 on page 5 to determine the optimal filter sets to use, or contact the instrument manufacturer for advice.

Note: If bands from the SYBR® Safe stained gel are to be excised and used in a ligation reaction, we recommend that the gel is illuminated using blue-light source (i.e., Safe Imager™ 2.0 Blue-Light transilluminator) and not a UV light source. In some instances, UV light sources in combination with SYBR® Safe stain can lead to reduced cloning efficiencies.

Stained gels can be photographed using Polaroid® 667 black-and-white print film and SYBR® Safe photographic filter (Cat. no. S37100). Molecular Probes' SYPRO® photographic filter (Cat. no. S6656) or a Kodak® Wratten #9 filter also work well. Using this film and one of these filters, SYBR® Safe DNA gel stain provides the same detection sensitivity as ethidium bromide and an appropriate filter. A standard ethidium bromide photographic filter should not be used with SYBR® Safe DNA gel stain. Gels stained with SYBR® Safe stain can also be imaged using a CCD camera or a laser-based scanner.

Table 5 Filter selection guide for use with SYBR® Safe stain

Instrument (Manufacturer)	Excitation Source	Emission Filter
AlphaImager (Alpha Innotech)	302 nm	SYB-500
AlphaImager HP (Alpha Innotech)	302 nm	SYB-500
AlphaDigiDoc RT (Alpha Innotech)	UV transilluminator	
Shroud, Camera Stand (Alpha Innotech)	UV transilluminator	SYB-100
DE500 or DE400 light cabinet 2.17″ diameter (Alpha Innotech)	UV transilluminator	SYB-500
DE500 or DE400 light cabinet 2″ diameter (Alpha Innotech)	UV transilluminator	SYB-400
VersaDoc Imaging Systems (Bio-Rad)	Broadband UV	520LP
Molecular Imager FX Systems (Bio-Rad)	488 nm	530 nm BP
Gel Doc Systems (Bio-Rad)	302 nm	520DF30 (#170-8074)
Typhoon 9400/9410 (GE Healthcare)	488 nm	520 BP 40
Typhoon 9200/9210/8600/8610 (GE Healthcare)	488 nm	526 SP
FluorImager (GE Healthcare)	488 nm	530 DF 30
Storm (GE Healthcare)	Blue (fluorescence mode)	
VDS-CL (GE Healthcare)	Transmission	UV Low
Ultracam/Gel Imager (Ultra-Lum)	UV	Yellow Filter (#990-0804-07)
Omega Systems (Ultra-Lum)	UV	520 nm
Polaroid Camera (Polaroid)	UV	SYBR® Safe Photographic Filter (S37100)
FOTO/Analyst Express/Investigator/Plus/ Luminary (FOTODYNE)	UV	Fluorescent Green (#60-2034)
FOTO/Analyst Minivisionary (FOTODYNE)	UV	Fluorescent Green (#62-4289)
FOTO/Analyst Apprentice (FOTODYNE)	UV	Fluorescent Green (#62-2535)
FOTO/Analyst Luminary (FOTODYNE)	UV	Fluorescent Green (#60-2056)
FCR-10 (Polaroid)	UV	#3-4218
FUJI FLA-3000 (FUJI Film)	473 nm	520LP
BioDocIt/AC1/EC3/BioSpectrum (UVP)	302 nm	SYBR® Green (#38-0219-01) or SYBR® Gold (#38-0221-01)
Gel Logic (Kodak)	UV	535 nm WB50
Syngene Instruments (Syngene)	UV	500–600 nm Shortpass filter

Product List **Current prices may be obtained at www.invitrogen.com or from our Customer Service Department.**

Cat no.	Product Name	Unit Size
G6600	Safe Imager™ 2.0 Blue-Light Transilluminator	each
S33100	SYBR® Safe DNA gel stain in 0.5X TBE	1 L
S33101	SYBR® Safe DNA gel stain in 0.5X TBE	4 L
S33102	SYBR® Safe DNA gel stain *10,000X concentrate in DMSO*	400 µL
S33103	SYBR® Safe DNA gel stain *10,000X concentrate in DMSO* *sample size*	20 µL
S33110	SYBR® Safe DNA Gel Stain Starter Kit *with 1 L of SYBR® Safe DNA gel stain in 0.5X TBE (S33100) and one photographic filter (S37100)*	1 kit
S33111	SYBR® Safe DNA gel stain in 1X TAE	1 L
S33112	SYBR® Safe DNA gel stain in 1X TAE	4 L
S37100	SYBR® Safe photographic filter	each
G5218-01	E-Gel® 1.2% with SYBR® Safe	18 gels
G6206-01	E-Gel® 1.2% with SYBR® Safe Starter Kit	1 kit
G5218-02	E-Gel® 2.0% with SYBR® Safe	18 gels
G6206-02	E-Gel® 2.0% with SYBR® Safe Starter Kit	1 kit

Contact Information

Corporate Headquarters
5791 Van Allen Way
Carlsbad, CA 92008
USA
Phone: +1 760 603 7200
Fax: +1 760 602 6500
Email: techsupport@lifetech.com

European Headquarters
Inchinnan Business Park
3 Fountain Drive
Paisley PA4 9RF
UK
Phone: +44 141 814 6100
Toll-Free Phone: 0800 269 210
Toll-Free Tech: 0800 838 380
Fax: +44 141 814 6260
Tech Fax: +44 141 814 6117
Email: euroinfo@invitrogen.com
Email Tech: eurotech@invitrogen.com

Japanese Headquarters
LOOP-X Bldg. 6F
3-9-15, Kaigan
Minato-ku, Tokyo 108-0022
Japan
Phone: +81 3 5730 6509
Fax: +81 3 5730 6519
Email: jpinfo@invitrogen.com

Additional international offices are listed at www.invitrogen.com

These high-quality reagents and materials must be used by, or directly under the supervision of, a technically qualified individual experienced in handling potentially hazardous chemicals. Read the Safety Data Sheet provided for each product; other regulatory considerations may apply.

Web Resources
Visit the Invitrogen website at www.invitrogen.com for:

- Technical resources, including manuals, vector maps and sequences, application notes, Meds, FAQs, formulations, citations, handbooks, etc.
- Complete technical support contact information
- Access to the Invitrogen Online Catalog
- Additional product information and special offers

SDS
Safety Data Sheets (SDSs) are available at www.invitrogen.com/sds.

Certificate of Analysis
The Certificate of Analysis provides detailed quality control and product qualification information for each product. Certificates of Analysis are available on our website. Go to www.invitrogen.com/support and search for the Certificate of Analysis by product lot number, which is printed on the product packaging (tube, pouch, or box).

Limited Warranty
Invitrogen (a part of Life Technologies Corporation) is committed to providing our customers with high-quality goods and services. Our goal is to ensure that every customer is 100% satisfied with our products and our service. If you should have any questions or concerns about an Invitrogen product or service, contact our Technical Support Representatives.
All Invitrogen products are warranted to perform according to specifications stated on the certificate of analysis. The Company will replace, free of charge, any product that does not meet those specifications. This warranty limits the Company's liability to only the price of the product. No warranty is granted for products beyond their listed expiration date. No warranty is applicable unless all product components are stored in accordance with instructions. The Company reserves the right to select the method(s) used to analyze a product unless the Company agrees to a specified method in writing prior to acceptance of the order.
Invitrogen makes every effort to ensure the accuracy of its publications, but realizes that the occasional typographical or other error is inevitable. Therefore the Company makes no warranty of any kind regarding the contents of any publications or documentation. If you discover an error in any of our publications, please report it to our Technical Support Representatives.
Life Technologies Corporation shall have no responsibility or liability for any special, incidental, indirect or consequential loss or damage whatsoever. The above limited warranty is sole and exclusive. No other warranty is made, whether expressed or implied, including any warranty of merchantability or fitness for a particular purpose.

Limited Use Label License No. 358: Research Use Only
The purchase of this product conveys to the purchaser the limited, non-transferable right to use the purchased amount of the product only to perform internal research for the sole benefit of the purchaser. No right to resell this product or any of its components is conveyed expressly, by implication, or by estoppel. This product is for internal research purposes only and is not for use in commercial services of any kind, including, without limitation, reporting the results of purchaser's activities for a fee or other form of consideration. For information on obtaining additional rights, please contact outlicensing@lifetech.com or Out Licensing, Life Technologies Corporation, 5791 Van Allen Way, Carlsbad, California 92008.

For research use only. Not intended for any animal or human therapeutic or diagnostic use.

Lonza Rockland, Inc.
www.lonza.com
biotechserv@lonza.com
Tech Service: 800-521-0390
Customer Service: 800-638-8174
Document # 18103-0807-8
Rockland, ME 04841 USA

NuSieve® 3-1 Agarose

Easy-to-handle gels for PCR product separation and blotting.

Introduction

NuSieve® 3:1 Agarose is a standard melting temperature agarose for resolving DNA fragments ≤1,000 bp. The high gel strength results in easy-to-handle gels, enhancing the convenience of gel processing and blotting. Performance testing of NuSieve® 3:1 Agarose ensures fine resolution of DNA fragments up to 1,000 bp.

Analytical Specifications

Gelling temperature (4%)	32.5°C-38.0°C
Melting temperature (4%)	≤90°C
Gel strength (4%)	≥1,400 g/cm²

Applications

- PCR† product separation and blotting
- Analytical electrophoresis of DNA and RNA fragments ≤1,000 bp

Suggested Agarose Concentrations

Size Range (Base Pairs)	Final Agarose Concentration (%) 1X TAE Buffer	1X TBE Buffer
500-1,000	3.0	2.0
100-500	4.0	3.0
10-100	6.0	5.0

Dye Mobility Table

Migration of double-stranded DNA in relation to Bromophenol Blue (BPB) and Xylene Cyanol (XC) in NuSieve® 3:1 Agarose Gels.

1X TAE Buffer		%	1X TBE Buffer	
XC	BPB	Agarose	XC	BPB
950	130	2.5	700	70
650	80	3.0	500	40
350	40	4.0	250	20
200	30	5.0	140	8
120	20	6.0	90	4

Precautions

Always wear eye protection when dissolving agarose and guard yourself and others against scalding solutions. Refer to Material Safety Data Sheet for additional safety and handling information.

Microwave Instructions for Agarose Preparation

1. Choose a beaker that is 2-4 times the volume of the solution.
2. Add **chilled** 1X or 0.5X electrophoresis buffer and a stir bar to the beaker.
3. Slowly sprinkle in the agarose powder while the solution is rapidly stirred.
4. **Remove the stir bar if not Teflon® coated.**
5. Soak the agarose in the buffer for 15 minutes before heating. This reduces the tendency of the agarose solution to foam during heating.
6. Weigh the beaker and solution before heating.
7. Cover the beaker with plastic wrap.
8. Pierce a small hole in the plastic wrap for ventilation. **For agarose concentrations >4%, the following additional steps will further help prevent the agarose solution from foaming during melting/dissolution:**
   A. Heat the beaker in the microwave oven on **Medium** power for 1 minute.
   B. Remove the solution from the microwave.
   C. Allow the solution to sit on the bench for 15 minutes.
9. Heat the beaker in the microwave oven on **Medium** power for 2 minutes.
10. Remove the beaker from the microwave oven. **Caution: Any microwaved solution may become superheated and foam over when agitated.**
11. **GENTLY** swirl the beaker to resuspend any settled powder and gel pieces.
12. Reheat the beaker on **HIGH** power until the solution comes to a boil.
13. **Hold at boiling point for 1 minute** or until all of the particles are dissolved.
14. Remove the beaker from the microwave oven.
15. **GENTLY** swirl the beaker to thoroughly mix the agarose solution.
16. After dissolution, add sufficient hot distilled water to obtain the initial weight.
17. Mix thoroughly.
18. Cool the solution to 50°C-60°C prior to casting.

Plate Instructions for Agarose Preparation

1. Choose a beaker that is 2-4 times the volume of the solution.
2. Add **chilled** electrophoresis buffer and a stir bar to the beaker.
3. Slowly sprinkle the agarose powder while the solution is rapidly stirred.
4. Weigh the beaker and solution before heating.
5. Cover the beaker with plastic wrap.
6. Pierce a small hole in the plastic wrap for ventilation.
7. Bring the solution to a boil while stirring.
8. Maintain gentle boiling until all the agarose is dissolved (approximately 10 minutes).
9. Add sufficient hot distilled water to obtain the initial weight.
10. Mix thoroughly.
11. Cool the solution to 50°C-60°C prior to casting.

Ordering Information:

Catalog No.	**Size**
50091	25 g
50090	125 g
50094	500 g

For more information on NuSieve® 3:1 Agarose, contact Technical Service at (800) 521-0390 or visit our website at **www.Lonza.com.**

Related Products:
DNA Ladders
DNA Markers
RNA Markers
GelStar® Nucleic Acid Gel Stain
NuSieve® GTG® Agarose
AccuGENE® TBE and TAE Buffers
The Sourcebook

For Laboratory Use.

†The PCR process may be covered by one or more third-party patents.

web SCIENTIFIC

A division of Thistle Scientific

User Instructions

*micro*CLEAN

Product description

*micro*CLEAN is an extremely rapid, simple, one tube (or plate) one and a half spin DNA clean-up reagent. It cleans up or concentrates any type of double stranded DNA. No mess 15 minute protocol. It requires No additional manipulations such as phenol extractions, ethanol precipitations, glass milk or spin columns. There is no limit to the size or volume of the target product DNA to be purified. After using *micro*CLEAN the recovered DNA can be resuspended in any desired volume of water or 10/1 TE buffer.

This product is supplied as a clear solution in aliquots of 1ml.

Procedure

- Just add an equal volume of *micro*CLEAN to DNA sample
- Mix by pipetting
- Leave at room temperature for 5 mins, then

For tubes

1) Spin at high speed (10 to 13K in microfuge) for 7 mins
2) Remove supernatant
3) Spin briefly again and remove all liquid (important)
4) Resuspend pellet in the appropriate volume of TE
5) Leave 5 mins to rehydrate DNA

For plates

1) Spin at 2000 to 4000 g for 40 mins
2) Place plate upside down onto tissue paper in the centrifuge holder
3) Pulse centrifuge to <40 g for 30 secs
4) Resuspend pellet in the appropriate volume of TE
5) Leave 5 mins to allow DNA to rehydrate

Clean DNA is now ready for further processing

Note: *micro*CLEAN-ed DNA is very clean and you should not see the pellet!

Precautions/Health and Safety

This product is for R&D use only. Please consult Material Safety Data Sheet for information on hazards and safe handling practices.

Storage and Stability

Store either at 4°C or at Room Temperature

Web Scientific, Radway Green Business Centre, Crewe, Cheshire, CW2 5PR. www.webscientific.co.uk
Tel: 01270 875 172 Fax: 01270 878 186 Email: info@webscientific.co.uk

Web Scientific

Radway Green Venture Park
Crewe, CW2 5PR

SAFETY DATA SHEET

DATE PRODUCED
July 2013

1. **CHEMICAL IDENTIFICATION**:
*micro*CLEAN

2. **COMPOSITION/INFORMATION ON INGREDIENTS**:
Proprietary Mix

3. **HAZARDS IDENTIFICATION**:
LABEL PRECAUTIONARY STATEMENTS
Irritant, irritating to eyes, respiratory system and skin.

4. **FIRST AID MEASURES**:
In case of contact with eyes, immediately flush eyes with copious amounts of water for at least 15 minutes.
In case of contact with skin, immediately wash skin with soap and copious amounts of water.
If inhaled, remove to fresh air and monitor breathing. If not breathing give artificial respiration. If breathing is difficult, give oxygen.
If swallowed, wash out mouth with water provided person is conscious. Call a physician.

5. **HANDLING AND STORAGE**:
This product should be handled only those persons who have been trained in laboratory techniques and that it is used in accordance with the principles of good laboratory practice. Storage conditions at +4°C.

6. **PHYSICAL AND CHEMICAL PROPERTIES**
Colourless slightly viscous liquid.

7. **REACTIVITY**
INCOMPATIBILITIES
Strong oxidising agents.
HAZARDOUS COMBUSTION OR DECOMPOSITION PRODUCTS
Toxic fumes of Carbon Monoxide, Carbon Dioxide, Hydrogen Chloride.
HAZARDOUS POLYMERISATION
Will not occur.

8. **TOXICOLOGICAL INFORMATION**:
Harmful if swallowed.

9. **TRANSPORT INFORMATION**:
Non-hazardous for road, sea and air freight.

10. **REGULATORY INFORMATION:**
EUROPEAN INFORMATION
Irritant
(R 36/37/38)
Irritating to eyes, respiratory system and skin.
(S26)
In case of contact with eyes, rinse immediately with copious amounts of water and seek medical advice.
(S36)
Wear suitable protective clothing

ExoSAP of PCRs prior to sequencing = élimine tous les produits du mix PCR qui pourraient empêcher le séquençage (primers...)

Make ExoSAP master mix:

EXO1/SAP Stock solution (per sample)

1.55µl dH_2O per PCR sample
0.15µl Exonuclease 1 per sample (10u/ul, ie 1.5 units)
0.30µl shrimp Alkaline Phosphatase (SAP) per sample (1u/ul, ie. 0.3units)

Add 2ul ExoSAP mix to PCR reaction (usually use 10ul of PCR + 2ul ExoSAP, allowing 2x5ul aliquots of clean PCR for for+rev sequencing reactions)

Run the "EXO-SAP" thermal cycler program, which is as follows:

- 37°C for 30 minutes
- 80°C for 15 minutes
- 10°C forever

2. Réaction de séquençage

Générer et imprimer une série.

Utiliser un des templates de séquençage pour le 3500 ou le 3500 XL disponibles dans : N:\ DIAGMOL.
Rentrer les indications nécessaires :

- dans la colonne « Sample Name » : en général no ADN_GENEexon123_primer
- dans la case qui contient le nom du template : en général, initiales et date

Sauver dans le répertoire approprié (format .xls), imprimer et pipeter les réactions de séquence selon cette liste.

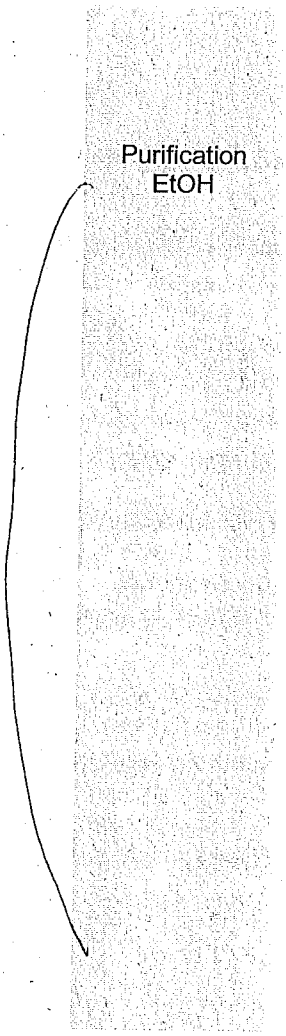
ATTENTION : ne pas mettre de caractères spéciaux dans ces fichiers (sont acceptés : chiffres, lettres sans accents, espaces, ., _, -).

Dans la Zone Postanalytique du Laboratoire

- Pipeter les réactions de séquence dans une plaque 96 puits ABI (pour la purification par le robot)
- Pipeter **1µl de produit PCR purifié par puits**
- Préparer les "Sequencing Mix" en fonction du nombre de tubes, garder sur glace.
  Par tube : 2 µl primer universel (M13F/R, N13hF/R (**1µM**)
  ou L/R (**10 ou 120ng/µl**), selon MON
  0.5µl Big Dye Terminator v3.1
  2 µl Big Dye Terminator v3.1 5x Sequencing Buffer
  µl cH_2O
- Pipeter 9µl de "Sequencing Mix" dans chaque puits
- Lancer le programme (se référer au document DIAG.4.1.FO.0003) :
  - 095 seq sur PCR ABI
  - « SEQUENCE » sur PCR BIOMETRA.

2µl PRIMERS m13
2µl ADN mio

Program #
DIAGMOL 12: SEQUENCE



S'assurer qu'il reste suffisamment d'eau pour les runs suivants, sinon remplir le bidon avec de l'eau déminéralisée (il est préférable qu'elle dégaze une nuit), et vider le bidon inférieur (déchets).

3.2 Purification EtOH

Protocole alternatif en cas de problème ou panne avec le robot Biomek.

a. Ajouter puits par puits (ne pas faire de mix)
   - 1µl de NaAc/EDTA (1/2vol 3M $C_2H_3NaO_2$ + 1/ vol 0.5M EDTA)
   - 40µl éthanol 95%
b. Inverser quelques fois la plaque pour mélanger (~50X)
c. Vortexer 1 min
d. Répéter cette étape une deuxième fois
e. Laisser la plaque 10 minutes à température ambiante à l'obscurité pour précipiter les produits d'extension
f. Centrifuger 30 minutes à 3000 rpm
g. Centrifuger la plaque à l'envers, 1 min à 1200 rpm, acc. :2, déc. :1
h. Ajouter à chaque puits 100µl de 70% éthanol
i. Inverser quelques fois la plaque pour mélanger
j. Centrifuger 10 minutes à 3000 rpm
k. Centrifuger la plaque à l'envers, 1 min à 1200 rpm, acc. :2, déc. :1
l. Laisser sécher à température ambiante 5-10 minutes, à l'obscurité
m. Resuspendre dans 12µl de Hi-Di formamide
n. Vortexer la plaque
o. Transférer les produits de séquençage sur une autre plaque, ajouter un septa
p. Centrifuger (pour enlever les éventuelle bulles ou gouttes restées sur les bords des puits

Utilizzo dell'Etanolo nel caso del protocollo Sanger

1_Il meglio sarebbe avere dell'etanolo preparato fresco in giornata

2_La concentrazione di etanolo è fondamentale

Step1 :95%

Step2 :70%

Se al primo step ho EtOH <90% non sic rea il pellet e perdo tutto (le sequenze)

Se al secondo step ho EtOH <60% il pellet si risospende e perdo tutto (le sequenze)

UNIVERSITÀ DEGLI STUDI DI FERRARA

Dottorati di ricerca

Il tuo indirizzo e-mail

chiara.olcese@gmail.com

Oggetto:

Dichiarazione di conformità della tesi di Dottorato

Io sottoscritto Dott. (Cognome e Nome)

Olcese Chiara

Nato a:

Treviso

Provincia:

Treviso

Il giorno:

27/11/1988

Avendo frequentato il Dottorato di Ricerca in:

Biochimica, Biologia Molecolare e Biotecnologie

Ciclo di Dottorato

28

Titolo della tesi:

Molecular basis of Primary Ciliary Dyskinesia

Titolo della tesi (traduzione):

Base Molecolare della Discinesia Ciliare Primaria

Tutore: Prof. (Cognome e Nome)

Bernardi Francesco

Settore Scientifico Disciplinare (S.S.D.)

BIO/18

Parole chiave della tesi (max 10):

Cilia, genes, Next-generation sequencing, whole-exome sequencing, PIH1D3 Cilia, geni, sequenziamento di nuova generazione, sequenziamento dell'intero esoma, PIH1D3

Consapevole, dichiara

CONSAPEVOLE: (1) del fatto che in caso di dichiarazioni mendaci, oltre alle sanzioni previste dal codice penale e dalle Leggi speciali per l'ipotesi di falsità in atti ed uso di atti falsi, decado fin dall'inizio e senza

fine di assicurarne la conservazione e la consultabilità da parte di terzi; (3) della procedura adottata dall'Università di Ferrara ove si richiede che la tesi sia consegnata dal dottorando in 2 copie di cui una in formato cartaceo e una in formato pdf non modificabile su idonei supporti (CD-ROM, DVD) secondo le istruzioni pubblicate sul sito: http://www.unife.it/studenti/dottorato alla voce ESAME FINALE – disposizioni e modulistica; (4) del fatto che l'Università, sulla base dei dati forniti, archivierà e renderà consultabile in rete il testo completo della tesi di dottorato di cui alla presente dichiarazione attraverso l'Archivio istituzionale ad accesso aperto "EPRINTS.unife.it" oltre che attraverso i Cataloghi delle Biblioteche Nazionali Centrali di Roma e Firenze; DICHIARO SOTTO LA MIA RESPONSABILITA': (1) che la copia della tesi depositata presso l'Università di Ferrara in formato cartaceo è del tutto identica a quella presentata in formato elettronico (CD-ROM, DVD), a quelle da inviare ai Commissari di esame finale e alla copia che produrrò in seduta d'esame finale. Di conseguenza va esclusa qualsiasi responsabilità dell'Ateneo stesso per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi; (2) di prendere atto che la tesi in formato cartaceo è l'unica alla quale farà riferimento l'Università per rilasciare, a mia richiesta, la dichiarazione di conformità di eventuali copie; (3) che il contenuto e l'organizzazione della tesi è opera originale da me realizzata e non compromette in alcun modo i diritti di terzi, ivi compresi quelli relativi alla sicurezza dei dati personali; che pertanto l'Università è in ogni caso esente da responsabilità di qualsivoglia natura civile, amministrativa o penale e sarà da me tenuta indenne da qualsiasi richiesta o rivendicazione da parte di terzi; (4) che la tesi di dottorato non è il risultato di attività rientranti nella normativa sulla proprietà industriale, non è stata prodotta nell'ambito di progetti finanziati da soggetti pubblici o privati con vincoli alla divulgazione dei risultati, non è oggetto di eventuali registrazioni di tipo brevettale o di tutela. PER ACCETTAZIONE DI QUANTO SOPRA RIPORTATO

Firma del dottorando

Ferrara, li _21/01/2016_____ (data) Firma del Dottorando _Chiara Olcese_

Chiara Olcese

Firma del Tutore

Visto: Il Tutore Si approva Firma del Tutore ____Francesco Bernardi_____