Year : 2013

Targeted sequence capture and Ultra High Throughput sequencing for gene discovery in inherited diseases

DI GIOIA SILVIO ALESSANDRO

UNIL | Université de Lausanne
Faculté de biologie
et de médicine

Département de Génetique Médicale

Targeted sequence capture and Ultra High Throughput sequencing for gene discovery in inherited diseases

Thèse de doctorat en science de la vie (PhD)

Présentée à la Faculté de biologie e médicine

de l'Université de Lausanne par

SILVIO ALESSANDRO DI GIOIA

Diplômé en Biotechnologie Mèdicale, Université de Rome "Sapienza", Italie

Jury

Prof. Luc PELLERIN, president

Dr. Carlo RIVOLTA, thesis director

Prof. Anneke DEN HOLLANDER, expert

Prof. Christian GRIMM, expert

Lausanne, 2013

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président	Monsieur	Prof. Luc **Pellerin**
Directeur de thèse	Monsieur	Dr Carlo **Rivolta**
Experts	Madame	Prof. Anneke **Den Hollander**
	Monsieur	Prof. Christian **Grimm**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Silvio Alessandro Di Gioia

Master en biotechnologie médicale de l' Université "Sapienza" Rome, Italie

intitulée

Targeted sequence capture and Ultra High Throughput sequencing for gene discovery in inherited diseases

Lausanne, le 12 septembre 2013

pour Le Doyen
de la Faculté de Biologie et de Médecine

Prof. Luc Pellerin

ACKNOWLEDGMENTS

It is not easy to enclose in few words four year-path of professional and personal enrichment. These were wonderful years: each person I met would deserve an entire book. For space reason I shall limit my acknowledgments, although I aware they do not constitute an exhaustive list of all the person I happened to meet and admire during my PhD years.

To them, first, this thesis is dedicated.

To Dr. Carlo Rivolta, whose great passion for genetics and his profound humanity have been a beacon and example, even in the most difficult moments of this journey.

To Prof. Ronald Roepman, which made me feel at home during my short visiting in Nijmegen, an experience I will never forget.

To the collaborators, with whom I have traveled this journey: Prof. Yvan Arsenijevic, and Dr. Corinne Kostic from Hospital Jules Gonin of Lausanne, Dr. Michaël Hofer, Dr. Marie Cochard and Dr. Andrea Superti-Furga, form CHUV in Lausanne.

To all the members of the laboratories of Medical Genetics, some like Adri and Goranka former colleagues, others actuals like Paola, Giulia, Luca, JJ, Hanna, Kostas, Pietro and Rosanna, Roman, Nathalie. Each single joke, hilarity and advice will be unforgettable memories.

To Katie Chapman, who had the burdensome task to edit my thesis.

To my friends of ITalaus, with whom I promoted the bright side of Italy and the friends of the theater association Pourquoi Pas, with whom I discovered another side of me.

To the old friends, like Valeria, Rosanna, Livio and Rosi, and all those with whom I spent wonderful moments while playing boardgames and having dinner. I hope I will enjoy such company for much longer.

To my sister: no matter how far you are, you will never be far from my heart.

To my parents: the thing I regret the most during these years was to be so far away from you. I have missed you, but I have always carried you with me over the years. I owe you everything in my life. Even this.

Thanks.

INDEX

RESUME

L'introduction des technologies de séquençage de nouvelle génération est en vue de révolutionner la médecine moderne. L'impact de ces nouveaux outils a déjà contribué à la découverte de nouveaux gènes et de voies cellulaires impliqués dans la pathologie de maladies génétiques rares ou communes. En revanche, l'énorme quantité de données générées par ces systèmes ainsi que la complexité des analyses bioinformatiques nécessaires, engendre un goulet d'étranglement pour résoudre les cas les plus difficiles. L'objectif de cette thèse a été d'identifier les causes génétiques de deux maladies héréditaires utilisant ces nouvelles techniques de séquençage, couplées à des technologies d'enrichissement de gènes. Dans ce cadre, nous avons développé notre propre méthode de travail (pipeline) pour l'alignement des fragments de séquence (reads). Suite à l'identification de gènes, nous avons réalisé une analyse fonctionnelle pour élucider leur rôle dans la maladie.

Dans un premier temps, nous avons étudié et identifié des mutations impliquées dans une forme récessive de la rétinite pigmentaire qui est à ce jour la dégénérescence rétinienne héréditaire la plus fréquente. En particulier, nous avons constaté que des mutations faux-sens dans le gène FAM161A étaient la cause de la rétinite pigmentaire préalablement associé avec le locus RP28. De plus, nous avons démontré que ce gène avait des fonctions au niveau du cil du photorécepteur, complétant le large spectre des cilliopathies rétiniennes héréditaires.

Dans un second temps, nous avons exploré la possibilité qu'un syndrome, relativement fréquent en pédiatrie de fièvre récurrente, appelé PFAPA (acronyme de fièvre périodique avec adénite stomatite, pharyngite et cervical aphteuse) puisse avoir une origine génétique. L'étiologie de cette maladie n'étant pas claire, nous avons tenté d'identifier le spectre génétique de patients PFAPA. Comme nous n'avons pas pu mettre à jour un nouveau gène unique muté et responsable de la maladie chez tous les individus dépistés, il semblerait qu'un modèle génétique plus complexe suggérant l'implication de plusieurs gènes dans la pathologie ait été identifié chez les patients touchés. Ces gènes seraient notamment impliqués dans des processus liés à l'inflammation ce qui élargirait l'impact de ces études à d'autres maladies auto-inflammatoires.

ABSTRACT

Next generation sequencing technologies are literally leading to a revolution in contemporary biology and medicine. The impact of these new tools has already been instrumental in the discovery of new genes and pathways involved in the etiology of rare or common inherited diseases. Conversely, the huge amount of data produced, and the complexity of the following bioinformatics analysis remain a bottleneck for solving the most challenging cases. The purpose of this thesis is to use these new sequencing techniques, in combination with DNA enrichment technologies, for the identification of the genetic cause of two hereditary diseases. In this framework we developed our own pipeline for reads mapping. Whenever possible, we also performed a deeper functional analysis to elucidate the role of this gene in the context of the disease.

In the first part of the thesis we investigated and identified the genetic injury involved in an autosomal recessive form of retinitis pigmentosa, the most common inherited retinal degeneration. In particular, we found that non-sense mutations in the gene *FAM161A* were the cause of retinitis pigmentosa linked to the previously-mapped locus RP28. Coherently with our workflow, we studied this novel disease gene, and we could demonstrate that it had functions at the level of the photoreceptor connecting cilium, expanding the spectrum of retinal inherited ciliopathies.

In the second part we explored the possibility that a relatively common pediatric syndrome with recurrent fever, but not with a clear etiology, PFAPA syndrome (acronym for Periodic Fever with Aphthous stomatitis, Pharyngitis and cervical Adenitis), could have a genetic background. We performed a whole exome genetic screening of 11 PFAPA affected patients to identify the gene(s) causing the disease. Although we did not identify any single gene mutated and shared among all screened individuals, a more complex genetic pattern was analyzed, implying oligogenic or complex inheritance of the trait. In particular the contribution of rare variants in inflammation related genes was investigated. This observation could also explain the pleiotropic spectrum of phenotypes found in several different autoinflammatory diseases.

PREFACE

LIVING THE "*NEXT GEN*"

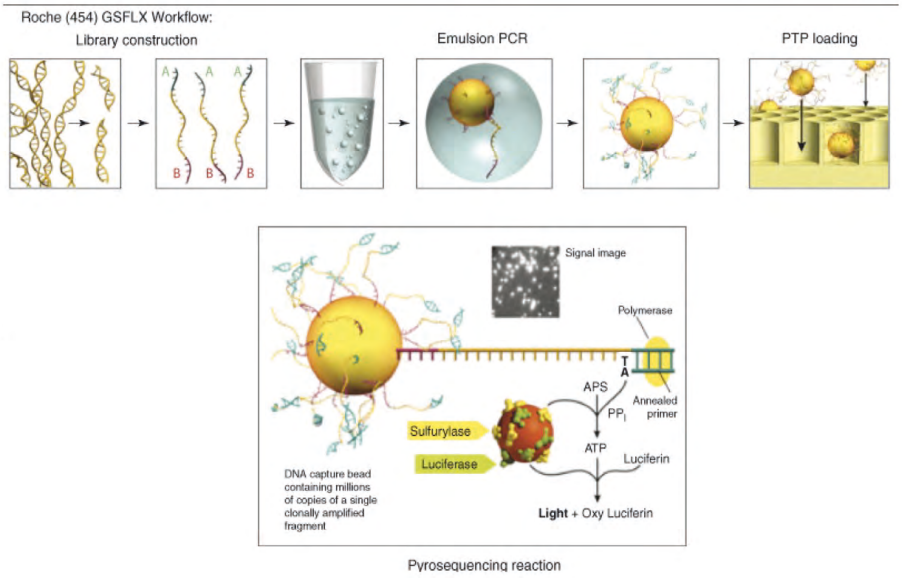
"Each generation goes further than the generation preceding it because it stands on the shoulders of that generation. You will have opportunities beyond anything we've ever known."
- Ronald Reagan, American president and actor, 1965

In 1977, the Nobel laureate Frederick Sanger presented in *Proceedings of the National Academy of Science of the United States of America* his work on DNA sequencing technique using radiolabeled "dideoxy" nucleotides that he used to sequence a long DNA fragment of a bacteriophage genome [1]. The subsequent development of this basic idea led to the modern "automated Sanger sequencing" with capillary electrophoresis, which was one if the driving factors behind the genetic revolution. This golden age of genetics culminated with the complete sequencing of the human genome, in 2001, released independently by two different groups [2, 3]. It took about 13 years to have a completed "draft" of the human genome, with the collaboration of different groups around the world. The knowledge acquired from this project helped researchers to identify the genetic cause of several different inherited diseases, and revealed the complexity and the high number of variants that are present in different individuals.

A second revolution in genetics began with the introduction of next-generation sequencing (NGS) technologies. These techniques were initially developed for the *de-novo* sequencing of unknown bacterial genomes, and thus almost completely primer free [4]. A few companies recently developed different machines for NGS, which mostly differ on the template preparation and the chemistry used for sequencing. The first NGS platform released on the market was the 454 Genome Sequences FLX, developed by Roche [5]. This system uses an emulsion-based clonal PCR for template preparation [6] and a modified pyrosequencing protocol for sequence reading (Fig. 1a). It is very efficient for *de-novo* sequencing and for the unrivaled length of reads (almost 600 bp) but on the other hand, it has some drawbacks, mostly connected to mistakes in the interpretation of the homopolymers-containing DNA sequences [7]. Another well-known platform for NGS has been produced by Life Technology and it is known as the SOLiD system [4]. The template is prepared by emulsion PCR, whereas the chemistry used for sequencing consists essentially in sequential steps of ligation by using dye-labeled couples of oligonucleotides (Fig. 1b) [8].

a.

ROCHE454



b.

Life Technology, SoLID

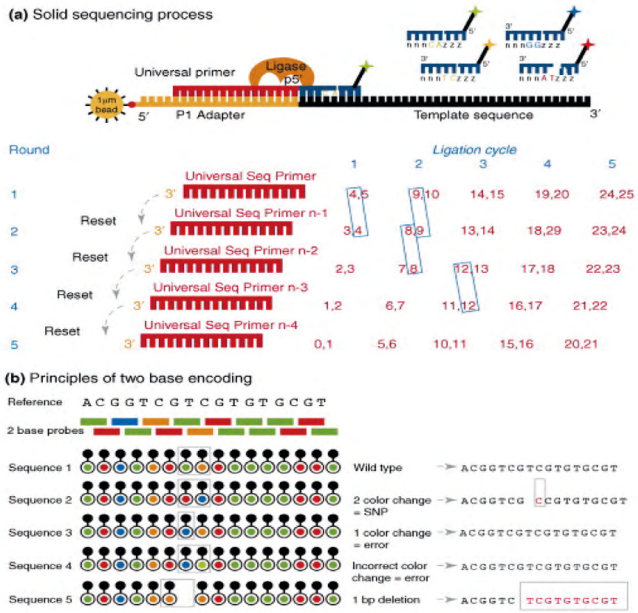


Figure 1. Schematic representation of two different NGS methods.
a) Different phases for Roche 454 library production and sequencing chemistry are depicted. The first two images show double stranded DNA sharing and adapter ligation. The third to the fifth images square show the special in situ PCR system called "emulsion PCR" that is performed on specific beads. These beads are then collected and loaded onto a silica membrane. The signal for a specific base is obtained by a pyrosequencing reaction. In this case the addition of a base causes the elongation of the single strand fragment and an emission of light. A CCD camera registers the emitted light and the software converts in sequence reads (adapted from Roche website). **b)** A depiction of the sequencing method for SOLiD technology. Again the template DNA is fragmented, denaturated, and ligated to a specific linker attached to beads. The sequencing reaction is mediated by special ligation reactions in which specific color marked dinucleotides are added to the growing sequence. The subsequent phase is called "primer reset", in which the previous primer is removed and another primer with a base less respect the previous one is annealed to the adapter. This is repeated six times. For this reason every single base for any read is called at least twice giving a better accuracy compared to other machines (Adapted from Life Technology website).

Among the advantages of this method are the high accuracy and the reduced error rates, making the sequencing process slower with respect to its competitors. The third platform, and the most popular of the three, is Solexa Genome Analyzer by Illumina Inc. It has the best ratio in output rate and running time, compensated by a lower multiplexing capability [9]. The process of template preparation is easy, and requires the shearing of DNA to generate shorter fragments that are then ligated to specific adapters. These fragments are then immobilized on a silicate surface, where *in-situ* "bridge" PCR [10] is performed (Fig. 2). This reaction generates clusters of DNA fragments on the silicate surface that are then denaturated as single stranded DNA. After the addition of fluorescent dye terminators there is the elongation step of one single base, with a modality similar to classical Sanger reaction. A CCD camera reads the emitted fluorescence and saves a picture. This synthesis/terminator step is repeated for several cycles. Software then analyzes images and converts them into a short nucleotide sequence, the so-called "read".

These techniques have been intensively used to investigate single nucleotide polymorphisms (SNPs) diversity among individuals [11] or in cancer genetics, to identify genetic injuries linked to a particular type of cancer [12, 13]. Indeed, NGS re-sequencing has had a major impact on human genetics, as demonstrated by the large number of novel genes recently associated to rare inherited diseases [14]. The combination of targeted enrichment techniques, such as whole-exome capturing [15] or custom targeted sequencing with NGS, allowed researchers to investigate only exons or regions of interest in a particular family, making the identification of the causative mutation easier and faster.

In this dissertation I have made extensive use of NGS technology to identify the causative mutations for two different and non-related diseases. The first, and major, part of this thesis focuses on the discovery and characterization of a new gene involved in retinitis pigmentosa. The second part attempts to identify the genetic basis of a common pediatric recurrent fever known as *P*eriodic *F*ever with *A*phthous stomatis, *P*haryngitis and cervical *A*denitis *(*PFAPA).

NEXT GENERATION SEQUENCING: BIOINFORMATICS TOOLS

"Computers are useless. They can only give you answers.", Pablo Picasso, Spanish painter, 1967

NGS is the current golden standard for mutational screening in both rare and common diseases. The combination of exome capturing with NGS has solved several difficult cases of rare syndromes for which few families were available [16]. We have discussed the different machines and techniques that can perform NGS, but we have not discussed the important step between sequencing and variant identification: bioinformatics analysis. This step is not a simple technical issue, but a critical part of the sequencing process. In fact, due to the large amount of raw data generated by the sequencing techniques, a correct alignment to the reference sequence and a perfect calling of all variants is the difference between finding and not finding the causative change. Therefore it is not surprising that a large portion of recent bioinformatics research has focused its attention on improving or generating ever better or more sophisticated systems to respond these above-mentioned tasks [17-19]. A general pipeline for NGS has consecutive steps: 1) base-calling with quality score calculation; 2) read-mapping; 3) quality recalibration; 4) genotyping and SNPs calling. These steps are now described separately with the most common software used for these analyses.

The base-calling step is the first bioinformatic analysis performed on the original output generated by the sequencing machine. The software analyzes the output and converts it into a sequence of DNA. Moreover, it also gives an estimation of the quality of the "calls", a technical term that indicate the base "called" by the machine in a given position by analysis of the output. This is called per-base quality score calculation. Usually this value is indicative of the quality of a single base without considering the surrounding ones. In general this step is performed directly by in-house software produced by the sequencing machine company, although several different specific methods have been recently developed to improve the quality of the calls by 30% [20].

The second step for a correct call is the alignment. In this step the large number of sequences are aligned to a reference genome. The quality of the alignment is crucial for the identification of the causative variant, since bad alignment or skipping of a real insertion/deletion (indels) of some bases could generate false positive or negative results. Another important element to take in account for a good alignment is the time necessary for the computer to accomplish this task. A good aligner should be also easily accessible also to not expert bioinformaticians, in terms of parameters and functionality set up. Up to now two different systems have been used for alignment algorithm. One is based on the "Burrows-Wheeler transform" (BTW) algorithm, and it is used by software like Bowtie [21], BWA [22], and SOAP2 [23]. This algorithm is based on data compression and permutation to simplify the large amount of data, making the

alignment fast, and without requiring huge computer memory. Although the permutation system is very good in aligning repetitive sequences is not so accurate for indels and gaps, and is less sensitive compared to the "hashing"-based algorithms. This second system is based on a "hash table", and is 'state of the art' in sequence aligners. Put simply, the software creates hash tables of the reference genome and reorders it in an intuitive way. Reads are then used to search these tables and to align them in the best way. It requires more memory and time, but this strategy is far more accurate than BTW-based algorithms. Moreover, the newly released software using this system (e.g. MAQ [18] and Novoalign) are much more implemented and fast compared to older softwares.

Quality score recalibration is the third step and it is necessary to have a correct variant call. In fact the per-base quality score is not overly precise and it does not take into account the environment where the called base is present. To resolve this issue, the quality score is recalibrated according to the reference information and the list of known SNPs. Some alignment software, like Novoalign, perform this step directly during the alignment, whereas for other systems specific tools are present such as for the Genome Analysis Tool Kit (GATK) [24] or in SOAP2 package.

Genotyping and variant calling follow the previously described procedures. These steps are composed of two different parts, although the same tool usually performs them at the same time. Firstly, a correct genotype must be given to each base of the alignment such as homozygous (homo) or heterozygous (het). Classical systems usually simply count the number of aligned reads having the mismatch base compared to the reference to determine the genotype (if around 50% reads have it then the change is called as het, if around 100% it is called as homo), but the new recent algorithms take into account other different options to make less mistakes. Recent systems use a probabilistic model for calculating the genotype based on quality score. Moreover, they also consider the percentage of errors in a specific region, and they correct the quality score according to this observation [18, 23]. For example, the use of a SNPs database is important for calculating the minor allele frequency, and for recalibrating the genotype calling according the observed ratio of het/homo. Other systems take in account known linkage disequilibrium data for any single variant to calculate the probabilistic value. Also the use of genotype information based on large sequencing experiments could help in this sense. The variant calling is then based on the genotypes and it gives the position and the type of change considering different filtering variants. This step generates a VCF

file (Variant Call Format, the standard format of variant file) that has to be deciphered and annotated to infer essential information and to find the correct gene.

This bioinformatics analysis however is the easiest part in the identification of a mutation. The challenging part starts just after this, and requires smartness, obstinacy and a dust of good luck.

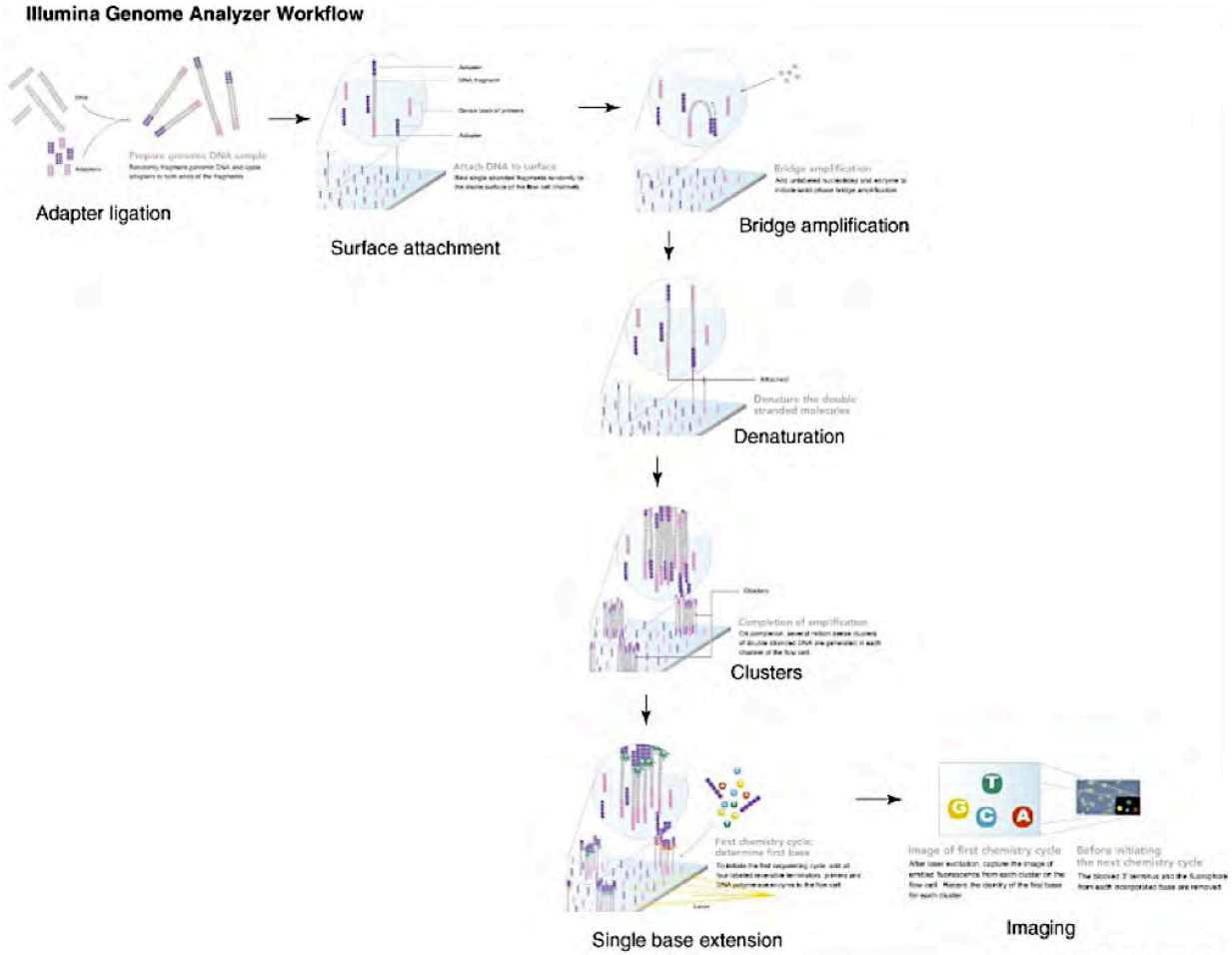


Figure 2. Schematic representation of ILLUMINA sequencing workflow.
Depiction of the sequencing method for Illumina GA. Template DNA is sonicated to obtain short fragments, where specific adapters are ligated. These fragments are then attached to a silica surface where they can hybridize with complementary sequences. Then an *in situ* reaction called "Bridge PCR" is performed to obtain a double stranded molecule. After denaturation two copies of the same fragment will be near to each other on the silica surface. The next step is a combination of different amplification steps important for the formation of specific clusters of single strand DNA to allow clonal enrichment. These clusters are the starting point for the sequencing reaction using dye terminators. Starting from a universal primer, single fluorescent dye terminators are added to the machine and the emitted fluorescence is read by a CCD camera. For each cycle a new dye is added and the fluorescent image is converted into a sequence. (adapted from Mardis et al., 2008)

I

IDENTIFICATION AND CHARACTERIZATION OF A NEW GENE INVOLVED IN AUTOSOMAL RECESSIVE RETINITIS PIGMENTOSA

I.1 INTRODUCTION

1.1 THE EYE

"*The tongue can hide the truth, never the eyes." Mikhail Bulgakov, Russian writer 1891*

The organ involved in vision is the eye. Despite its small size, the eye contains some of the most complex machinery present in the body. Anatomically, it is made up of three different layers, each having a specific function (Fig. 3a)[25]:

1. The fibrous tunic, which is comprised of the cornea and the sclera. It is composed of dense connective tissue. It has both a protective function and a role in maintaining the shape of the eye.
2. The vascular tunic, which is comprised of the iris, the ciliary bodies and the choroid. The iris defines the size of the pupil and the ciliary bodies are involved in controlling lens shape. The choroid has a vascular function and provides oxygen or nutrients to the outer retina.
3. The nervous tunic, which corresponds to the inner sensory tissue including the retina.

Adjacent to the nervous tunic there is a large cavity, filled with a gel-like substance called vitreous. The vitreous is composed of 99% water, with the other 1% consisting of collagen, hyaluronic acid and other extracellular matrix proteins [26]. The main function of the vitreous is to keep the shape and structure of the eye that is required to have the correct focus point.

Protected by the sclera and blood supplied by the choroid there is the retina, which is the sensitive and most important tissue of the eye and belongs to the central nervous system. During development, the retina (about 0.2 mm thick) is derived from an out-pocketing of the neural tube, called the optic vesicle, which leads to the formation of the optic cup. The external part of the optic cup forms the retinal pigment epithelium (RPE), whereas the inner part leads to the formation of the retina (Fig. 3b). The RPE is an epithelial tissue directly connected to the choroid and has three main functions: Firstly, the presence of melanin pigment in RPE cells is necessary for

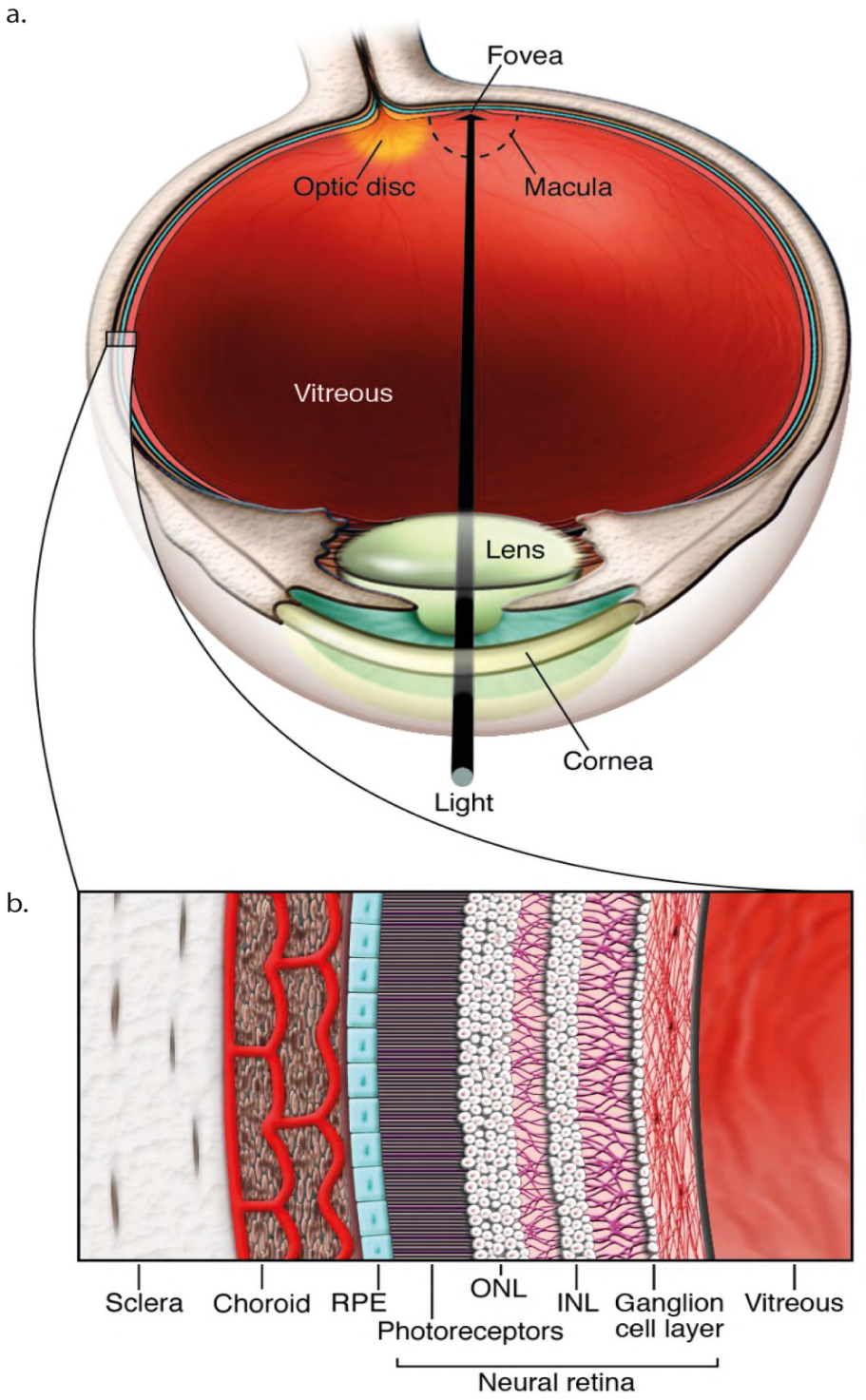


Figure 3. The eye and the retina
a) A schematic view of the eye. Light enters through the pupil that is protected by a crystalline tissue called cornea. The lens concentrates light-waves at the level of macula, a highly pigmented part of the eye. The center of the macula is called the fovea and in humans it contains the highest density of cones and corresponds to the point of best visual acuity. The optic disc is the blind point of the eye. **b)** Longitudinal section of the eye. It is possible to distinguish the peculiar laminar structure of the retinal tissue. ONL, outer nuclear layer; INL, inner nuclear layer; RPE, retinal pigmented ephitelium (Adapted from Den Hollander et al., 2010).

the absorption of light passing over the retina, preventing it from being reflected off of the back of the eye [27]; secondly, it has an important trophic function, since it is in direct contact with the choroid and provides the necessary molecules for photoreceptor feeding and renewal; finally, it has a direct role in the regeneration of the photoreceptor outer segment via the phagocytosis of the apical part of photoreceptor cells [28]. Conversely, the retina is an ordered laminated structure composed of neurons with different specialization and shapes (Fig. 3b). The cell bodies and the nuclei of these neurons are well separated, forming different layers. The most external part of the retina, in direct contact with the vitreous, is composed of the nuclei of ganglion cells. These cells project their dendrites into the brain forming the optic nerve. Where these dendrites form the optic nerve there is the so-called "blind spot" of the eye (the optic disc), where no single photosensitive cell is present [29]. Ganglion cells touch dendrites from intermediate neurons at the level of the inner plexiform layer (IPL). Here both the soma of amacrine cells and the dendrites of bipolar cells, which have their nuclei in the inner nuclear layer (INL), can be found. Whereas bipolar cells have a direct function in transmitting the signal from the photoreceptor to ganglion cells, amacrine cells work more as message collectors and integrate the signal from rod-connected bipolar cells. The INL also contains the cell bodies of horizontal cells, which are involved in the modulation of the signals from photoreceptors under different light conditions. Bipolar and horizontal cells connect to photoreceptor dendrites in the outer plexiform layer (OPL). Two different types of cells make up the photoreceptor layer; rods and cones, which are directly connected with the RPE layer. Photoreceptors are the photosensitive cells in the retina and their apical part is where phototransduction reactions occur.

1.2 PHOTORECEPTOR CELLS AND THE VISION

"What we call the beginning is often the end. And to make an end is to make a beginning. The end is where we start from." - T. S. Eliot, English poet, 1888

Photoreceptors are particular neuronal cells [30] with a specific function: to capture light and transmit the signal to neurons. Each photoreceptor contains photopigments, composed of a protein moiety called opsin, plus a small light-absorbing compound called chromophore. Although photoreceptor cells present species-specific features at the structural level, the photoreaction cascade is well conserved, excluding the photopigment structure [31]. Originally photoreceptors evolved from a single sensitive cell, but during evolution they diverged into two different branches, distinguishable by different modifications of the apical membrane in order to increase

light absorption. These subtypes are the ciliary photoreceptors, derived from a modified primary cilium, or the rhabdomeric photoreceptors, from a modification of the surface in a microvillar way to form a rhabdome (Fig. 4)[32]. Rhabdomeric photoreceptors are common in arthropods (like drosophila) and mollusks, whereas ciliary photoreceptors are found in fish and vertebrates as well as in mammals.

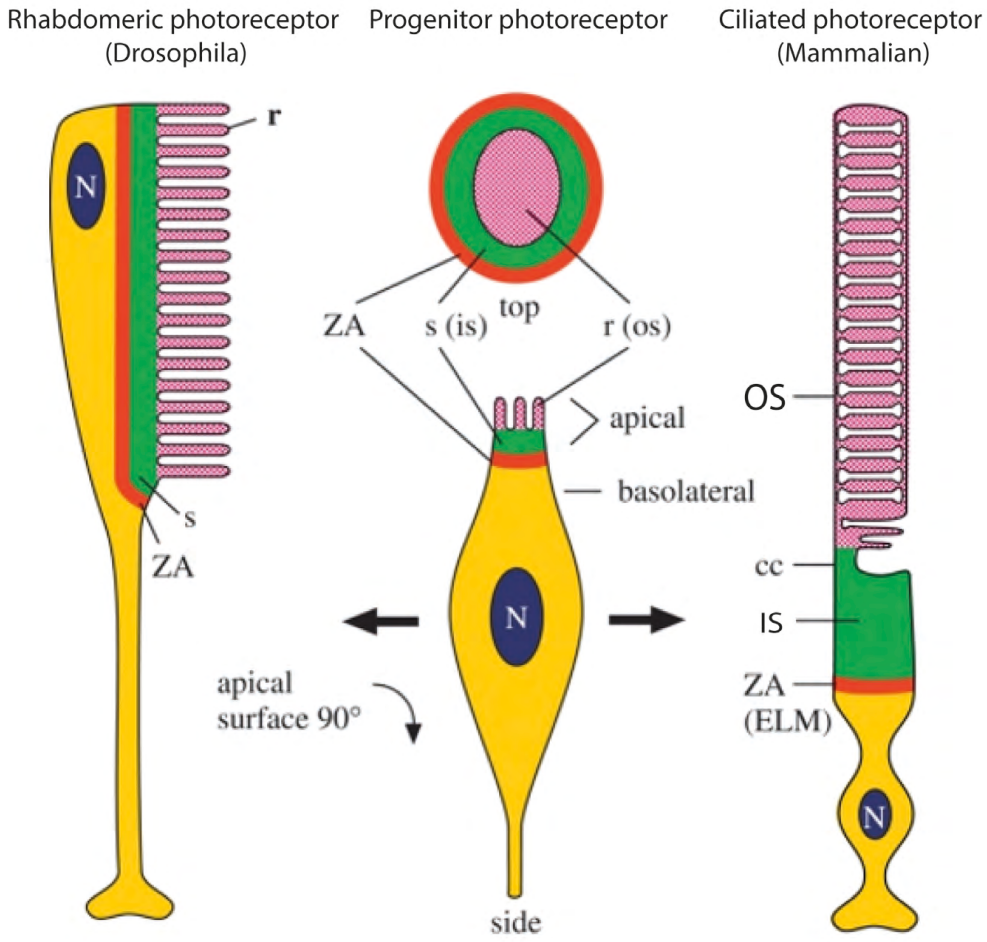


Figure 4. Photoreceptor cells originate from a common ancestor photosensitive cell.
The figure shows the differences and similarities between ciliated and rhabdomeric photoreceptor. Both originate from a photosensitive progenitor. Some elements of the cell are conserved in both types of photoreceptors. The IS of ciliated photoreceptors correspond to the stalk of rhabdomeric ones (green area). The OS correspond to the rhabdome (indicated in pink), whereas the red part corresponds to the zonula adherens (or external limiting membrane in mammalian photoreceptors) that delimitate the apical membrane with the basolateral membrane. It is possible to observe that the peculiar polarization of the sensitive cell is maintained during evolution.
r, rhabdome; s, fly stalk; ZA, zonula adherens, N, nucleus; OS, outer segment; cc, connecting cilium; IS, inner segment; ELM, external limiting membrane. (Adapted from Lamb et al., 2009).

As previously mentioned, vertebrates have two kinds of ciliary photoreceptors called rods and cones, respectively (Fig. 5), because of their resemblance to such geometric shapes. In brief, cones are involved in daylight and color vision, whereas rods mediate night or dim-light vision. This functional difference is achieved by using distinct opsin molecules that are capable of absorbing various wavelengths of light. Rods contain a specific opsin, called rhodopsin. This molecule can absorb a broad range of wavelengths, and has a great sensitivity, able to detect one single photon. Rhodopsin has an absorbance peak at 510 nm. Conversely cones contain a different opsin, less sensitive than rhodopsin in photon capturing, but much more specific for discrete wavelengths. In humans there are three different opsins, called S-, M- and L-opsin, with peaks of 430, 530 and 570 nm, respectively, that are responsible for trichromic vision (red, green and blue -RGB). In the cones of rodents, only two diverse opsins are present, S- and LM-opsin, giving them dichromic vision. In humans the majority of cones are localized to the fovea, which is the area with the best acuity of the eye. In the fovea, each cone is connected to a single bipolar cell and to a single ganglion cell, giving great benefits to signal transmission quality and acuity. Conversely, the periphery of the eye is enriched in rods and each ganglion or bipolar cell is connected to more than one photoreceptor. Therefore rods are more abundant than cones (they outnumber them by more than 20 fold) and are distributed all over the retina surface with the exception of in the central fovea region (fovea centralis).

The most important pathway for photoreceptor biology, and for which the entire cell is specialized, is the phototrasduction cascade. It begins in the outer segment of the photoreceptor and involves opsins. Although this pathway is similar between rods and cones, small differences are present [31]. In rods, when a single photon enters the cell, it causes the photoisomerization of the chromophore, *11-cis-retinal*, that becomes *all-trans-retinal* (Fig. 6b). This chemical shift induces a conformational modification in rhodopsin tridimensional structure that leads to the activation of a specific G-protein, called transducin, by catalyzing the exchange from GDP to GTP. This stimulates the activation of a membrane-bound phospodiesterase (PDE), which is able to break down cGMP (specifically, it is the γ subunit of the PDE that performs hydrolysis). In darkness there is a high concentration of cGMP in the cells and this activates the cGMP-gated cation channels with consequent depolarization of the cell and the release of the neurotransmitter glutamate. This is called the “dark current”. In the presence of light there is a hyperpolarization of the cells with the subsequent blocking or reduction of glutamate transmission to bipolar cells. The activation of rhodopsin is followed by a

recovery phase in which photopigment is regenerated and the transducin pathway deactivated. The recovery is started by the decay of activated rhodopsin into an inactive state via phosphorylation, mediated by the Rhodopsin Kinase (encoded by the gene GRK1). This event induces the binding of the protein arrestin to rhodopsin that arrests its activation. Transducin has an intrinsic GTPase activity and autonomously converts itself from the active GTP to the inactive GDP state. This causes the inactivation of PDE and the regeneration of normal concentrations of cGMP by retinal guanylyl-cyclase, restoring the dark current. A very important step for the regeneration of photoreceptor light-sensitivity is the "visual cycle" that happens between the photoreceptors (both rods and cones) and the RPE cells, or between cones and Müller cells [33] (Fig. 6a). A Retinol dehydrogenase converts the *all-trans-retinal* into *all-trans-retinol*, which subsequently moves from the photoreceptors to the RPE or Müller cells via the *inter-photoreceptor retinoid-binding protein* (IRBP). In these cells, the retinol is sequestered by *cellular retinol-binding protein* (CRBP) and is converted in *11-cis-retinal* by different chemical reactions (in the RPE these are catalyzed by some important enzymes such as *lecithin: retinol acyl-transferase* LRAT, and the *isomer-hydrolase* RPE65). The *11-cis-*retinal is then transported back to the photoreceptor cell where it can restore rhodopsin[34].

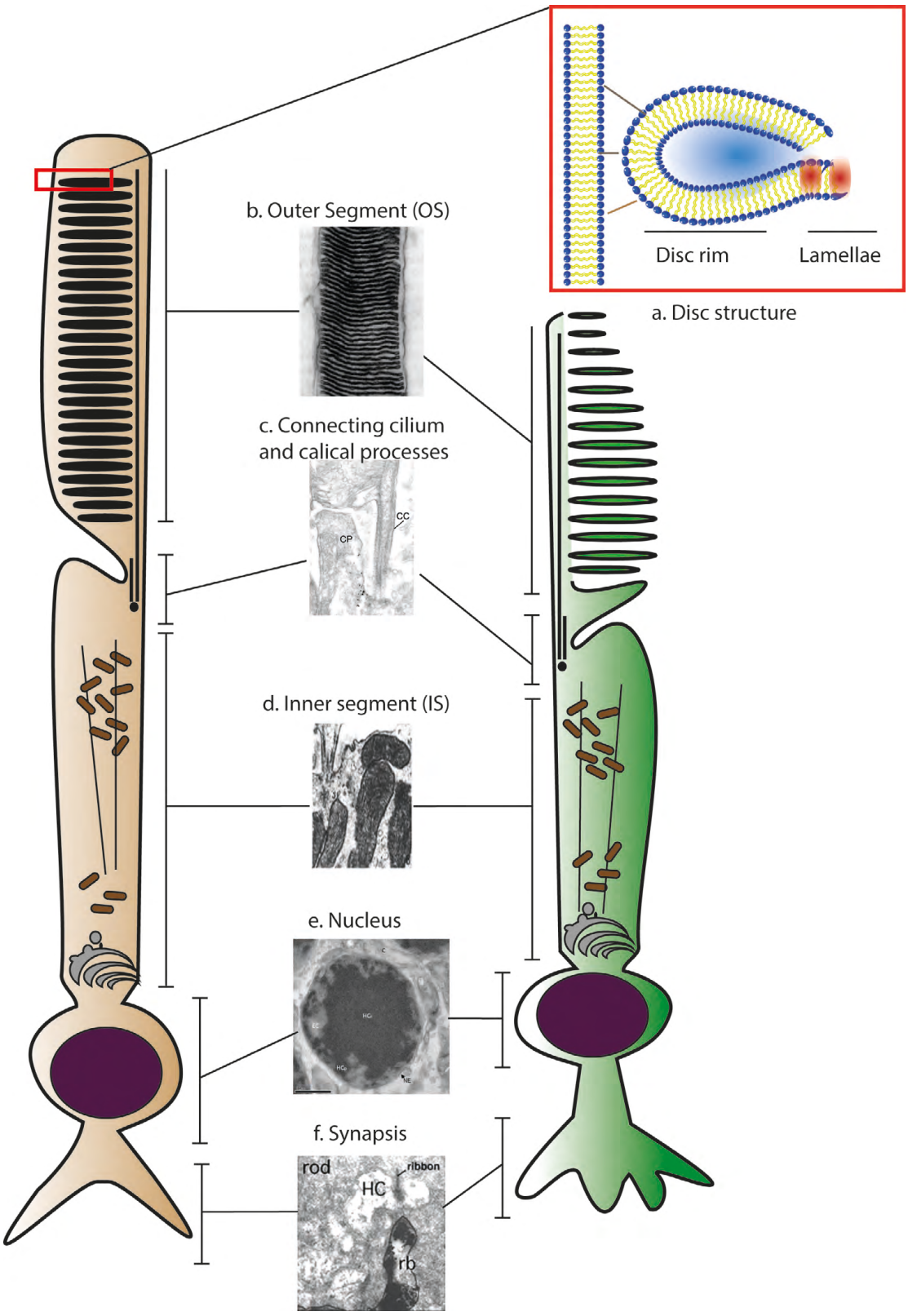


Figure 5. Photoreceptor cells.
The structure of a cone and rod cell is depicted. **a)** Magnification of the disc rim. Rhodopsin molecules are indicated in red. **b)** EM image of a disc stack. **c)** Connecting cilium and calical process (CC and CP) (from www.phys.ufl.edu). **d)** Inner segment of a rod where is possible to see large mitochondria (from Steinberg et al., 1980). **e)** The nucleus (from Kizilyaprack et al, 2010). **f)** Synaptic ribbon of a rod (Missotten et al., 1965).

1.3 STRUCTURE AND BIOLOGY OF PHOTORECEPTORS

"The universe is built on a plan, the profound symmetry of which is somehow present in the inner structure of our intellect." - Paul Valeri, French poet, 1871.

The evolution of photoreceptors is strictly connected with their function [32]. In particular, the need to increase surface area for light capturing in ciliary photoreceptors was resolved by the formation of an apical compartment packed with several membrane discs, known as the outer segment (OS). There are four distinct parts in photoreceptor cells: the already-mentioned OS, the inner segment (IS), the nucleus and the synaptic terminus (Fig. 3abcd). In rods, the OS has a cylindrical shape and it is composed of more than 2000 highly organized, flattened, membranous discs tightly-packed and surrounded by their own plasma membrane (Fig 5b). These discs are perpendicular to the entering light, maximizing photon capture [35]. They are made up of a central thinner part called the "lamellae" and the outer part of the disc, the rim, which are very different in terms of molecular composition (Fig. 5a). The disc rim is connected to the plasma membrane of the OS through specific filaments [36]. New discs are formed at the base of the OS, close to the ciliary compartment and they move progressively towards the apical part where they mature [37]. The lamellae are filled with rhodopsin molecules, increasing the probability of capturing light, whereas no rhodopsin is present in the disc rims [38].

The IS can be considered as the industrial area of the photoreceptor. This is the portion of the cell where the metabolism, biosynthesis and all other cellular pathways occur. The IS is extremely organized in terms of internal structure. The apical part is protruding forming a microvilli structure that surrounds the base of the OS. These structures resemble a calyx of a flower, and for this reason are called "calycal processes" (Fig 5c). In addition, at the apical end, the IS forms a resistant ridge-like structure. Together these two systems work as physical support to the cilium [27] (Fig. 5c). From a metabolic point of view, all the molecules, including rhodopsin, are assembled in the Golgi apparatus, located close to the nuclear part of the IS (Fig. 5d). Conversely the upper part, called "ellipsoid", contains many mitochondria, which are necessary to fill the high-energy demands of the photoreceptors (Fig. 5d). Moreover, the IS directly interacts with Müller glia cells for the exchange of nutrients and in particular for lactate production, which directly enters into the photoreceptor's pyruvate cycle [39]. The IS also has a very well organized microtubule network (cytoplasmic microtubules), necessary for the transport of molecules to different structures within the photoreceptor. The importance of this intracellular activity is clearly demonstrated by

the transport of rhodopsin: Rhodopsin-containing vesicles are generated in the Golgi and moved away through the cell. In greater detail, with the interaction of the rhodopsin protein complex [40] and the protein *ArfGAP with SH3 domain ankyrin repeat and PH domain* (ASAP1), a budding vesicle is formed from the Golgi. This vesicle is then transported though the microtubules network to the apical part of the cell by a dynein driven motor [41]. It has been demonstrated that the suppression of cytoplasmic dynein motors causes the accumulation of post-Golgi vesicles in the IS [42]. After a long "walk" through the IS this shuttle can finally reach the ciliary region, which it has to cross to reach the OS.

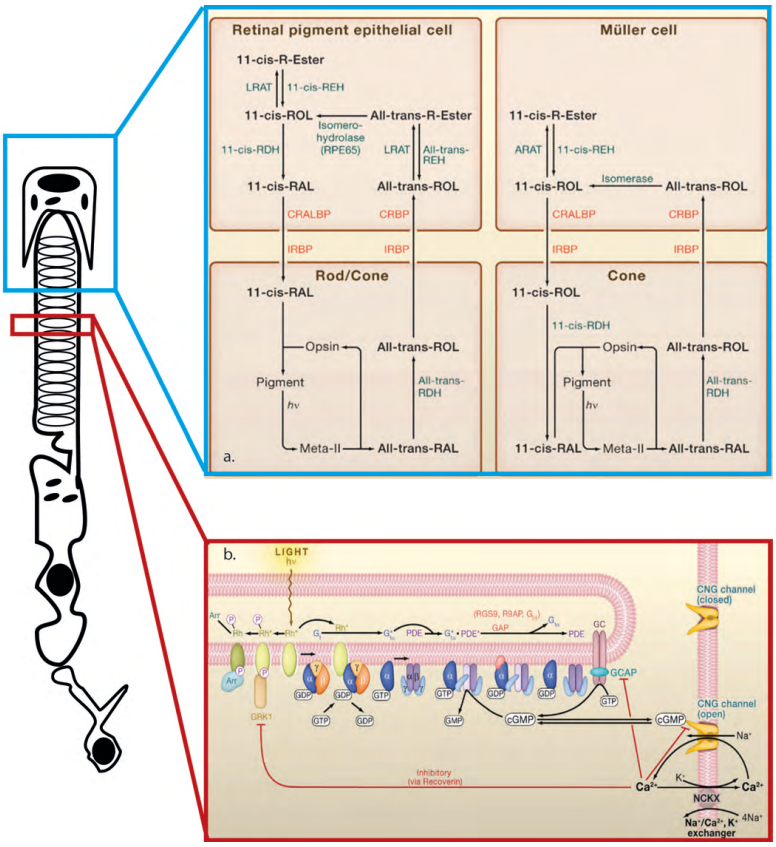


Figure 6. Phototransduction and the visual cycle.
Schematic representation of a photoreceptor cell with two of the most important pathways: the visual cycle (blue box) and the phototransduction cascade (red box). **a)** The visual cycle to restore the rhodopsin molecule. There are two different cycles, one that occurs between cone and rods and RPE cells, and the other that is exclusively for cones and Müller cells. hv, photon; RAL, retinal; ROL, retinol; R-ester, retinyl ester; IRBP, interphotoreceptor retinoid-binding protein; CRBP, cellular retinol-binding protein; CRALBP, cellular retinalheyde-binding protein; RDH, retinol dehydrogenase; REH, retinyl ester hydrolase; LRAT, lecithin:retinol acyl transferase; ARAT, acylCoA:retinol acyl transferase. (adapted from Muniz et al., 2007) **b)** The phototransduction cascade happens in the disc membrane and photoreceptor OS membrane. The light causes a conformational change in rhodopsin (Rh) that becomes active (Rh*). This change requires the G-coupled protein transducin that catylizes GDP to GTP. The alpha subunit of the transducin activates phospodiesterase (PDE) that hydrolyzes cGMP into GMP, reducing the concentration of cGMP in the cell. This causes the closure of the sodium channel (CNG channel) and hyperpolarization of the membrane. Also, the arrestin (Arr) and recoverin pathway, important for inhibiting the process, are depicted in the box. GRK1, G-protein-coupled receptor kinase 1; CNG, cyclic nucleotide-gated; GCAP, guanylatr cyclase activator protein (adapted from Purves et al., 2001).

1. 4 THE PHOTORECEPTOR CONNECTING CILIUM

"Only Connect!" - E.M Forster, English writer, 1910

The cilium is the backbone of the photoreceptor and it has several fundamental functions for photoreceptor biology (Fig. 7a). It can be considered as a modified primary cilium, an organelle found in almost all post-mitotic vertebrate cells. The primary cilium has a structure similar to the motile cilia or bacterial flagella. Although some primary cilia have conserved their motile functions, the majority of mammalian cilia are immotile and have evolved mostly as sensory organelles [43]. Three different parts form the photoreceptor connecting cilium: the basal body (BB), the connecting cilium (CC) and the axoneme (Fig. 7bc). The axoneme is the OS part of the cilium. It is composed of microtubule doublets arranged in a classical 9+0 circular composition and forms the backbone of the whole OS. In the very apical part, however, the axoneme has a 9+0 singlet microtubule structure. The CC corresponds to the transition zone of a cellular primary cilium. It is composed of a 9+0 microtubule doublet and it works as a thin bridge between the IS and the OS (this is the reason why it is called connecting cilium). In the part of the CC close to the IS there is a fibrous structure that cross-links each microtubule doublet. This structure is called a ciliary necklace and forms a pocket around the CC, working as a molecular diffusion barrier [44]. At the base of the CC there is the basal body. This structure originates from the "mother" centriole, which is the oldest centriole derived from mitosis. A specific cloud of proteins surrounding the BB called "appendage" distinguishes this structure from the normal cellular centrioles. These proteins are involved in ciliogenesis [45]. In addition to this, the photoreceptor BB is connected to a protein called rootletin that forms a specific track (called a rootlet) to the synaptic terminus and it has the important function of anchoring the photoreceptor's ER [46].

The cilium has a central function both in the development of the cell and in the mature cell. In development, it directly drives the formation of the OS, with the raising cilium protruding from the basal body and continuing to grow until it reaches the correct size. Around postnatal day 8-10, the developing OS is filled with disc-like membranes that are not yet organized into a regular shape [47]. These discs are then rearranged perpendicularly to the ciliary backbone to form the discs stacks, typical of a mature photoreceptor. The rearrangement is directly mediated by the axoneme microtubules, which work as anchoring points for disc alignment. The final phase

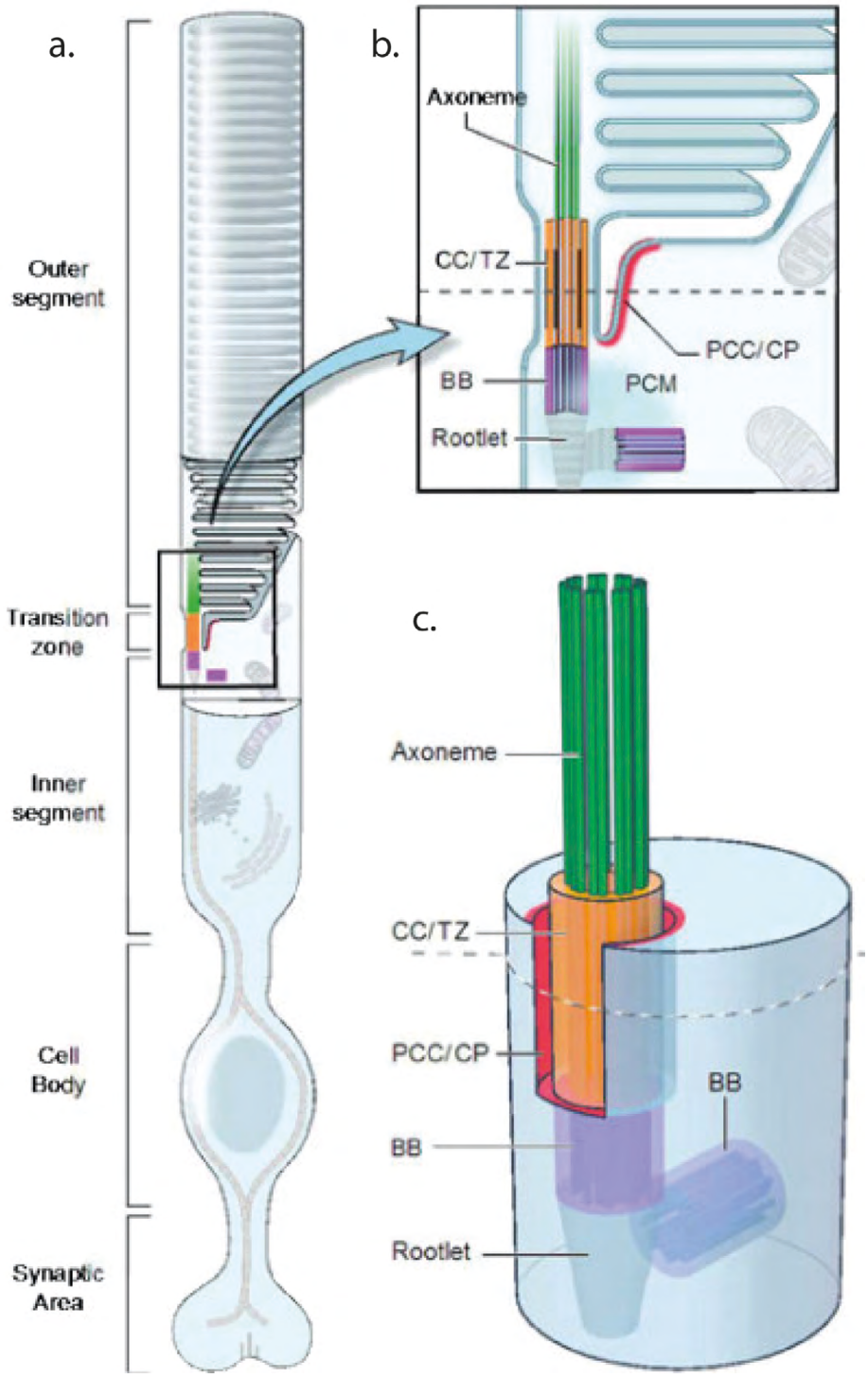


Figure 7. Photoreceptor connecting cilium
A schematic view of the photoreceptor and of the CC area. **a)** Representation of the photoreceptor. The cilium contain an IS compartment that corresponds roughly to the BB (purple), the transition zone between the IS and OS (orange), with the OS part called the axoneme (green). **b)** Magnification of the black box area. It is possible to distinguish the pericentrosomal proteins (PCM) that form a cloud around the BB; the periciliary complex or ciliary pocket (PCC/CP) in between the transition zone; calycal process and the rootlet connected to the BB. **c)** 3D reconstruction of (b). (from Rachel et al, 2012).

of development, starting after day 10, corresponds to the elongation of the axoneme and of the OS to their mature lengths. A protein called RP1 drives the axoneme elongation and the correct stacking of outgrowing discs [48]. This molecule is essential for cilium biology, as demonstrated by the fact that mutations in this gene cause retinitis pigmentosa [49]. It has been clearly shown that RP1 is a microtubule-associated protein (MAP), enhancing the addition of tubulin subunits to the growing cilium [50], and when mutated causes uncontrolled growth of the cilium. During development, newly formed discs are constantly produced at the ciliary region, while RPE cells phagocytose the apical part of the photoreceptor and the oldest and most distal discs. It has been observed, in addition to cilium structure, that the formation of functional stacked discs depends also on the presence of integer opsin molecules. In fact, although knock-out mice for the rhodopsin gene show a normal formation of the primitive cilium, they do not develop a complete OS [51]. Moreover, mutations of the signal motive present at the C-terminal of the rhodopsin gene in humans cause severe retinal degeneration [52].

The formation of stacked discs also takes place in mature photoreceptors throughout life, since light has a damaging effect on the OS and the cell needs a continual renewal of its proteins and membranes. This function is also accomplished by the cilium. However, since the OS and the cilium do not have any biosynthetic machinery, all the proteins and molecules necessary for phototransduction, ciliogenesis and disc formation are produced in the IS and transported to the OS through the cilium. This process is known as *IntraFlagellar Transport* (IFT) and it was first discovered in unicellular organisms [53], although this machinery is well conserved among different species [54]. The IFT is a bidirectional transport of IFT particles along the cilium and axoneme (Fig. 8). It can be considered as a specialized microtubule transport that uses specific motors and it can carry multiprotein IFT particles to the top of the cilium and back. The anterograde transport is mediated by a specific protein called kinesin II (in human this protein is composed of two subunits encoded by the KIF3A, and KIF3B or KIF3C genes [55] plus a non motor protein KAP1). Null mice with a mutation in the KIF3A gene have retinal degeneration, with poor differentiation of the photoreceptor OS, probably due to the accumulation of OS proteins in the IS [56]. Conversely retrograde transport, from the axoneme to the basal body, is mediated by another motor protein called cytoplasmic dynein 2 (which has a heavy chain, encoded by the gene DHC1b/2 in mammalian, and a light chain encoded by D2LIC [57]).

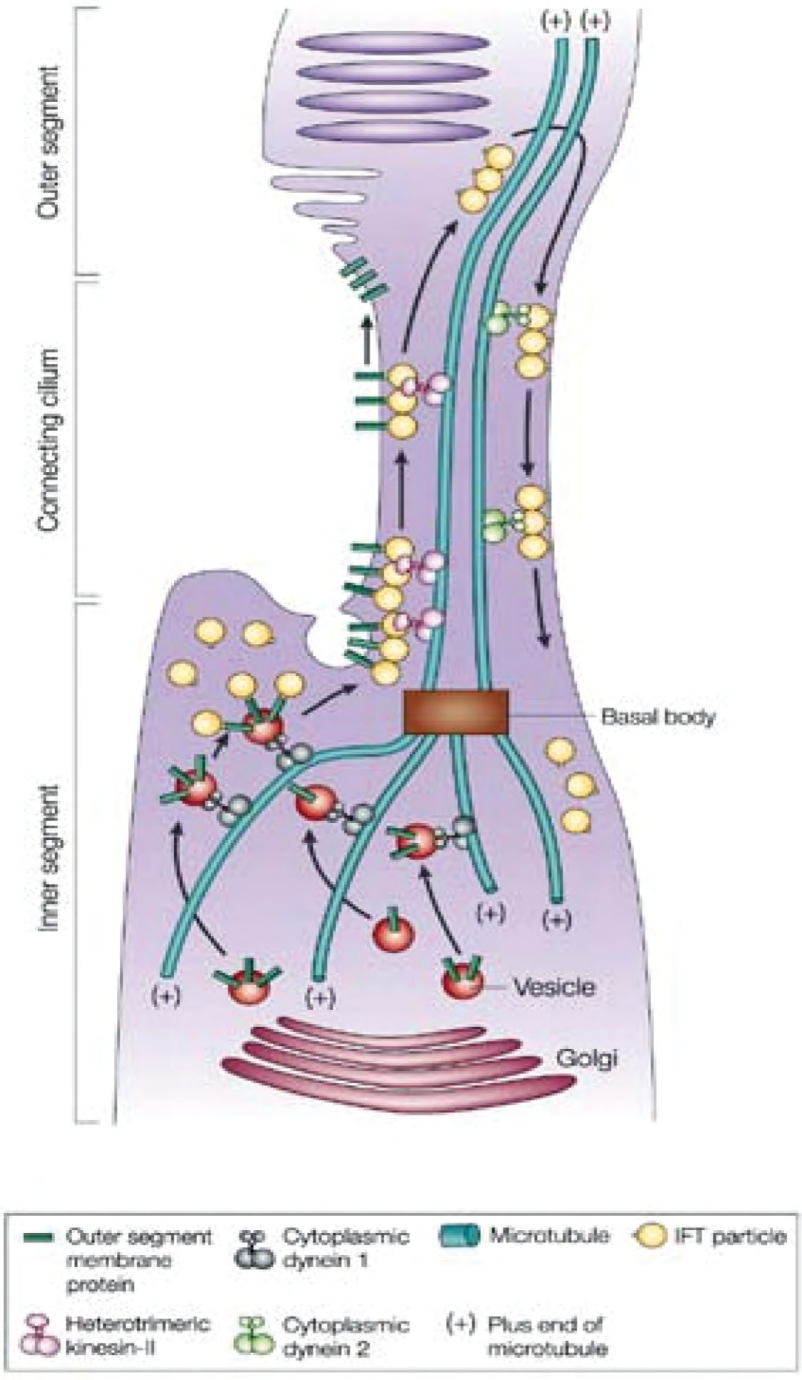


Figure 8. The photoreceptor "traffic" between Golgi and OS.
Model for vesicle movement and IFT in ciliated photoreceptor cells. The vesicles containing membrane proteins for the OS formation (such as rhodopsin) are generated at the Golgi and transferred through cytoplasmatic microtubules at the peri-cilary ridge, where they fuse with photoreceptor membranes. There they go through the basal body by passing through different checkpoints and are sequestered by IFT particles that directly interact with heterodimeric kinesin-II. They go through the cilium until they reach the OS and at that point IFT particles free the cargo. IFT particles are recycled and transported back to the BB thanks to a cytoplasmic dynein II. (Adapted from Rosenbaum et al., 2002).

In addition to these motors, many other proteins are cargo for the IFT. In general the IFT complex can be fractionated by sucrose density gradient and dissociated into two different complexes, named A and B [54]. Very little is known about the composition of these complexes, although proteins containing mainly protein-protein interaction domains, like TPR repeats and coiled-coil domains, are their major constituents. Immunofluorescence and electron microscopy experiments exclusively localize these complexes to the basal body, probably demonstrating their accumulation at the base of the cilium. Although this protein family is widely expressed in every ciliated cell, only one subgroup has been identified in photoreceptors and their depletion impairs photoreceptor development. For instance, IFT88 disruption in zebrafish leads to the complete failure of primary cilium formation [58]. Moreover, a hypomorphic mutation in the same gene causes syndromic ciliary defects in mice, associated with rapid photoreceptor degeneration [59]. In general, the disruption of the IFT complex leads to the accumulation of opsin molecules in the IS, which has a strong toxic effect on the photoreceptor cell. This effect could either be due to the impairment of the primary cilium formation or via the direct damaging of the IFT complex for opsin transport. Recently it has been shown that a conditional knockout for another IFT component, IFT20, causes the accumulation of opsin and rhodopsin in cones and rods respectively, with subsequent retinal degeneration, without affecting the OS formation or ciliogenesis [60]. This, as well as other evidences, support a direct implication of a component in the IFT machinery for the transport of phototransduction molecules.

Other important molecular complexes are directly or indirectly linked to the cilium and to the IFT machinery. In general, mutations in genes that make up a part of these complexes cause widespread and severe diseases, for which retinal degeneration is the main one or just one of the features. Since cilium function or development is severely impaired, these diseases are generally recognized as ciliopathies. One example is the Bardet-Biedl Syndrome (BBS) complex. Mutations in genes encoding proteins belonging to this molecular complex (called BBSome) cause Bardet-Biedl syndrome, which is characterized by hypogonadism, obesity, polydactyly, mental disability, cystic kidneys and retinal degeneration [61]. The BBSome is localized at the base of the cilium, and has a fundamental function in vesicle trafficking to the cilium in association with Rab8 [62]. Recently a new and direct role in IFT trafficking and the turnaround of IFT particles has been clearly demonstrated for some of BBS proteins [63]. Another important complex is the *N*ephrono*PH*tisis *P*rotein complex (NPHP) that is composed of several different proteins (up to 11 loci are mapped on the human genome [64]), for which the major

phenotype is nephrophtisis, the most common cause for early onset renal failure. Often, but not always, mutations in these genes also cause an associated retinal degeneration, and in this case the disease is usually called Senior-Løcken syndrome. For instance, mutations in *NPHP10*, which encodes for the protein *S*erologically *D*efined *C*olon *C*ancer *A*ntigen 8 (*SDCCAG8*) causes renal-retinal ciliopathy [65]. Additionally, *NPHP1* (encoding for nephrocystin) knockout mice show problems with IFT88 localization and an impairment in OS development [66]. Although the function of these proteins and of other NPHP proteins has not yet been clearly elucidated, they are all localized to the BB or CC of the photoreceptor, linking their function to cilium maintenance or development. In addition to the renal-retinal phenotype, mutations in NHPH genes can be also connected to neurological defects, such as mental retardation and ataxia. In this case, the syndrome is called Joubert syndrome [64]. Recently, mutations in the protein *O*ro-*F*acial-*D*igital 1 (*OFD1*) have been shown to lead to a X-linked form of Joubert syndrome [67]. This protein controls the length and structure of distal centrioles and it is important in recruiting IFT88 to the centrosome [68]. Another important NPHP gene is *NPHP6* that encodes for the protein *CE*ntrosomal *P*rotein 290 (*CEP290*). Mutations in this gene are the cause of different ciliopathies such as Joubert syndrome, Senior-Løken syndrome [69, 70], Meckel-Grüber syndrome [71](one of the most severe form of syndromic ciliopathies) and BBS [72]. In addition, it has been clearly demonstrated that hypomorphic mutations in this gene cause a severe early onset retinal degeneration, known as Leber congenital amaurosis, for more than 25% of patients affected by this disease [73]. Similarly, a naturally occurring mouse mutant carrying an in-frame deletion in CEP290, called rd16, presents an early-onset retinal degeneration and mild olfactory dysfunction without any other ciliopathy associated phenotype [74]. Although the function of CEP290 has not been completely clarified, several studies connect this protein to ciliogenesis. Specifically, CEP290 knockdown causes defective ciliogenesis with the mislocalization of Rab8a protein [75]. In addition to this primary effect, several studies demonstrate an active role of CEP290 in the delocalization of cellular inhibitors of ciliogenesis [76, 77].

1.5 Retinitis pigmentosa and related retinal distrophies

"The only thing worse than being blind is having sight but no vision."- Helen Keller, American writer and political activist, first deaf-blind person to earn a BA degree, 1900 circa.

Retinitis pigmentosa (RP) comprises a set of inherited retinal diseases characterized by the progressive degeneration of both types of photoreceptors: the

cones and the rods. The worldwide prevalence is 1 in 4000, with more than 1 million individuals affected overall [78]. In a minority of cases, RP can also be a part of syndromes that target several other organs. The disease presents highly variable symptoms. Some patients develop symptomatic loss of vision in childhood, whereas others remain totally asymptomatic until mid-adulthood. Usually the initial symptoms are associated with rod photoreceptor death leading to difficulties with dark adaption and night blindness. The disease develops until there is a complete loss of peripheral vision, tunnel vision and sometimes loss of central vision. Due to its variable phenotype, diagnosis is mostly based on electrophysiological and imaging analysis. For example, fundus analysis by slit-lamp examination or by ophthalmoscopy, reveals a typical intraretinal pigmentation, the so-called bone spicule deposits, due to the migration of the retinal pigmented epithelium into the neural part retina in response to photoreceptor cell death [79] (Fig. 9a). Another important instrument for diagnosis is the electroretinogram (ERG), which measures the electrical response of the retina to flashes of lights with an electrode on the cornea (defined as a-wave and b-wave amplitude). Patients with RP have a reduction in the electric response amplitude of rods and cones and a delay in response timing, both in scotopic and photopic conditions (Fig. 9c) [80].

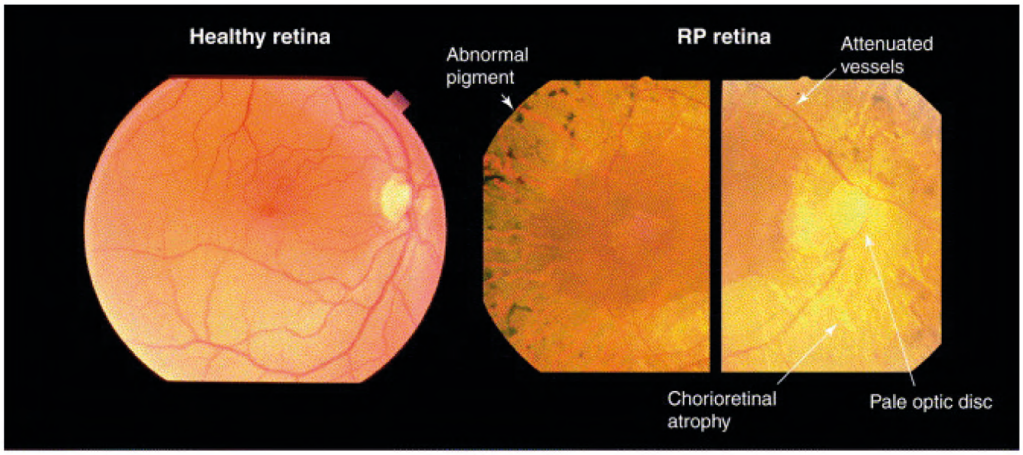
Another non-syndromic retinal dystrophy, closely related to RP, is called Leber congenital amaurosis (LCA), taking its name from the person that first described it [81]. This form of retinal degeneration presents different clinical features, including early severe visual loss, sensory nystagmus, amaurotic pupils and a missing ERG response. It is more rare than RP with an incidence of 1 newborn in every 30.000, but accounts for a large number of children born blind [82]. It can be considered as one of the most severe forms of non syndromic inherited retinal blindness, and it is usually an autosomal recessive condition. Mutations in some genes involved in LCA have also been associated with RP, demonstrating a genetic overlap between these two inherited diseases of retinal degeneration [83, 84].

RP can be present as one symptom in other syndromic disease, such as for example in Usher syndrome, which is associated with mild to severe deafness and is the major cause of inherited deafness-blindness in children. Usher syndrome is a rare disease with an incidence of 1 in 25,000 newborn children [85] and it is inherited as an autosomal recessive trait. In general, genes mutated in Usher syndrome are mostly connected to structural proteins and have an important function for the correct shaping and mechanical functioning of mature stereocilia in auditory epithelia [86]. However,

the reason why they can lead to RP is still debated. In fact, mouse models targeting proteins in Usher syndrome do not show retinal degeneration or abnormalities. The localization of Usher proteins in the proximity of the CC [87, 88] could suggest a functional connection with this organelle, and this hypothesis is also reinforced by the observed interaction of several Usher proteins with ciliary/BB proteins [89, 90]. Recently it has been shown that USH1 protein complex localizes to the calycal process in human photoreceptor cells [91], structures that are closely linked to the CC cilium area. These structures are very well developed in primates, but not in rodents. A role in the formation of the calycal processes could partially explain the missing retinal phenotype in the mutant usher mouse model compared to the observed phenotype in humans.

From a genetic point of view, RP is usually transmitted as a Mendelian trait (autosomal recessive, autosomal dominant, or X-linked), although in a small portion of cases it can display a non-Mendelian pattern of inheritance (digenic trait, di/triallelic trait or maternal inheritance) [92-94]. One distinguishing characteristic of RP, and in general of retinal dystrophies, is their genetic heterogeneity. According to the RETNET database (www.retnet.org) 16 different genes are associated with autosomal dominant RP (adRP), 3 to X-linked RP, and 29 with autosomal recessive RP (arRP) and 7 genes have been associated to both autosomal and recessive RP. Because of this heterogeneity, the molecular diagnosis is difficult, with a large portion of patients worldwide for which the causative gene has not yet been identified. Genes involved in RP and in retinal dystrophies are mostly linked to three major processes of the rod photoreceptors and the RPE (Fig. 10). These include: the retinoid cycle, which takes place in the outer segment of the photoreceptor and in the RPE, the phototransduction cascade in the membranous disc of rod photoreceptors, and the ciliary transport in the connecting cilium, involved in the transport process between the cytoplasm and the outer segment of the photoreceptor [95]. Other genes are involved in photoreceptor development and regulation (like NRL [96] and CRX [97]), or are ubiquitously expressed across tissues (e.g. pre-mRNA processing proteins-PRPs) [98-100].

a.



b.

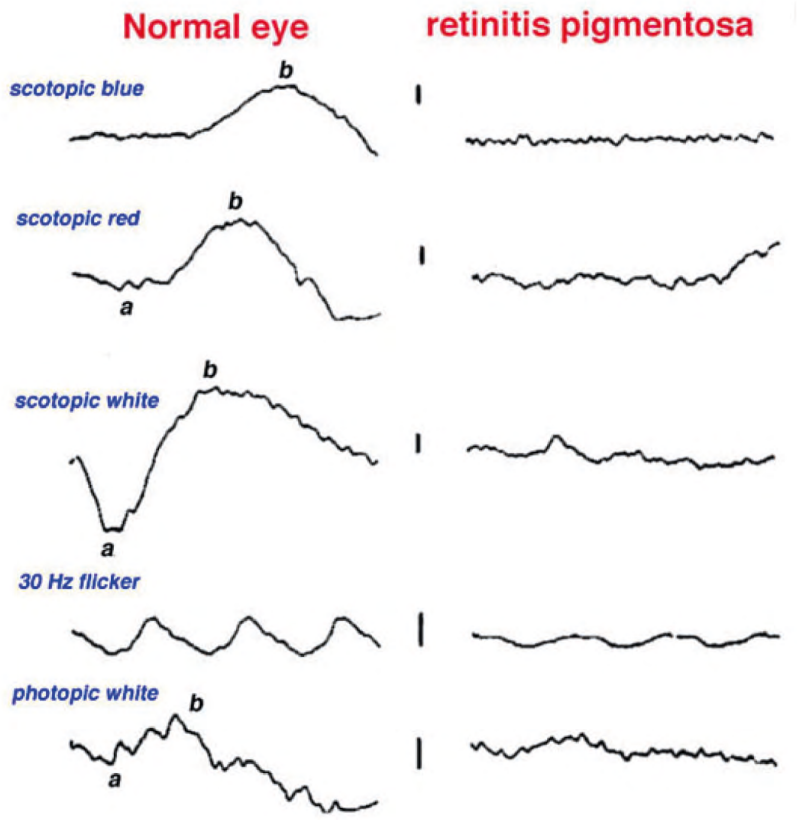


Figure 9. Retinitis pigmentosa, ocular phenotype.
Fundus image and ERG of a control and RP patient. **a)** Left panel, normal retina fundus image. Right panels, typical defects of a RP retina: diffuse pigmentation; pale optic disc; thin blood vessels (from Chapple et al., 2001) **b)** ERG comparison between a control and a RP patient. It is possible to distinguish the differences in amplitude of the a-wave (a) and b-wave (b) both for scotopic and photopic condition in a control eye. These waves are totally abolished or severely reduced in RP retina (from www.webvision.med.utha.edu).

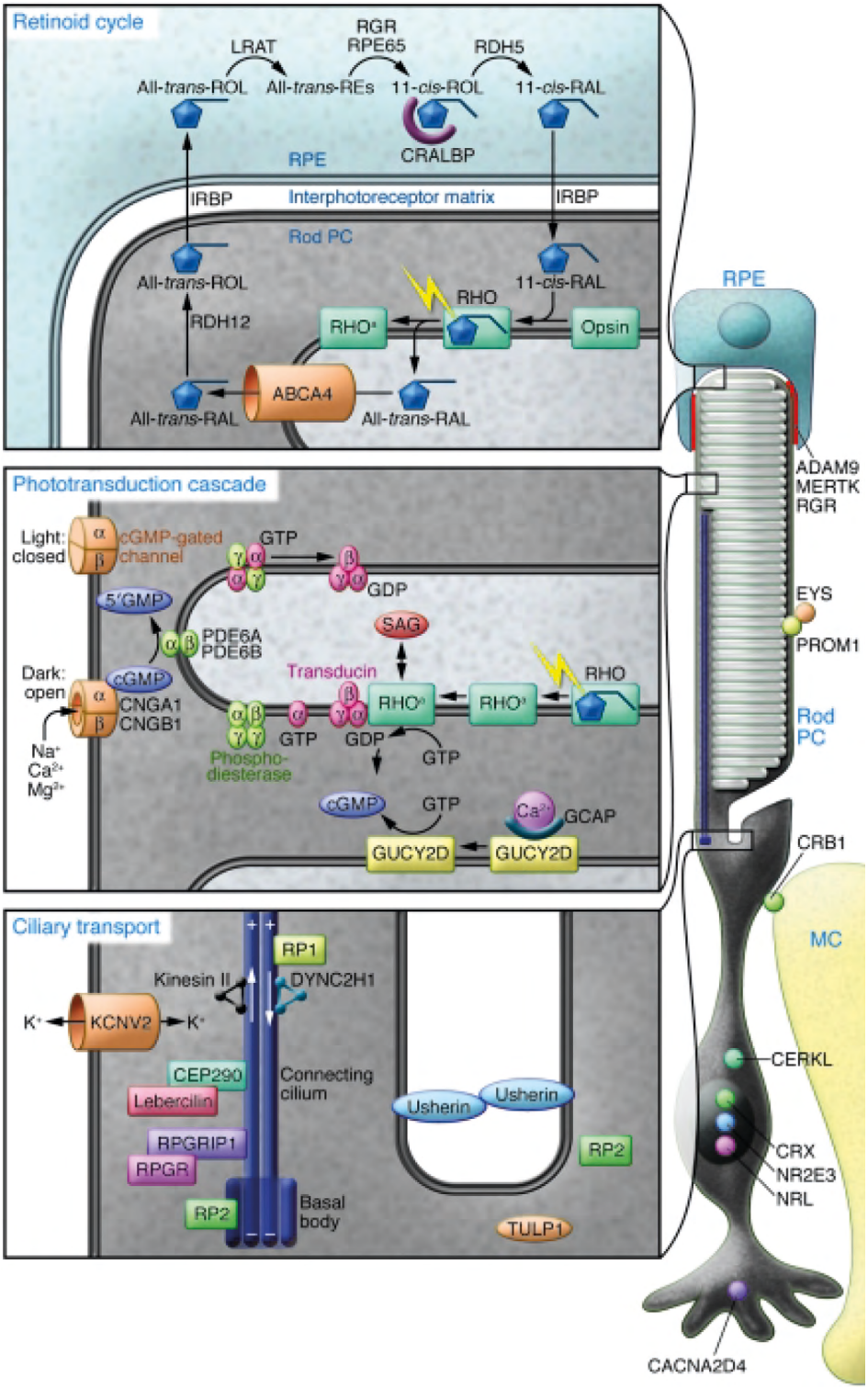


Figure 10. Localization and function of the genes mutated in retinal degenerations. Schematic diagram of a rod photoreceptor and its major biochemical pathways. Upper panel: visual cycle; middle panel: phototransduction cascade; lower panel: ciliary transport. All proteins implied in retinal degeneration are marked in these panels or on the photoreceptor picture. Their real or hypothetical localization is indicated. (From Den Hollander et al., 2010).

1.6 AUTOSOMAL RECESSIVE RP AND THE LOCUS RP28

"If you wanted to dissect the structure of living cells, genetic analysis was an extremely powerful method, so my interest turned to that". – Joshua Lederberg, American molecular biologist, Nobel Prize for medicine, 1958

The most common form of RP is autosomal recessive caused by mutations in several different genes (Figure 11). Despite this genetic heterogeneity, a few of them are responsible for a significant number of cases: *USH2A* (1-7%)[101], *EYS* (11%) [102], *ABCA4* (2-5%) [103], *PDE6A* (4%) [104], and *PDE6B* (4%) [105], *RPE65* (2-11%) [84], while others are rarely found to be mutated in patients . The function of autosomal recessive linked genes (Table 1) is mostly related to photoreceptor metabolism, phototransduction cascade, signaling, trafficking of intracellular proteins, ciliary function, or the cytoskeleton and structure of the photoreceptor. Although most of these genes are expressed only in the retina, others, like *IDH3B*, are ubiquitously expressed, but others organs do not seem to be affected [106].

The locus RP28 was identified in 1999 by genome-wide homozygosity mapping in an Indian family with multiple consanguineous marriages [107]. The family included 4 affected members with arRP. A locus of 16 cM (14Mb) located between 2p11 and 2p15 on the short arm of chromosome 2 and with a maximum LOD score of 3.38 was identified. From a clinical point of view, the age of onset of the affected members in the family varied from 5 to 15 years, and fundus analysis showed classical RP features. The locus was then confirmed and refined in 2004 by the analysis of another, apparently unrelated Indian family by microsatellites mapping of known RP loci [108]. This latter analysis confirmed that the RP28 locus was linked to arRP and narrowed the locus interval to a region of about 1 cM between 2p14 and 2p15. The clinical analysis of affected patients revealed an age of onset between 8-18 years, with classical RP features. Although not well clarified, the identification of heterozygous SNPs in the suspected homozygous region for this second family seems to suggest the possibility that this microsatellites linkage could correspond to a false positive peak.

Table 1. Identified recessive genes and their functions

Gene	Cromosome	Function	Gene	Cromosome	Function
USH2A	1q41	Cell interaction	**RGR**	10q23.1	Vitamin A metabolism
PDE6B	4p16	Phototransduction cascade	**NR2E3**	15q23	Nuclear transcription factor
PDE6A	5q33.1	Phototransduction cascade	**NRL**	14q11.2	Transcription factor
RPE65	1p31.2	Vitamin A metabolism	**RP1**	8q12.1	Microtubule associated protein - cilium
CNGA1	4p12	Phototransduction cascade	**BEST1**	11q12.3	RPE ion channel
CNGB1	16q13	Phototransduction cascade	**C2ORF71**	2p23.2	Unknown Function- cilium
ABCA4	1p22	Vitamin A metabolism	**EYS**	6q12	Unknown Function- Extracellular matrix protein
CRB1	1q31.3	Cell interaction	**FAM161A**	2p15	Unknown Function - cilium
LRAT	4q32.1	Vitamin A metabolism	**IDH3B**	20p13	Krebs cycle protein
MERTK	2q13	Phagocytosis	**IMPG2**	3q12.3	Retnal intracellular matrix
TULP1	6p21.31	Transport of Rhodopsin	**PDE6G**	17q25.3	Phototransduction cascade
RHO	3q22.1	Phototransduction cascade	**RBP3**	10q11.22	Vitamin A metabolism
RLBP1	15q26.1	Vitamin A metabolism	**SPATA7**	14q31.3	Unknown function
CERKL	2q31.3	Ceramide converting enzime	**TTC8**	14q32.11	ciliary functions
SAG	2q37.1	Phototransduction cascade	**ZNF513**	2p23.3	Unknown function- transcription factor
C8ORF37	8q22.1	Unknown function - cilium	**PRCD**	17q25.1	Unknown Function
MAK	6p24.2	Protein kinase - cilium	**CLRN1**	3q25.1	Membrane protein
PROM1	4p15.32	Disc morphogenesis			

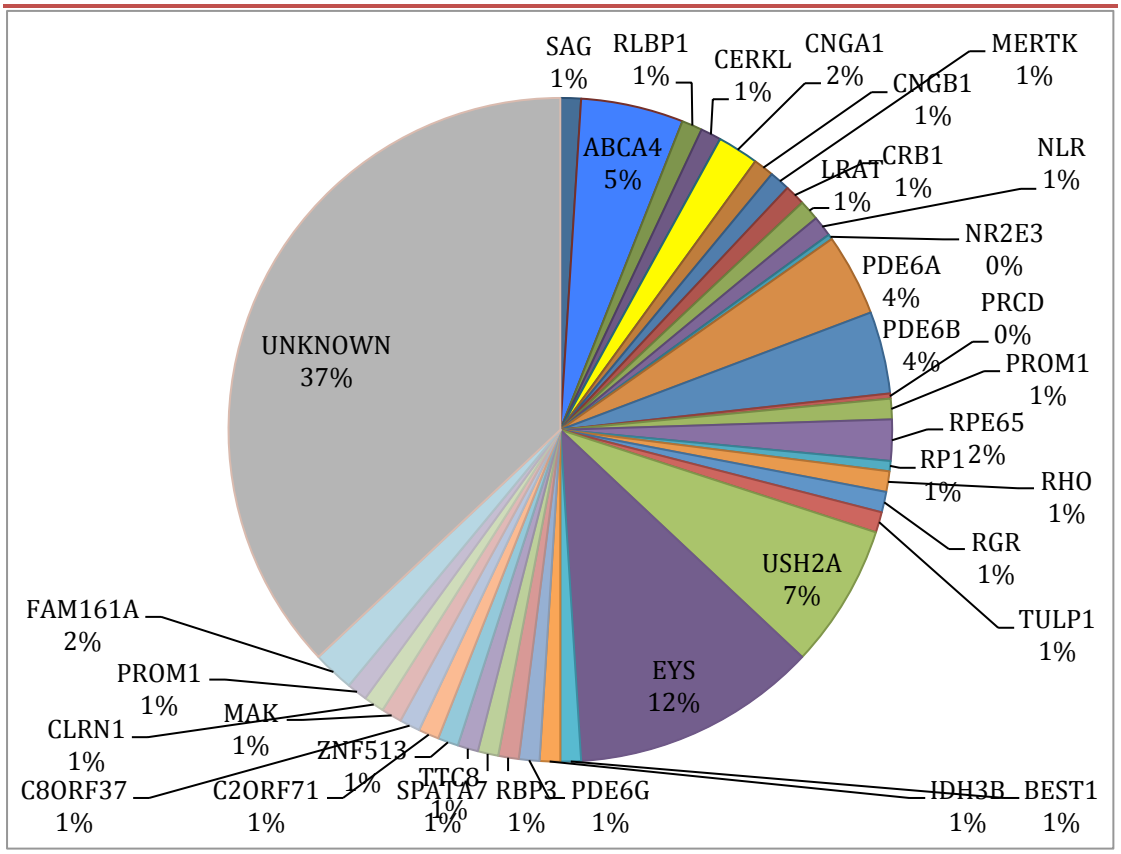


Figure 11. Frequencies of genes mutated in non-syndromic arRP.
The frequencies of the genes were derived from the literature. Whenever mutations could be identified only in one or few cases, their prevalence was set to 1%.

I.2 PAPER

ULTRA HIGH THROUGHPUT SEQUENCING EXCLUDES MDH1 AS CANDIDATE GENE FOR RP28-LINKED RETINITIS PIGMENTOSA.

Thomas Rio Frio,[1] Sylwia Panek,[1] Christian Iseli,[2] Silvio Alessandro Di Gioia,[1] Arun Kumar,[3] Andreas Gal,[4] Carlo Rivolta[1]

[1]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland; [2]Ludwig Institute for Cancer Research and Swiss
Institute of Bioinformatics, Lausanne, Switzerland; [3]Department of Molecular Reproduction, Development and Genetics, Indian
Institute of Science, Bangalore, India; [4]Institut für Humangenetik, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany

Abstract

PURPOSE:
Mutations in IDH3B, an enzyme participating in the Krebs cycle, have recently been found to cause autosomal recessive retinitis pigmentosa (arRP). The MDH1 gene maps within the RP28 arRP linkage interval and encodes cytoplasmic malate dehydrogenase, an enzyme functionally related to IDH3B. As a proof of concept for candidate gene screening to be routinely performed by ultra high throughput sequencing (UHTs), we analyzed MDH1 in a patient from each of the two families described so far to show linkage between arRP and RP28.
METHODS:
With genomic long-range PCR, we amplified all introns and exons of the MDH1 gene (23.4 kb). PCR products were then sequenced by short-read UHTs with no further processing. Computer-based mapping of the reads and mutation detection were performed by three independent software packages.
RESULTS:
Despite the intrinsic complexity of human genome sequences, reads were easily mapped and analyzed, and all algorithms used provided the same results. The two patients were homozygous for all DNA variants identified in the region, which confirms previous linkage and homozygosity mapping results, but had different haplotypes, indicating genetic or allelic heterogeneity. None of the DNA changes detected could be associated with the disease.
CONCLUSIONS:
The MDH1 gene is not the cause of RP28-linked arRP. Our experimental strategy shows that long-range genomic PCR followed by UHTs provides an excellent system to perform a thorough screening of candidate genes for hereditary retinal degeneration.

For this paper my role was marginal, in fact I just performed confirmations of UHT data by Sanger sequencing, but it represented my first contact with the RP28 family.

Molecular Vision 2009; 15:2627-2633 <http://www.molvis.org/molvis/v15/a280>
Received 11 August 2009 | Accepted 1 December 2009 | Published 8 December 2009

Ultra high throughput sequencing excludes *MDH1* as candidate gene for *RP28*-linked retinitis pigmentosa

Thomas Rio Frio,[1] Sylwia Panek,[1] Christian Iseli,[2] Silvio Alessandro Di Gioia,[1] Arun Kumar,[3] Andreas Gal,[4] Carlo Rivolta[1]

(The first two authors contributed equally to this work.)

[1]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland; [2]Ludwig Institute for Cancer Research and Swiss Institute of Bioinformatics, Lausanne, Switzerland; [3]Department of Molecular Reproduction, Development and Genetics, Indian Institute of Science, Bangalore, India; [4]Institut für Humangenetik, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany

Purpose: Mutations in IDH3B, an enzyme participating in the Krebs cycle, have recently been found to cause autosomal recessive retinitis pigmentosa (arRP). The *MDH1* gene maps within the *RP28* arRP linkage interval and encodes cytoplasmic malate dehydrogenase, an enzyme functionally related to IDH3B. As a proof of concept for candidate gene screening to be routinely performed by ultra high throughput sequencing (UHTs), we analyzed *MDH1* in a patient from each of the two families described so far to show linkage between arRP and *RP28*.
Methods: With genomic long-range PCR, we amplified all introns and exons of the *MDH1* gene (23.4 kb). PCR products were then sequenced by short-read UHTs with no further processing. Computer-based mapping of the reads and mutation detection were performed by three independent software packages.
Results: Despite the intrinsic complexity of human genome sequences, reads were easily mapped and analyzed, and all algorithms used provided the same results. The two patients were homozygous for all DNA variants identified in the region, which confirms previous linkage and homozygosity mapping results, but had different haplotypes, indicating genetic or allelic heterogeneity. None of the DNA changes detected could be associated with the disease.
Conclusions: The *MDH1* gene is not the cause of *RP28*-linked arRP. Our experimental strategy shows that long-range genomic PCR followed by UHTs provides an excellent system to perform a thorough screening of candidate genes for hereditary retinal degeneration.

Retinitis pigmentosa (RP; OMIM 268000) is a hereditary and progressive form of retinal degeneration, with an estimated prevalence of one patient in 4,000 people [1]. Affected individuals experience the constant and unstoppable death of photoreceptors, a phenomenon that results in increasing loss of sight and in many instances legal or complete blindness [2]. Genetically, RP is a highly heterogeneous condition, since around 50 genes or loci, most of which act as individual Mendelian entities, have been implicated so far (RetNet).

The *RP28* locus associated with autosomal recessive RP (arRP) has previously been mapped to chromosome 2p14-p15 through the analyses of two consanguineous but apparently unrelated Indian families [3,4]. The candidate region spans 1.06 cM and includes 15 genes, 14 of which are expressed in the retina. None of these genes has been previously associated with retinal degeneration or is known to have a specific function in the retina. The *IDH3B* gene, which encodes for the β subunit of isocitrate dehydrogenase 3 (NAD+ dependent, EC 1.1.1.41), was recently found to be associated with arRP, indicating a link between the Krebs cycle and retinal disease [5]. One of the genes within the *RP28* linkage interval, *MDH1*, encodes for the cytosolic form of malate dehydrogenase (EC 1.1.1.37), which is directly connected to the Krebs cycle via the malate-aspartate shuttle. The gene products of *IDH3B* and *MDH1* are related at an additional functional level, since malate dehydrogenase can convert the product of isocitrate dehydrogenase [6] (Figure 1). Therefore, we reasoned that *MDH1* could correspond to the *RP28* locus, and mutations in its sequence could be responsible for the disease in a manner similar to that of pathogenic changes in *IDH3B*.

A common approach to the mutational screening of candidate genes consists of sequencing their exons and immediate intron boundaries. However, since pathogenic mutations can sometimes be located deep within introns, as was recently shown for retinal degeneration genes as well [7,8], we decided to analyze the full *MDH1* sequence. To circumvent problems linked to sequence length and composition and to test the potential of parallel sequencing in routine mutation detection, we used long-range PCR (LR-

Correspondence to: Carlo Rivolta, Department of Medical Genetics University of Lausanne, Rue du Bugnon 27, 1005 Lausanne, Switzerland; Phone: +41(21) 692-5451; FAX: +41(21) 692-5455; email: carlo.rivolta@unil.ch

TABLE 1. PRIMER PAIRS USED IN LR-PCRS.

Nucleotide sequence (5′-3′)	Size of PCR product (bp)
F: TGTCCGGTCGTCCCAACTTATCAATTC	11,429
R: CTGGTCACTGGCTCCTTGGCATACTTATCTAT	
F: CAAGGAGAACTTCAGTTGCTTGACTCGTTT	6,752
R: AACACCATAGGAGTTGCCATCAGAGATAACAC	
F: TGAGGATTAGGTTTCCCTGGCCTACTTCAC	6,204
R: TCAATTGTGCTACCCAGGTCAGGCTATGA	

Abbreviations: F represents forward primer; R represents reverse primer.

PCR) amplification of genomic DNA followed by ultra high throughput sequencing (UHTs).

METHODS

DNA samples and long-range PCR: The genomic DNA used in this study was part of the collection of samples that originally allowed mapping of the *RP28* locus. Samples were obtained in accordance with the ethical guidelines regulating these previous investigations [3,4]. Specifically, they belonged to patient V-4 from family PMK146 [3] and patient IV-7 from family IIS-2 [4]. The entire *MDH1* gene (nine exons and eight introns) as well as an additional 0.5 kb upstream and 4.7 kb downstream (23,459 bp in total, from nucleotide 63,669,094 to nucleotide 63,692,552 of chromosome 2, NCBI *Homo sapiens* Build 36.3, NC_000002.10) were amplified by three tiled LR-PCRs, using the primer pairs described in Table 1. PCR reaction mixes were set as specified by the standard protocol of TaKaRa LA Taq (TaKaRa Bio Inc., Shiga, Japan) using GC Buffer I, except the final reaction volume was reduced to 10 µl, and the concentration of each primer to 0.1 µM. The thermal profile common to all sets of primers consisted of an initial step at 94 °C for 1 min, followed by 30 cycles at 98 °C for 5 s and 68 °C for 15 min, and a final step at 72 °C for 10 min. Prior to sequencing, LR-PCR products were checked on agarose gel, quantified using the ImageJ software [9], and pooled in roughly equimolar quantities.

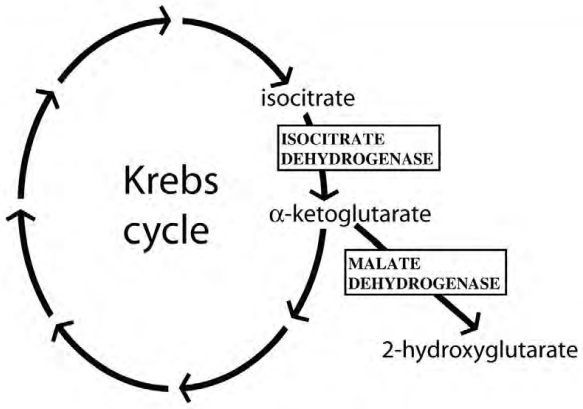


Figure 1. Schematic representation of the Krebs cycle. The specific functions of two enzymes, isocitrate dehydrogenase (NAD+ dependent, partly encoded by *IDH3*) and the *MDH1*-encoded malate dehydrogenase, are shown.

Ultra high throughput (UHT) and conventional Sanger sequencing: UHTs was performed with an Illumina Genome Analyzer (Illumina, San Diego, CA) for the DNA of patient IV-7/IIS-2 and an Illumina Genome Analyzer II (Illumina) with the paired-end procedure for the DNA of patient V-4/PMK146, according to the manufacturer's protocols and starting from around 400 ng of LR-PCR products.

Conventional Sanger sequencing was performed using exon-specific primers (Table 2) and the BigDye Terminator v1.1 cycle sequencing kit (Applied Biosystems, Foster City, CA) on LR-PCR products purified by treatment with ExoSAP-IT (USB, Cleveland, OH). Sequencing reactions were purified on Performa DTR columns (Edge BioSystems, Gaithersburg, MD) and run on an ABI-3130XL (Applied Biosystems).

Sequence analysis: Mapping and analysis of UHTs reads were performed independently with the use of three different software packages: FetchGWI/align0 [10,11], Maq [12], and the CLC Genomics Workbench (CLC bio, Aarhus, Denmark). In all instances, analyses were performed on the original calls generated by the Genome Analyzer. The procedure followed when using the FetchGWI/align0 packages was composed of multiple sequential steps. First, all the reads were filtered according to the Phred quality score of each nucleotide, replacing nucleotides with low scores by "Ns" and discarding reads with more than three Ns. Second, fetchGWI was used to find unique exact matches of all the retained reads. Reads that mapped within the target regions were directly used, while those mapping outside of these regions were kept as a negative control group. Third, among reads for which it was impossible to find any matches (either unique or repeated), those sharing a common 12-mer (except mono- or di-nucleotide repeats) with the selected regions were kept and aligned against them by a global Smith-Waterman procedure (by align0). Fourth, reads that had a good score (no more than three mismatches) were verified not to produce a better alignment score against the negative control group kept during the first step. Finally, DNA variants were derived from

Table 2. Primers used in Sanger sequencing.

Nucleotide sequence (5′-3′)	Exon
GCTCATCCTCAGGGACTACTTTGCAATC	1
CCAGGGTTTGGATCACCACATACTGAAC	2
GTGGTGATTCCTCTCACTGTGTCTGTTAGC	3
CTCAGGCTCCTGAAATGTATATCAGTGTG	4
ATCAAGTAGGAAGTCCAGCCTCT	5
GTCCACAGTTGTTACCACTGTTAAGCTG	6
GCCAGTCATGATCTAGTGTGATCTGATGTG	7
CCTGGTGCTGATGATAGTTCCTTACACA	8
GTAGAGATGGGGTGTCACTATTTG	9

sequences retained at step #1 and from the global alignment outputs produced at step #3. Mapping of sequences and mutation detection via Maq were accomplished as follows. First, Maq performed ungapped alignments of the reads on the reference sequence by retaining only reads with one or no mismatches in the first 24 bases and a maximum of one mismatch for the remaining bases. After mapping, Maq generated a consensus sequence and calculated the Phred quality score at each position of this sequence. Results from the analysis allowed the calling of genomic variations compared to reference sequence. The requirements for calling a variation were a minimum Phred quality score of 40; no insertions or deletions in the range of five bases; only an additional single nucleotide polymorphism (SNP) within a ten-base window; a minimum neighbor Phred quality score of 20; and a minimum coverage of 3×. The CLC Genomics Workbench-assisted mapping and mutation detection were conducted according to the general procedure recommended by the software house. More specifically, the raw sequences first underwent a filtering and trimming process based on the quality of their scores (quality score cutoff=0.01, maximum number of ambiguous nucleotides=2). All reads satisfying these criteria were then mapped on the 23,459-bp reference sequence mentioned above using the Reference Assembly algorithm within the UHTs module of the CLC package (mismatch cost=2, indel costs=3). Detection of variants was performed by identifying on the final mapping all bases for which 60% or more of the calls differed from the reference sequence, regardless of the base coverage.

Electropherograms generated by Sanger sequencing were analyzed either by the Staden software package [13] or the CLC Genomics Workbench (CLC bio).

RESULTS

We sequenced 23,459 bp of human chromosome 2, which span the *MDH1* gene, in patients V-4/PMK146 and IV-7/IIS-2 by UHTs. All three software packages used to filter and align the reads provided essentially the same results and identified the same DNA variants, all of which were intronic (Table 3). As expected, affected members from the two consanguineous families displayed homozygosity for all DNA changes present in the region. This supports the hypothesis that two copies of the same recessive mutation were inherited by patients from a single ancestor who was common to both their paternal and maternal branches, as previous microsatellite analyses have suggested [3,4]. Since all of the identified variants represented either known SNPs or changes with no predicted effects on splicing [14], none of them was considered potentially pathogenic. Sanger sequencing of all *MDH1* exons and their intron vicinities confirmed the absence of DNA variants with respect to the reference sequence, apart from seven SNPs already detected by UHTs: rs10469944, rs2305157, rs7606045, rs2604613, rs262472, rs262473 (in V-4/PMK146) and g.42648177T>A (in IV-7/IIS-2).

The large majority of UHTs reads were of good quality, as only a minimal part of them were discarded by the filtering procedures. Furthermore, filtered sequences could be easily mapped back to the reference sequence of the human genome, with an overall success rate of about 97%. The average length of the reads (30.5 bases) was also satisfying and, given the amount of sequences that were generated, produced a very high mean base coverage. More specifically, sequencing runs produced 5.2 million and 4.8 million raw sequences when the LR-PCRs from patients V-4/PMK146 and IV-7/IIS-2, respectively, were used as a template. Following the Maq procedure, 5.1 million and 4.5 million sequences from these two samples were retained after filtering for quality, generating an average coverage of 7,600× and 6,700× per given base. Slightly lower values were obtained with the CLC bio and FetchGWI/align0 packages. For CLC bio, 4.9 million (V-4/PMK146) and 3.8 million (IV-7/IIS-2) reads from the raw sequences satisfied filtering criteria, producing an average coverage of 7,000× and 4,700×, respectively. For the FetchGWI/align0 procedure, the corresponding values were 5.0 million and 3.1 million sequences, resulting in a mean coverage of 6,500× and 3,700×.

To determine the limits of UHTs as a potential tool for routine mutation detection in a homozygous context, we simulated various coverage values by randomly selecting 500,000, 50,000, 25,000, 10,000, and 5,000 filtered reads from the V-4/PMK146 sample. We then remapped them onto the 23,459-bp reference contig with the CLC bio software, exactly as performed on the original set of sequences. Although the average coverage was substantially different

Table 3. Variants detected in patients V-4/PMK146 and IV7/IIS-2.

Patient	Variant #	dbSNP entry	HGVS name (position)
V-4/PMK146	1	rs10469944	NT_022184.14:g.42632605C>G
	2	rs10469945	NT_022184.14:g.42632709T>C
	3	rs6546018	NT_022184.14:g.42633355G>T
	4	rs6546019	NT_022184.14:g.42633373A>G
	5	rs1446569	NT_022184.14:g.42634247C>G
	6	rs4671519	NT_022184.14:g.42636331C>T
	7	rs2305157	NT_022184.14:g.42638718T>C
	8	rs10865340	NT_022184.14:g.42639982C>G
	9	rs10865341	NT_022184.14:g.42639983A>T
	10	rs11125979	NT_022184.14:g.42640239T>G
	11	rs4671069	NT_022184.14:g.42641337C>A
	12	rs1255	NT_022184.14:g.42641867T>C
	13	rs7606045	NT_022184.14:g.42642461T>G
	14	rs964880	NT_022184.14:g.42642865C>A
	15	rs2121351	NT_022184.14:g.42644294G>A
	16	rs262470	NT_022184.14:g.42646492C>T
	17	rs262471	NT_022184.14:g.42646802G>A
	18	rs2604613	NT_022184.14:g.42648709G>A
	19	rs262472	NT_022184.14:g.42649549G>A
	20	rs262473	NT_022184.14:g.42649746G>T
	21	rs262474	NT_022184.14:g.42650552A>G
	22	rs262489	NT_022184.14:g.42652200G>A
	23	rs262490	NT_022184.14:g.42653584A>G
IV-7/IIS-2	1	NR	NT_022184.14:g.42645344C>G
	2	NR	NT_022184.14:g.42648177T>A

NR, not referenced in public databases.

across the sets, as well as proportional to the simulated coverage (Table 4), we did not observe any dramatic reduction in power for the correct detection of SNPs. Specifically, whenever coverage of at least 2× could be preserved, all SNPs were correctly called (Table 4). Below this threshold, false positives began to appear, as expected, since technical sequencing errors could not be verified by any overlapping sequence (data not shown).

DISCUSSION

Identification of RP genes can be performed via genome-wide linkage analyses (when large families with multiple affected individuals are available), candidate gene screenings (when several dozen unrelated patients can be investigated), or a combination of both techniques. Good candidate genes share sequence similarities with other disease genes, relate to the same biochemical pathway, or are expressed mainly or exclusively in the affected tissue or organ [15,16]. To increase the chances of gene identification, the systematic screening of all genes present within an interval defined by linkage mapping has also been adopted, allowing the discovery of mutations in genes that, a priori, were not considered prime candidates [17,18].

Regardless of the strategy used to select the candidate genes, in most cases only exonic and nearby intronic sequences are investigated, for a few practical reasons. Specifically, exons and surrounding splicing signals are more likely to harbor disease-causing mutations than deep intronic regions. In addition, they are shorter and cheaper to analyze as well as easier to sequence, since they rarely contain repeats or low-complexity regions. However, by definition, this strategy prevents the identification of deep intronic mutations.

In this study, we perform a mutational screening on a candidate gene that maps within a previously identified linkage interval by using UHTs to analyze its introns and exons. We exclude *MDH1* as a candidate gene for *RP28*-linked arRP and determine that patients from the only RP28 families identified so far carry different haplotypes in this region. Under the assumption that the disease is caused by an *RP28* mutation in both families, our finding suggests that these Indian pedigrees are truly unrelated. The absence of a common disease haplotype implies the absence of a founder mutation and, furthermore, does not exclude the possibility that two distinct arRP loci could lie in this region. Therefore, the 13 other genes present in the *RP28* interval and expressed in the retina remain to be screened, likely by a global and blind

Table 4. Variants detected in patient V-4/PMK146 as a function of base coverage.

SNP	Reference base	Calls in patient V-4/PMK146 (absolute coverage of the base)					
		4,000,000 sequences	500,000 sequences	50,000 sequences	25,000 sequences	10,000 sequences	5,000 sequences
rs10469944	C	G (3,901)	G (332)	G (34)	G (22)	G (7)	G (2)
rs10469945	T	C (3,602)	C (375)	C (35)	C (22)	C (6)	C (4)
rs6546018	G	T (3,445)	T (315)	T (30)	T (18)	T (8)	T (5)
rs6546019	A	G (3,403)	G (316)	G (24)	G (13)	G (5)	G (4)
rs1446569	C	G (3,407)	G (327)	G (36)	G (9)	G (5)	G (2)
rs4671519	C	T (4,079)	T (375)	T (35)	T (15)	T (6)	T (4)
rs2305157	T	C (2,671)	C (267)	C (28)	C (11)	No Call	No Call
rs10865340	C	G (2,227)	G (221)	G (27)	G (11)	G (3)	G (2)
rs10865341	A	T (2,228)	T (221)	T (27)	T (11)	T (3)	T (2)
rs11125979	T	G (1,899)	G (172)	G (10)	G (9)	G (2)	No Call
rs4671069	C	A (3,053)	A (275)	A (27)	A (14)	A (7)	No Call
rs1255	T	C (2,588)	C (249)	C (25)	C (15)	C (5)	No Call
rs7606045	T	G (2,390)	G (227)	G (19)	G (13)	G (8)	G (2)
rs964880	C	A (5,158)	A (480)	A (52)	A (24)	A (12)	A (7)
rs2121351	G	A (6,998)	A (684)	A (51)	A (29)	A (14)	A (7)
rs262470	C	T (6,461)	T (620)	T (60)	T (29)	T (13)	T (5)
rs262471	G	A (6,379)	A (545)	A (56)	A (31)	A (13)	A (6)
rs2604613	G	A (5,445)	A (514)	A (69)	A (32)	A (9)	A (4)
rs262472	G	A (7,326)	A (732)	A (70)	A (35)	A (12)	A (6)
rs262473	G	T (3,401)	T (308)	T (33)	T (13)	T (7)	T (3)
rs262474	A	G (7,445)	G (635)	G (51)	G (35)	G (20)	G (5)
rs262489	G	A (6,908)	A (528)	A (47)	A (33)	A (9)	A (6)
rs262490	A	G (4,928)	G (469)	G (39)	G (16)	G (9)	G (2)
W.S. Avg. Cov.		7,053	659	64.5	32.4	12.8	6.3
SNP Avg. Cov.		4,319	399	38.5	20	8.3	4.1

No Call, base call not performed (coverage<2); W.S. Avg. Cov., average base coverage for the whole sequence; SNP Avg. Cov., average base coverage of all detected SNPs.

sequencing strategy. Unlike *MDH1*, no biologic feature currently points to another promising candidate. Since several genes have already been shown to cause RP despite their apparent irrelevance to retinal physiology as well as their ubiquitous expression, the hypothesis that a non-obvious candidate could be the *RP28* gene is not particularly surprising.

However, in addition to these specific results, our work can be considered a proof of concept experiment for the use of highly parallel sequencing techniques for mutation detection in monogenic diseases. The assumption that UHTs could replace conventional sequencing in human gene screening procedures is not completely straightforward, since the former technique has a series of defects (e.g., short read length and relatively poor accuracy in base calling) that render it not particularly suitable for the analysis of the human genome [19]. Although we have willingly reduced the complexity of our analyses by selecting homozygous DNA regions in patients from consanguineous families, our results indicate that mutation detection via UHTs could be performed even when coverage is relatively low. Factors such as heterozygozity (e.g., in dominant diseases) or pooling of multiple samples unquestionably increase the noise and decrease the power of UHTs analyses. However, recent studies have shown that, when specific countermeasures are adopted, these confounding elements can be reduced or eliminated [20-22].

In summary, we believe that the strategy used for the analysis of the *MDH1* gene likely represents one of the technical approaches to future candidate gene screening in monogenic conditions. Since retinitis pigmentosa and allied diseases show great genetic heterogeneity, the screening of large sets of patient DNA samples is required. Therefore, UHTs, in combination with systematic long-range PCR [23-25] or genomic sequence capturing [26,27], can be particularly relevant in the analysis of this disorder.

ACKNOWLEDGMENTS

We would like to acknowledge Dr. Gábor Csárdi for Python programming and Fasteris SA, Plan-les-Ouates, Switzerland, for help on UHT sequencing. Part of the analyses were performed at the Vital-IT Center for high-performance computing of the Swiss Institute of Bioinformatics. Our work was supported by the Swiss National Science Foundation

(grant # 320030–121929) and the CEC (EVI-GENORET LSHG-CT-2005–512036).

REFERENCES

1. Hartong DT, Berson EL, Dryja TP. Retinitis pigmentosa. Lancet 2006; 368:1795-809. [PMID: 17113430]
2. Berson EL. Retinitis pigmentosa. The Friedenwald Lecture. Invest Ophthalmol Vis Sci 1993; 34:1659-76. [PMID: 8473105]
3. Gu S, Kumaramanickavel G, Srikumari CR, Denton MJ, Gal A. Autosomal recessive retinitis pigmentosa locus RP28 maps between D2S1337 and D2S286 on chromosome 2p11-p15 in an Indian family. J Med Genet 1999; 36:705-7. [PMID: 10507729]
4. Kumar A, Shetty J, Kumar B, Blanton SH. Confirmation of linkage and refinement of the RP28 locus for autosomal recessive retinitis pigmentosa on chromosome 2p14-p15 in an Indian family. Mol Vis 2004; 10:399-402. [PMID: 15215745]
5. Hartong DT, Dange M, McGee TL, Berson EL, Dryja TP, Colman RF. Insights from retinitis pigmentosa into the roles of isocitrate dehydrogenases in the Krebs cycle. Nat Genet 2008; 40:1230-4. [PMID: 18806796]
6. Rzem R, Vincent MF, Van Schaftingen E, Veiga-da-Cunha M. L-2-hydroxyglutaric aciduria, a defect of metabolite repair. J Inherit Metab Dis 2007; 30:681-9. [PMID: 17603759]
7. den Hollander AI, Koenekoop RK, Yzer S, Lopez I, Arends ML, Voesenek KE, Zonneveld MN, Strom TM, Meitinger T, Brunner HG, Hoyng CB, van den Born LI, Rohrschneider K, Cremers FP. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. Am J Hum Genet 2006; 79:556-61. [PMID: 16909394]
8. Rio Frio T, McGee TL, Wade NM, Iseli C, Beckmann JS, Berson EL, Rivolta C. A single-base substitution within an intronic repetitive element causes dominant retinitis pigmentosa with reduced penetrance. Hum Mutat 2009; 30:1340-7. [PMID:19618371]
9. Abramoff MD, Magelhaes PJ, Ram SJ. Image Processing with ImageJ. Biophotonics International 2004; 11:36-42.
10. Iseli C, Ambrosini G, Bucher P, Jongeneel CV. Indexing strategies for rapid searches of short words in genome sequences. PLoS One 2007; 2:e579. [PMID: 17593978]
11. Myers EW, Miller W. Optimal alignments in linear space. Comput Appl Biosci 1988; 4:11-7. [PMID: 3382986]
12. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 2008; 18:1851-8. [PMID: 18714091]
13. Staden R, Beal KF, Bonfield JK. The Staden package, 1998. Methods Mol Biol 2000; 132:115-30. [PMID: 10547834]
14. Pedersen AG, Nielsen H. Neural network prediction of translation initiation sites in eukaryotes: perspectives for EST and genome analysis. Proc Int Conf Intell Syst Mol Biol 1997; 5:226-33. [PMID: 9322041]
15. Dryja TP. Human genetics. Deficiencies in sight with the candidate gene approach. Nature 1990; 347:6142215691. [PMID: 2215691]
16. Dryja TP. Gene-based approach to human gene-phenotype correlations. Proc Natl Acad Sci USA 1997; 94:12117-21. [PMID: 9342372]
17. McKie AB, McHale JC, Keen TJ, Tarttelin EE, Goliath R, van Lith-Verhoeven JJ, Greenberg J, Ramesar RS, Hoyng CB, Cremers FP, Mackey DA, Bhattacharya SS, Bird AC, Markham AF, Inglehearn CF. Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). Hum Mol Genet 2001; 10:1555-62. [PMID: 11468273]
18. Chakarova CF, Papaioannou MG, Khanna H, Lopez I, Waseem N, Shah A, Theis T, Friedman J, Maubaret C, Bujakowska K, Veraitch B, Abd El-Aziz MM. Prescott de Q, Parapuram SK, Bickmore WA, Munro PM, Gal A, Hamel CP, Marigo V, Ponting CP, Wissinger B, Zrenner E, Matter K, Swaroop A, Koenekoop RK, Bhattacharya SS. Mutations in TOPORS cause autosomal dominant retinitis pigmentosa with perivascular retinal pigment epithelium atrophy. Am J Hum Genet 2007; 81:1098-103. [PMID: 17924349]
19. Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. Genomics 2008; 92:255-64. [PMID: 18703132]
20. Ingman M, Gyllensten U. SNP frequency estimation using massively parallel sequencing of pooled DNA. Eur J Hum Genet 2009; 17:383-6. [PMID: 18854868]
21. Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA, Homer N, Huentelman MJ. Identification of genetic variants using bar-coded multiplexed sequencing. Nat Methods 2008; 5:887-93. [PMID: 18794863]
22. Comabella M, Craig DW, Camiña-Tato M, Morcillo C, Lopez C, Navarro A, Rio J, Biomarker MS. Study Group, Montalban X, Martin R. Identification of a novel risk locus for multiple sclerosis at 13q31.3 by a pooled genome-wide scan of 500,000 single nucleotide polymorphisms. PLoS One 2008; 3:e3490. [PMID: 18941528]
23. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. Whole-genome patterns of common DNA variation in three human populations. Science 2005; 307:1072-9. [PMID: 15718463]
24. Harismendy O, Frazer K. Method for improving sequence coverage uniformity of targeted genomic intervals amplified by LR-PCR using Illumina GA sequencing-by-synthesis technology. Biotechniques 2009; 46:229-31. [PMID: 19317667]
25. Yeager M, Xiao N, Hayes RB, Bouffard P, Desany B, Burdett L, Orr N, Matthews C, Qi L, Crenshaw A, Markovic Z, Fredrikson KM, Jacobs KB, Amundadottir L, Jarvie TP, Hunter DJ, Hoover R, Thomas G, Harkins TT, Chanock SJ. Comprehensive resequence analysis of a 136 kb region of human chromosome 8q24 associated with prostate and colon cancers. Hum Genet 2008; 124:161-70. [PMID: 18704501]
26. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 2009; 27:182-9. [PMID: 19182786]
27. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA. Direct selection of human genomic loci by microarray hybridization. Nat Methods 2007; 4:903-5. [PMID: 17934467]

I.3 PAPER

NONSENSE MUTATIONS IN *FAM161A* CAUSE *RP28*-ASSOCIATED RECESSIVE RETINITIS PIGMENTOSA.

Thomas Langmann[1, 9], Silvio Alessandro Di Gioia[2, 9], Isabella Rau[3, 9], Heidi Stöhr[1], Nela S. Maksimovic[3, 11], Joseph C. Corbo[4], Agnes B. Renner[5], Eberhart Zrenner[6], Govindasamy Kumaramanickavel[7], Marcus Karlstetter[1], Yvan Arsenijevic[8], Bernhard H.F. Weber[1], Andreas Gal[3, 10], Carlo Rivolta[2, 10]

[1] Institute of Human Genetics, University of Regensburg, D-93053 Regensburg, Germany
[2] Department of Medical Genetics, University of Lausanne, CH-1005 Lausanne, Switzerland
[3] Institute of Human Genetics, University Medical Centre Hamburg-Eppendorf, D-20246 Hamburg, Germany
[4] Department of Pathology and Immunology, Washington University School of Medicine, Saint Louis, MO 63110, USA
[5] Department of Ophthalmology, University Medical Center Regensburg, D-93053 Regensburg, Germany
[6] Centre for Ophthalmology, Institute for Ophthalmic Research, University of Tübingen, D-72076 Tübingen, Germany
[7] Department of Genetics and Molecular Biology, Sankara Nethralaya, Chennai - 600 006, India
[8] Unit of Gene Therapy and Stem Cell Biology, Jules-Gonin Eye Hospital, University of Lausanne, CH-1004 Lausanne, Switzerland
[9] These authors contribute equally to this work

ABSTRACT

Retinitis pigmentosa (RP) is a degenerative disease of the retina leading to progressive loss of vision and, in many instances, to legal blindness at the end stage. The RP28 locus was assigned in 1999 to the short arm of chromosome 2 by homozygosity mapping in a large Indian family segregating autosomal-recessive RP (arRP). Following a combined approach of chromatin immunoprecipitation and parallel sequencing of genomic DNA, we identified a gene, FAM161A, which was shown to carry a homozygous nonsense mutation (p.Arg229X) in patients from the original RP28 pedigree. Another homozygous FAM161A stop mutation (p.Arg437X) was detected in three subjects from a cohort of 118 apparently unrelated German RP patients. Age at disease onset in these patients was in the second to third decade, with severe visual handicap in the fifth decade and legal blindness in the sixth to seventh decades. FAM161A is a phylogenetically conserved gene, expressed in the retina at relatively high levels and encoding a putative 76 kDa protein of unknown function. In the mouse retina, Fam161a mRNA is developmentally regulated and controlled by the transcription factor Crx, as demonstrated by chromatin immunoprecipitation and organotypic reporter assays on explanted retinas. Fam161a protein localizes to photoreceptor cells during development, and in adult animals it is present in the inner segment as well as the outer plexiform layer of the retina, the synaptic interface between photoreceptors and their efferent neurons. Taken together, our data indicate that null mutations in FAM161A are responsible for the RP28-associated arRP.

In this paper, for which I participated equally as first author, I performed the gene capture of the whole RP28 locus and I first discovered the nonsense mutation in FAM161A in an affected member of the original RP28 family. I performed the RT-PCR analysis on human tissue and designed the custom antibody. The immunohistochemistry was performed as a collaboration between myself and Yvan Arsenijevic.

REPORT

Nonsense Mutations in *FAM161A* Cause *RP28*-Associated Recessive Retinitis Pigmentosa

Thomas Langmann,[1,9] Silvio Alessandro Di Gioia,[2,9] Isabella Rau,[3,9] Heidi Stöhr,[1] Nela S. Maksimovic,[3,11] Joseph C. Corbo,[4] Agnes B. Renner,[5] Eberhart Zrenner,[6] Govindasamy Kumaramanickavel,[7] Marcus Karlstetter,[1] Yvan Arsenijevic,[8] Bernhard H.F. Weber,[1] Andreas Gal,[3,10,*] and Carlo Rivolta[2,10,*]

Retinitis pigmentosa (RP) is a degenerative disease of the retina leading to progressive loss of vision and, in many instances, to legal blindness at the end stage. The *RP28* locus was assigned in 1999 to the short arm of chromosome 2 by homozygosity mapping in a large Indian family segregating autosomal-recessive RP (arRP). Following a combined approach of chromatin immunoprecipitation and parallel sequencing of genomic DNA, we identified a gene, *FAM161A*, which was shown to carry a homozygous nonsense mutation (p.Arg229X) in patients from the original *RP28* pedigree. Another homozygous *FAM161A* stop mutation (p.Arg437X) was detected in three subjects from a cohort of 118 apparently unrelated German RP patients. Age at disease onset in these patients was in the second to third decade, with severe visual handicap in the fifth decade and legal blindness in the sixth to seventh decades. *FAM161A* is a phylogenetically conserved gene, expressed in the retina at relatively high levels and encoding a putative 76 kDa protein of unknown function. In the mouse retina, Fam161a mRNA is developmentally regulated and controlled by the transcription factor Crx, as demonstrated by chromatin immunoprecipitation and organotypic reporter assays on explanted retinas. Fam161a protein localizes to photoreceptor cells during development, and in adult animals it is present in the inner segment as well as the outer plexiform layer of the retina, the synaptic interface between photoreceptors and their efferent neurons. Taken together, our data indicate that null mutations in *FAM161A* are responsible for the *RP28*-associated arRP.

Retinitis pigmentosa (RP [MIM 268000]) is a hereditary blinding condition that affects approximately 1 million individuals worldwide. It is genetically heterogeneous, with mutations in more than 45 genes identified to date according to the RetNet database. Nevertheless, these mutations account for only about 50% of the existing disease alleles.[1] Many of the known RP genes are expressed in rod photoreceptors, although during the disease process both rod and cones progressively degenerate as a result of mechanisms that are only partly known.[2] The *RP28* locus (MIM 606068) was mapped to chromosome 2p11-p15 in a large consanguineous family (PMK146) from India segregating autosomal-recessive RP (arRP),[3] whereas the corresponding gene and the underlying genetic defect were not identified. Homozygosity mapping defined a 15 Mb interval between *D2S1337* and *D2S286* harboring 188 annotated sequences or putative genes. To search for the gene implicated in the *RP28* phenotype, we implemented a synergistic strategy based on ultrahigh-throughput sequencing (UHTs) of microarray-captured genomic DNA and chromatin immunoprecipitation coupled to UHTs (ChIP-Seq). Specifically, results from UHTs-based mutational screening of *RP28* patient DNA were merged with data from genome-wide analyses of murine in vivo targets for the retinal transcription factor Crx[4] for the identification of candidate genes localized within the *RP28* interval. This approach was based on our observation that phylogenetically conserved Crx-bound regions (CBRs) detected by ChIP-Seq were present in 95% (61/64) of retina-enriched genes and in the majority of known RP-associated genes, suggesting that nearly all identified retinal disease genes with retina-enriched expression are direct Crx targets.

Our study involved human subjects and was carried out in accordance with the tenets of the Declaration of Helsinki and the ethical guidelines of our institutions. Written informed consent was obtained from all participants. Traces of DNA from patient V-4 of family PMK146[3] were amplified by whole-genome amplification (QIAGEN, Venlo, The Netherlands). Fifteen micrograms of DNA were used in a sequence-capture experiment (SureSelect, Agilent, Santa Clara, CA, USA) with custom-designed probes targeting all of the 1643 exons (and their intronic vicinities) present in the *RP28* interval. The captured DNA was then further processed for UHTs with a Genome Analyzer II (Illumina, San Diego, CA, USA) and mapped back to the human reference sequence

[1]Institute of Human Genetics, University of Regensburg, D-93053 Regensburg, Germany; [2]Department of Medical Genetics, University of Lausanne, CH-1005 Lausanne, Switzerland; [3]Institute of Human Genetics, University Medical Centre Hamburg-Eppendorf, D-20246 Hamburg, Germany; [4]Department of Pathology and Immunology, Washington University School of Medicine, Saint Louis, MO 63110, USA; [5]Department of Ophthalmology, University Medical Center Regensburg, D-93053 Regensburg, Germany; [6]Centre for Ophthalmology, Institute for Ophthalmic Research, University of Tübingen, D-72076 Tübingen, Germany; [7]Department of Genetics and Molecular Biology, Sankara Nethralaya, Chennai - 600 006, India; [8]Unit of Gene Therapy and Stem Cell Biology, Jules-Gonin Eye Hospital, University of Lausanne, CH-1004 Lausanne, Switzerland
[9]These authors contributed equally to this work
[10]These authors contributed equally to this work
[11]Permanent address: Institute of Human Genetics, School of Medicine, University of Belgrade, 11000 Belgrade, Serbia
*Correspondence: gal@uke.de (A.G.), carlo.rivolta@unil.ch (C.R.)
DOI 10.1016/j.ajhg.2010.07.018.

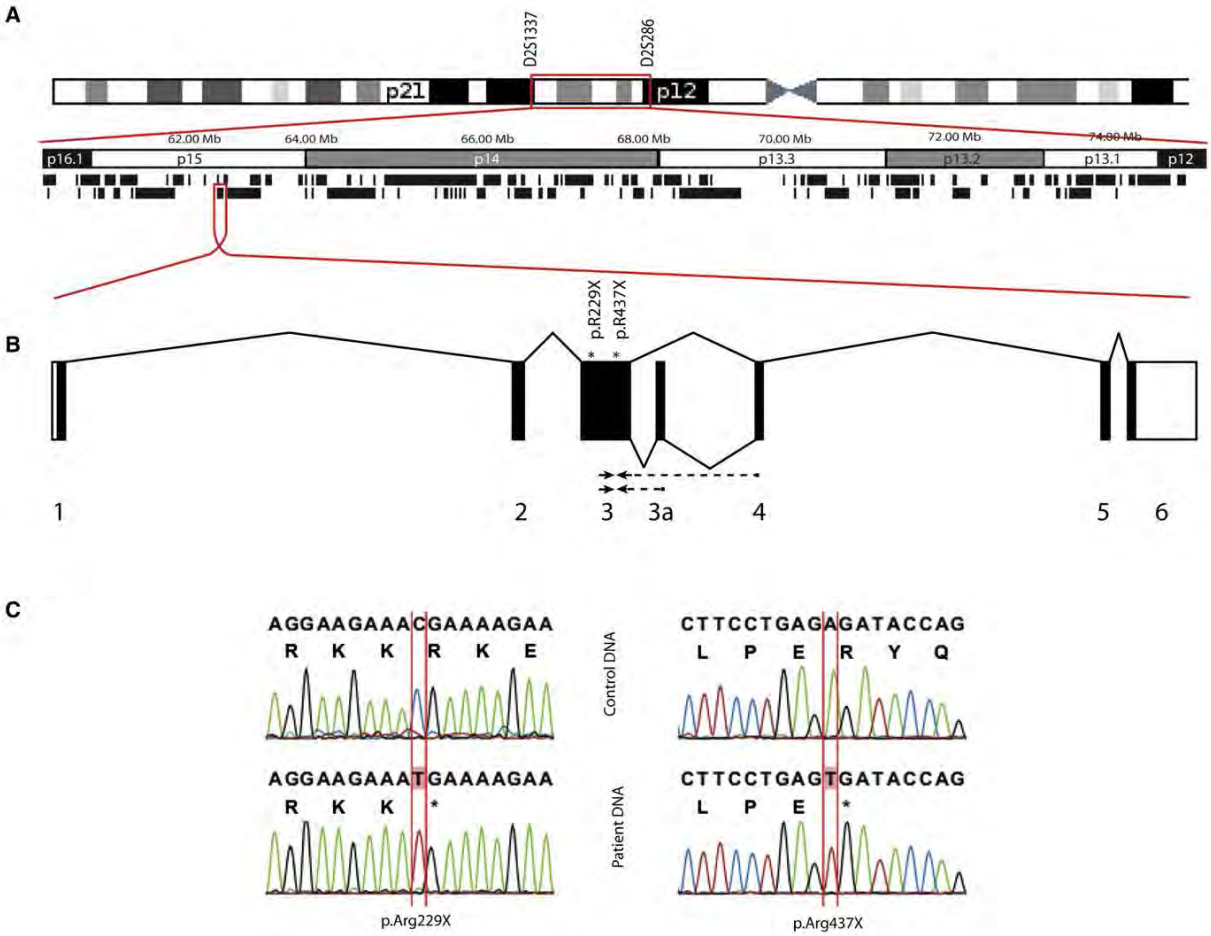


Figure 1. Schematic Representation of the *RP28* Interval and Mutations in *FAM161A*

(A) Linkage interval for *RP28* on chromosome 2p is delimited by microsatellite markers D2S1337 and D2S286 and spans over 15 Mb of DNA. Black boxes represent the 188 genes screened by UHTs.

(B) Structure of the two protein-coding isoforms of *FAM161A*, which are transcribed in the reverse direction with respect to the indicated orientation of chromosome 2, and location of the nonsense mutations detected in patients with arRP. Arrows indicate the primers used in quantitative RT-PCR experiments for detection of FAM161A mRNA expression in human tissues. The forward primer for both RT-PCR reactions lies on exon 3, whereas the reverse primer is complementary to either the 3-4 or the 3-3a exon-exon junctions, depending on the isoform analyzed.

(C) Electropherograms of the c.685C>T (p.Arg229X) and c.1309A>T (p.Arg437X) mutations and the corresponding region in controls.

NC_000002.11. We obtained 14.8 million reads, corresponding to approximately 500 Mb of contiguous sequence, 62% of which could be aligned to exons from the candidate chromosomal region. The reads were assembled and evaluated as described previously[5] and produced a 450-fold coverage per bp on average. In total, eight nonannotated homozygous variants were found with respect to the reference sequence (Table S1, available online).

In a complementary approach, alignment of mouse CBRs identified through ChIP-Seq and matched to the orthologous human *RP28* interval yielded 15 presumptive Crx target genes (Table S2). The gene with the strongest Crx binding, as reflected by 205 ChIP-Seq reads at the corresponding mouse locus, was also the top candidate from the UHTs mutational screening. It corresponded to the uncharacterized gene *FAM161A*, bearing a homozygous c.685C>T (NM_032180.2) substitution in exon 3 that should result in the nonsense change p.Arg229X (Figure 1). As expected, this mutation was present in all four affected members of family PMK146 in a homozygous state, whereas unaffected relatives of the patients were either heterozygous or carried wild-type alleles. These data suggest *FAM161A* to be the *RP28* gene previously mapped to 2p11-p15 in this family.

Further screening of *FAM161A* in 118 presumably unrelated patients from Germany with recessive or isolate forms of RP revealed the presence of another homozygous nonsense mutation (c.1309A>T [p.Arg437X]; Figure 1) in exon 3 of three patients (M09-0352, arRP173, and arRP323; Table 1) that cosegregated with the RP phenotype in the respective families. This mutation was absent in 400 ethnically matched control chromosomes ($p = 5.4 \times 10^{-3}$,

Table 1. *FAM161A* Mutations in Patients with Retinitis Pigmentosa

Index Patient ID	Ethnic Origin	Exon	Nucleotide Change	Allele Status	Protein Variant
V-4	Indian	3	c.685C>T	homozygous	p.Arg229X
M09-0352	German	3	c.1309A>T	homozygous	p.Arg437X
arRP173	German	3	c.1309A>T	homozygous	p.Arg437X
arRP323	German	3	c.1309A>T	homozygous	p.Arg437X

by chi-square). Both the c.685C>T and c.1309A>T changes represent likely functional null alleles with absence of the gene product, because the mRNA is predicted to undergo nonsense-mediated decay.[6] The three German patients carrying the p.Arg437X mutation originated from different geographical regions, but they shared the same *FAM161A* haplotype composed of six intragenic SNPs (Table S3). The G allele of rs11125895:A>G was present in German controls (n = 178) at a frequency of 0.167. The finding of homozygosity in these apparently unrelated patients suggests that p.Arg437X represents a mutation originating from a common founder. Other DNA changes detected in *FAM161A* included rare missense or isocoding changes of unknown pathogenicity and SNPs (Tables S4 and S5). Three of the nonsynonymous sequence variants affect evolutionarily conserved residues and were not detected in 400 control chromosomes. However, all were found in a heterozygous state only. Thus, the significance of these rare variants in the pathogenesis of the retinal phenotype of the patients remains to be elucidated.

It seems that the *FAM161A*-type retinal dystrophy shows no unique clinical features. The course of the disease is rather slow, with severe visual handicap in the fifth decade and legal blindness in the sixth to seventh decades. Patient M09-0352 was a 19-year-old female who complained about progressive constriction of the visual field, disturbed night vision, slight hearing problems, and hyposmia at her first visit. She was the offspring of unaffected parents with no history of consanguinity. Best corrected visual acuity was 20/25 in both eyes. Funduscopy showed typical findings of RP, such as bone spicules in the midperiphery, narrowed vessels, and optic disc pallor. Fundus autofluorescence was mottled and reduced in the midperiphery and showed a fine, small ring of increased autofluorescence around the fovea. Goldmann perimetry demonstrated distinct constricted visual fields in both eyes. The outer borders were at 10° excentricity. Full-field electroretinogram was not recordable. Patients arRP173 and arRP323 were 70 and 69 years old, respectively. These women were the only affected offspring of apparently unrelated parents. In both cases, age at onset was reported to be in the second to third decade. The diagnosis of RP was made at the ages of 30 (arRP323) and 42 (arRP173) years. Disease onset in the three affected siblings of the PMK146 family was reported to be around 15 years of age. At the ages of 42 and 47 years, respectively, patients V-2 and V-3 had a severe visual handicap but were not regarded to be legally blind. They presented with fundus changes typical for RP, including narrowing of the arteries, bone spicule pigments, and optic atrophy. The age at onset was considerably earlier (at the age of 5 years) in patient VI-2, with a more rapid disease progression: at the age of 15 years, she could count fingers close to face (OD) and had a visual acuity of 3/60 (OS). It therefore appears that *FAM161A* defects lead to a variable phenotype in terms of disease onset and progression. The differences could be due to genetic or environmental modifiers.

FAM161A is an uncharacterized sequence with multiple splicing variants, two of which are predicted to result in stable mRNA transcripts. The most abundant derives from six exons and encodes a predicted protein of 660 amino acid residues (76 kDa), whereas the second isoform contains a supplementary in-frame 168 bp exon

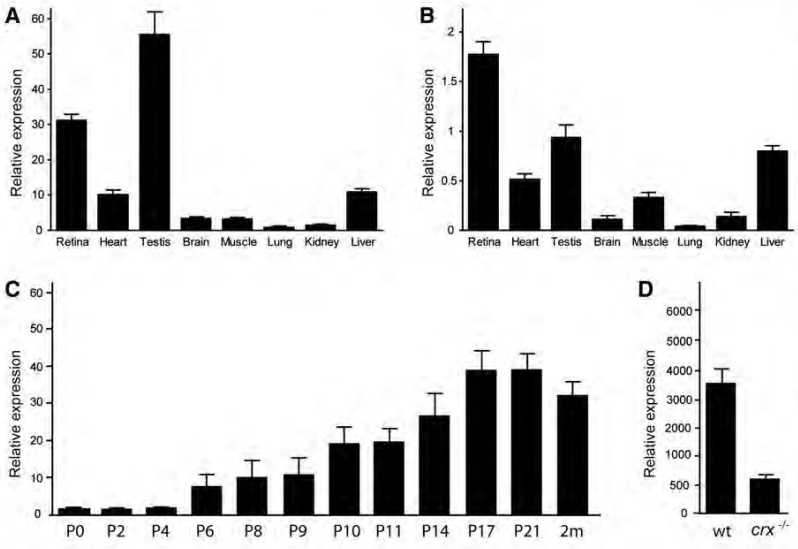


Figure 2. FAM161A mRNA Expression in Human and Mouse

(A and B) Quantitative RT-PCR in various human tissues of the FAM161A short isoform (A) and long isoform (B). Both graphs represent mRNA expression normalized to the housekeeping gene *ACTB*. Note that ranges of y axes are not the same.

(C) Fam161a expression in the mouse retina, as detected by quantitative RT-PCR with the use of the 18S rRNA as housekeeping gene at various developmental stages.

(D) Affymetrix GeneChip expression data from wild-type versus $Crx^{-/-}$ P14 retinas, as detected with the Fam161a-specific probe set 1443569_at.

Error bars indicate standard errors of the mean (A and B) and standard deviations (C and D), calculated in all instances from three independent experiments.

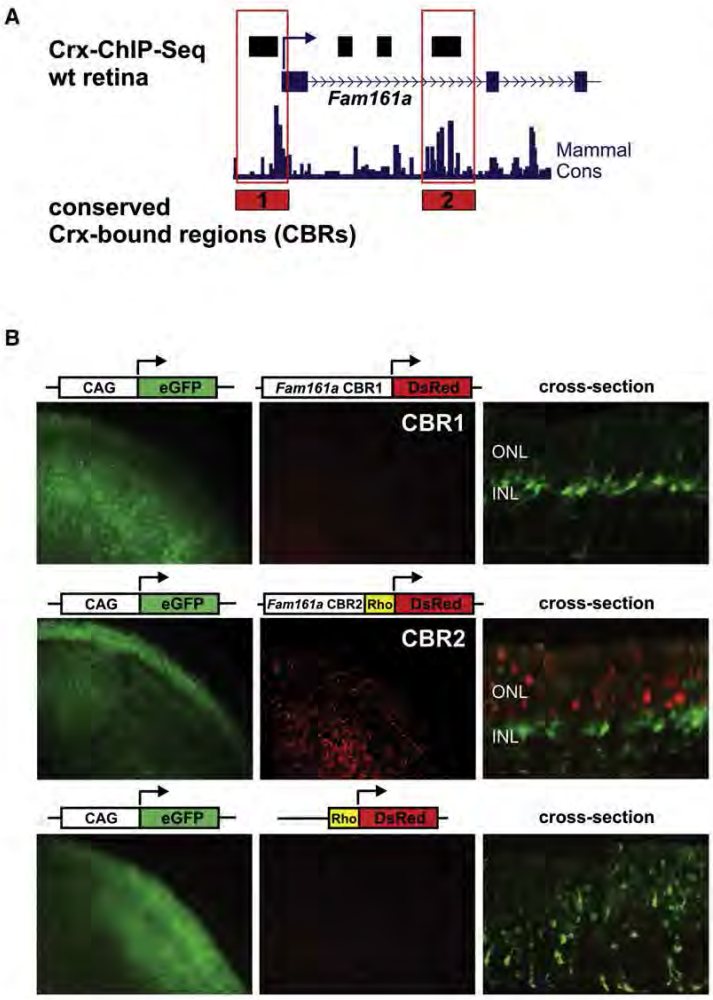


Figure 3. CBRs around *Fam161a* and CBR-Driven Reporter Expression in the Murine Retina
(A) *Fam161a* Crx-bound regions (CBR1 and CBR2, black boxes), as detected by ChIP-Seq experiments and phylogenetic conservation among mammals.
(B) Constructs carrying a constitutive CMV early enhancer-chicken β actin (CAG) promoter-eGFP fusion (electroporation control) and each of two *Fam161a* CBRs-DsRed fusions were coelectroporated in explanted P0 mouse retinas. CBR1 was cloned immediately upstream of DsRed, whereas CBR2 (intronic enhancer) was cloned upstream of a rhodopsin minimal promoter, which by itself is not active (lower panels). Fluorescence was measured after 8 days in whole mounts (left and central panels) and cross-sections (right panels). CBR1 alone does not seem to be active, whereas CBR2 can activate transcription in a Crx-dependent manner. CBR2 activity is detected in the outer nuclear layer, but not in the inner nuclear layer. This *cis*-regulatory element could therefore be responsible for the photoreceptor-specific expression of *Fam161a*.

between exons 3 and 4 defining additional 56 residues (Figure 1). Although in some species (including mouse) orthologous entries are incorrectly or only partly annotated, *FAM161A* is evolutionarily conserved in vertebrates (Figure S1) and has a paralog, named *FAM161B*. The putative protein products of *FAM161A* and *FAM161B* share approximately 30% identity over a 400 amino acid residue stretch. Data derived from in silico databases suggest that *FAM161B* is not expressed in the retina. The function of protein family 161 is currently unknown.

Quantitative RT-PCR of the major isoform of the FAM161A mRNA revealed elevated expression in the human retina and testis, whereas in other tissues FAM161A transcripts were present at lower abundance (Figure 2A). Considering that testis is an organ expressing almost 85% of all human genes at various levels,[7] biologically relevant *FAM161A* expression in humans may be confined to the retina. Although present at different levels, the two protein-coding human isoforms of *FAM161A* showed a similar expression profile (Figure 2B), and none of them could be defined as retina-specific, as is the case for other RP genes presenting with multiple splicing isoforms.[8–10]

In the mouse, Fam161a mRNA was highly expressed in the developing and adult retina and its presence was critically dependent on Crx (Figures 2C and 2D). As shown in Figure 3A, four CBRs are present within the *Fam161a* region, two of which are evolutionarily conserved and reside in the promoter (CBR1) and the first intron of the gene (CBR2). To analyze whether these CBRs represent photoreceptor-specific *cis*-regulatory elements, we performed electroporations of CBR-reporter fusions into explanted living-mouse postnatal day 0 (P0) retinas (Figure S2). Remarkably, although it is part of the *Fam161a* promoter region, CBR1 could not drive detectable expression in photoreceptors alone. In contrast, we found that CBR2 exerted strong photoreceptor-specific enhancer activity, as demonstrated by prominent DsRed fluorescence in the outer nuclear layer (ONL) (Figure 3B). These data establish *Fam161a* as a direct Crx target gene and highlight its retina-specific *cis*-regulatory regions.

We raised a polyclonal rabbit antibody against a conserved part of FAM161A (peptide: N-NPITGARVAQYE-C, 100% identity between human and mouse) and performed immunohistochemistry by using this antibody at a final dilution of 1:40 and biotinylated goat anti-rabbit IgGs (1:200). Specificity of the antibody for the target protein was tested by competition experiments with the peptide used for immunization, as well as by siRNA-mediated silencing of *FAM161A* in cultured cells (not shown). Detection was achieved with the ABC-HRP Kit (Vector, Burlingame, CA, USA) and the use of 3,3′-diaminobenzidine (DAB) (Dako, Carpinteria, CA, USA). Staining revealed a developmentally regulated expression of Fam161a within the mouse retina. Specifically, we observed a light, diffuse labeling in the neural basal layer at P0 (Figure 4A). At

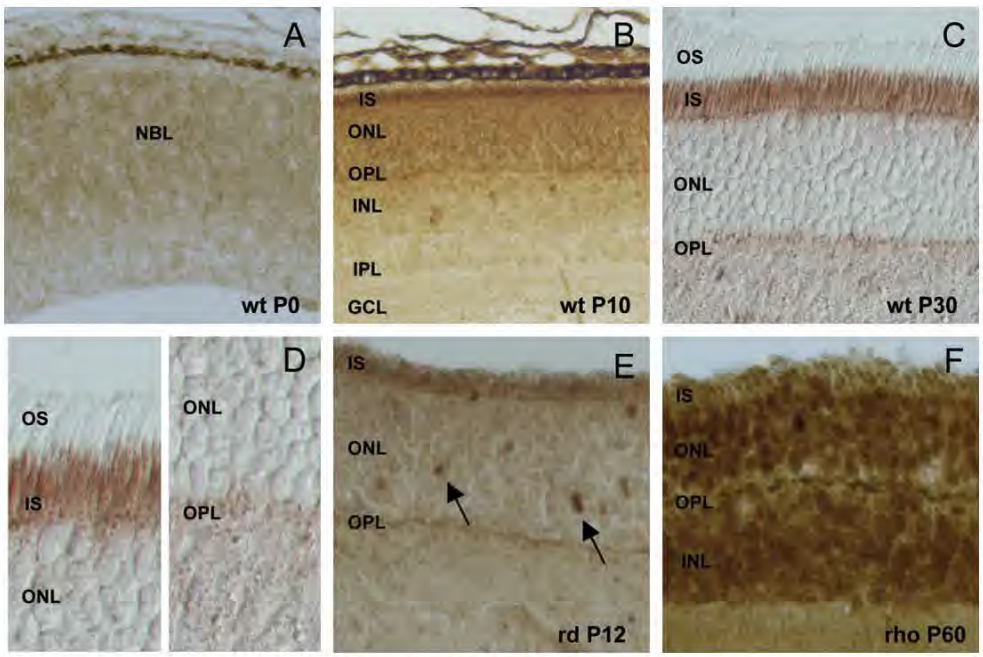


Figure 4. Fam161a Expression Pattern during Mouse Retinal Development and Retinal Degeneration
DAB immunostaining of Fam161a in murine retinas at various developmental stages, in wild-type animals (wt) and in models of retinal degeneration.
(A and B) Fam161a is minimally expressed at P0 (A), whereas at P10 it shows a marked presence in the outer nuclear layer and in the inner segments of photoreceptors (B). Note that the outer plexiform layer is also positive.
(C) In adult animals (P30), Fam161a reaches a well-defined localization in the inner segment of photoreceptors, as well as in the outer plexiform layer. It is completely absent from the outer segment of photoreceptors.
(D) Magnification of portions of (C), showing details of positive Fam161a staining.
(E and F) Fam161a localization in retinas from the *rd1* (rd) mouse at P12 and from the $Rho^{-/-}$ (rho) mouse at P60. Arrows indicate nuclear or perinuclear aggregates of Fam161a within photoreceptors. Magnification: A, B, and E, 200×; C and F, 400×; D, 800×. Abbreviations: NBL, neural basal layer; OS, outer segment of photoreceptors; IS, inner segment of photoreceptors; ONL, outer nuclear layer; OPL, outer plexiform layer; INL, inner nuclear layer; IPL, inner plexiform layer; GCL ganglion cell layer.

P10, a marked labeling was detected in all cell bodies of the outer nuclear layer (ONL), suggesting that Fam161a is expressed in both rods and cones. Note that the inner segments of the photoreceptors also stained positive for Fam161a (Figure 4B). In the adult retina (P30), Fam161a was present in the inner segments but absent in the outer segments of the photoreceptors (Figures 4C and 4D). It also showed small, strongly labeled structures at the level of the outer plexiform layer (OPL). However, further analysis is needed to identify in which subcellular structure of the OPL Fam161a is exactly localized. Several studies have shown that instability of some photoreceptor proteins may play an important role in the process of cell death even if the protein in question is only indirectly linked to the mutated gene targeted in a defined mouse model. For instance, mislocalization of S and M opsins is thought to play a role in cone degeneration in the *Rpe65*-deficient mouse (*RPE65* [MIM 180069]).[11] We therefore analyzed the *rd1* mouse to explore whether Fam161a expression or its subcellular localization may be affected in the early process of retinal degeneration. *Rd1* mice bear homozygous mutations in the gene encoding the beta subunit of the rod cGMP-phosphodiesterase (*PDE6B* [MIM 180072]) and show a rapid degeneration of rod photoreceptors starting at P10–11. Interestingly, the Fam161a expression pattern was markedly affected in the *rd1* mouse retina; several cell bodies and nuclei showed only dot-like positive signals (Figure 4E). Fam161a localization was also abnormal in rhodopsin $^{/}$ mice (*RHO* [MIM 180380]), which have a slower degeneration process. Specifically, Fam161a expression in adult *Rho* $^{/}$ retinas was very similar to that of immature wild-type retinas (P10), with a uniform and marked presence in all cells of the ONL (Figure 4F). Mislocalization of Fam161a in the retina of these genetically manipulated animals can be part of the pathology associated with the primary gene defect or a nonspecific consequence of the degenerative process in the retina. In particular, these observations may suggest a link between Fam161a and proteins involved in the phototransduction cascade. During adulthood, Fam161a is localized in the inner segment, where both Pde6b and rhodopsin migrate through to reach the outer segment. Because the absence of each of these two latter proteins results in maintaining Fam161a in the cell body,

Fam161a trafficking may be dependent on certain phototransduction proteins. However, this hypothesis is highly speculative and needs further experimental verification.

Analysis of a second consanguineous family from India has suggested refined boundaries of the putative *RP28*-chromosomal region and defined a smaller region of overlap, of ~3.5 Mb.[12] DNA samples from the original *RP28* family and the second Indian family showed different haplotypes, excluding a *FAM161A/RP28*-founder mutation segregating in both families.[5] Furthermore, linkage data do not support the assumption that *FAM161A* is implicated in the retinal dystrophy in the second family.

In this report, we show that mutations in *FAM161A* cause the *RP28*-associated form of arRP and that *FAM161A* mutations are present in the population of both India and Europe. The measured prevalence of *FAM161A*-associated RP in our cohort of German patients is in the range of 2%–3%, comparable to that of other arRP genes. Our data also indicate that these mutations are likely functional null alleles, making *FAM161A* an ideal target for constructing knockout animal models and for the option of gene replacement in the quest of seeking treatment for this disease.

Supplemental Data

Supplemental Data include two figures and five tables and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

This work was supported by the Swiss National Science Foundation (grant no. 320030-121929, to C.R.), the Gebert Rüf Foundation (Rare Diseases - New Technologies grant, to C.R.), the German Research Foundation (FOR1075 with grants to B.H.F.W., H.S., and T.L.), and Pro Retina (to B.H.F.W., T.L., and H.S.). N.M. received financial support from the DAAD aimed to promote the cooperation between the Universities of Hamburg and Belgrade. The authors would also like to thank Anne Maillard-Menoud for technical support, as well as scientists at Genotypic Technology Pvt. Ltd., Bangalore, India, and at Fasteris SA, Plan-les-Ouates, Switzerland.

Received: June 15, 2010
Revised: July 23, 2010
Accepted: July 27, 2010
Published online: August 12, 2010

Web Resources

The URLs for data presented herein are as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/

Retinal Information Network (RetNet), http://www.sph.uth.tmc.edu/retnet/

References

1. Hartong, D.T., Berson, E.L., and Dryja, T.P. (2006). Retinitis pigmentosa. Lancet *368*, 1795–1809.
2. Koenekoop, R.K. (2009). Why do cone photoreceptors die in rod-specific forms of retinal degenerations? Ophthalmic Genet. *30*, 152–154.
3. Gu, S., Kumaramanickavel, G., Srikumari, C.R., Denton, M.J., and Gal, A. (1999). Autosomal recessive retinitis pigmentosa locus RP28 maps between D2S1337 and D2S286 on chromosome 2p11-p15 in an Indian family. J. Med. Genet. *36*, 705–707.
4. Corbo, J.C., Lawrence, K.A., Karlstetter, M., Myers, C.A., Abdelaziz, M., Dirkes, W., Weigelt, K., Seifert, M., Benes, V., Fritsche, L.G., et al. (2010). CRX ChIP-seq reveals the cis-regulatory architecture of mouse photoreceptors. Genome Res., in press.
5. Rio Frio, T., Panek, S., Iseli, C., Di Gioia, S.A., Kumar, A., Gal, A., and Rivolta, C. (2009). Ultra high throughput sequencing excludes MDH1 as candidate gene for RP28-linked retinitis pigmentosa. Mol. Vis. *15*, 2627–2633.
6. Hentze, M.W., and Kulozik, A.E. (1999). A perfect message: RNA surveillance and nonsense-mediated decay. Cell *96*, 307–310.
7. Ramsköld, D., Wang, E.T., Burge, C.B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput. Biol. *5*, e1000598.
8. Bowne, S.J., Liu, Q., Sullivan, L.S., Zhu, J., Spellicy, C.J., Rickman, C.B., Pierce, E.A., and Daiger, S.P. (2006). Why do mutations in the ubiquitously expressed housekeeping gene IMPDH1 cause retina-specific photoreceptor degeneration? Invest. Ophthalmol. Vis. Sci. *47*, 3754–3765.
9. van Wijk, E., Pennings, R.J., te Brinke, H., Claassen, A., Yntema, H.G., Hoefsloot, L.H., Cremers, F.P., Cremers, C.W., and Kremer, H. (2004). Identification of 51 novel exons of the Usher syndrome type 2A (USH2A) gene that encode multiple conserved functional domains and that are mutated in patients with Usher syndrome type II. Am. J. Hum. Genet. *74*, 738–744.
10. Neidhardt, J., Glaus, E., Barthelmes, D., Zeitz, C., Fleischhauer, J., and Berger, W. (2007). Identification and characterization of a novel RPGR isoform in human retina. Hum. Mutat. *28*, 797–807.
11. Rohrer, B., Lohr, H.R., Humphries, P., Redmond, T.M., Seeliger, M.W., and Crouch, R.K. (2005). Cone opsin mislocalization in Rpe65-/- mice: a defect that can be corrected by 11-cis retinal. Invest. Ophthalmol. Vis. Sci. *46*, 3876–3882.
12. Kumar, A., Shetty, J., Kumar, B., and Blanton, S.H. (2004). Confirmation of linkage and refinement of the RP28 locus for autosomal recessive retinitis pigmentosa on chromosome 2p14-p15 in an Indian family. Mol. Vis. *10*, 399–402.

I.3 SUPPLEMENTARY MATERIALS

Figure S1. Multiple Alignments of FAM161A in Eight Vertebrate Species.
Since current annotations of FAM161A orthologues are generally poor, non-human entries shown derive from the conceptual translation of full transcripts, partial cDNA clones, ESTs, or "in silico spliced" sequences from genomic DNA, all available from public databases. Conserved residues are blue shaded.

CBR1 (Promoter, 932 bp)

AGGTTCTGTGGAAGGTGTGCCGCATAACGTGGAAAGTACTGGAACCTTAGGGTGGGGACAGGGGCGCTAGCT
GGGGTTCTAAGTAAAGTGTTATGCTGCAGGTGGCTGAGAGTTCACTGAAAAAGCATCCAAAGTGGCCAAAAA
CCTGAGGCCGATCAGGAGGTTGGCAGGGCTGAGCTGGGAGGGTGTGGTCCTTGTAGGCTCTCGCCCCCTC**TA**
ATCTGCCGCTAAGCTAGCTGCAGAACCGCTAGACAAGCTACCACATCTCACTCACTGGTCTGAGTGACAGAC
ACTTCCTATGTCTAGTGTCTTACTGGTCAGTCGAGAAAAAAGAGGGTGGGACAAATAAGTATTCTCCTTCTG
ATTGGTCAACCCGCTGGTTCTTGTCA**ATTA**GAGATGAACTGGGTTCTCCGCCATGTTGAATAGCGCGCTGCA
AAGCCTGTTCTTTTCTTGGGATGGTCACCCAACACCCGGTCCTTCCTGTCCCTCTGTGCCCTAATATGTTGA
TCCAGCTAGTTAAAACGCGCTTTCTGCTTTAAGTCCGTATACCGCGAGAAGTTCTCGTTTAGGTAGTTAGAC
AGAAATAAAGCTAGAAACCAGAGAAAAGCTCAGTCCCTTGCCGGAGACCTGGGTGTCATTCAAGTGAATCAG
TAATTGTTCCGGCGCTGGTGAGACTCCCATTGAGCA**ATTA**GCTTGAAGATTGCGTTTTACTTTAGCTTGCTG
TGGTAGTCGTTTGTGTAGCTTCAGACAGTTATGCTTCTAGAAACCGCAGGCTTGTATCAGAAACCATC**ATTA**
GACACTAGCTGGACTCGGGGTTCGCATTTGTGACAGTGAAAAAGGACGTTGCGTGCGTTATGGTCCGTCGCC
ACCTCCTGTAGCGTCCCTAGTAACCGAGCCCCTAGTAACCGCTCGCCCCAAGCCAGGATCTGTGAGCT

CBR1 cloning into "No basal-dsRed"

***Eco*RI / *Kpn*I**
forward tccccg**GAATTC**AGGTTCTGTGGAAGGTGTGC
reverse ttcgg**GGTACC**AGCTCACAGATCCTGGCTTG

CBR2 (Intron 1, 400 bp)

GTTCAGCATTGAAAAGAACATAAATTAGTAAAAGCTCCAGAATGTCAGATAAACTTTGTAGACTACAATACA
AACAGTTCCTTATGGTGCTGACACAAGGTAAGCCTCAGGTGCTGGGAGAGCTGATACATCTGTAAGTGGGTC
AGGCCACTCTCGGGAAG**ATTA**GCTTCTTCTGTAGTGGCACTCACAACACGGAAAGCTTAGTATCAAGGGGGA
GGACAGGACCAGCATTGCACATTGTGTTTTGCAAACCGGACAGTCTTGGGGCAGGGGGCGAAGGTGAAGCGA
GAATTGGGAGTTATTTCCTAGCCATTCATCTGTTTTAGGTTAAAATACAAGCAAACTAGGAGGGCCTATTTG
GAAAGTTCCGCTCGT**ATTA**ACAGTGGACTGGGCACTGTGC

CBR2 cloning into "Rho basal-dsRed"

***Eco*RV / *Kpn*I**
forward tcagtcacgat**GATATC**GTTCAGCATTGAAAAGAACATAAA
reverse ttcgg**GGTACC**GCACAGTGCCCAGTCCAC

Figure S2. Sequences and Primers for Cloning of *Fam161a* CBR1 and CBR2 into "No basal-dsRed" and "Rho basal-dsRed"

Crx consensus motifs are printed in bold and are underlined.

Table S1. Non-Annotated DNA Variants Detected Homozygously by Exon Sequence Capturing and UHTs for the 188 Genes Present in the *RP28* Interval

Human ENSEMBL ID	Gene Symbol	DNA Variant	Position	Frequency (%)	Predicted aa change
ENSG00000170264	*FAM161A*	G/A	c.685	100	Arg>End
ENSG00000124383	*MPHOSPH10*	T/A	c.82	89.2	Phe>Ile
ENSG00000144040	*SFXN5*	G/A	c.364	100	Gly>Ser
ENSG00000173209	*AHSA2*	T/A	c.393	90.9	Val>Asp
ENSG00000143971	*ETAA1*	C/A	c.2299	100	Lys>Gln
ENSG00000173209	*AHSA2*	G/T	c.402	100	Lys>Asn
ENSG00000169564	*PCBP1*	G/A	c.735	81.9	no (Gln=)
ENSG00000163170	*BOLA3*	C/T	c.23	100	no (Ala=)

Entries are ranked according to the pathogenicity score of the amino acid (aa) residue change, from most damaging (top) to neutral (bottom).

Table S2. Crx-Driven ChIP-Seq Reads in Murine Retinas

Human ENSEMBL ID	Human Gene Symbol	Mouse ENSEMBL ID	Mouse Gene Symbol	Crx-ChIP-Seq reads wt retinas
ENSG00000170264	*FAM161A*	ENSMUSG00000049811	*Fam161a*	205
ENSG00000119862	*HSPC159*	ENSMUSG00000042363	*1110067D22Rik*	124
ENSG00000170340	*B3GNT2*	ENSMUSG00000051650	*B3gnt2*	112
ENSG00000197329	*PELI1*	ENSMUSG00000020134	*Peli1*	104
ENSG00000143995	*MEIS1*	ENSMUSG00000020160	*Meis1*	77
ENSG00000082898	*XPO1*	ENSMUSG00000020290	*Xpo1*	66
ENSG00000186889	*TMEM17*	ENSMUSG00000049904	*Tmem17*	40
ENSG00000143951	*C2orf86*	ENSMUSG00000020319	*AV249152*	38
ENSG00000115484	*CCT4*	ENSMUSG00000007739	*Cct4*	37
ENSG00000115504	*EHBP1*	ENSMUSG00000042302	*Ehbp1*	33
ENSG00000173163	*COMMD1*	ENSMUSG00000051355	*Commd1*	25
ENSG00000119865	*CNRIP1*	ENSMUSG00000044629	*Cnrip1*	25
ENSG00000115946	*PNO1*	ENSMUSG00000020116	*Pno1*	21
ENSG00000204923	*FBXO48*	ENSMUSG00000044966	*Fbxo48*	19
ENSG00000143971	*ETAA1*	ENSMUSG00000016984	*Etaa1*	10

The transcription factor Crx (which is expressed in both rods and cones) is strongly bound to the *Fam161a* genomic region, as reflected by 205 ChIP-seq reads. The number of reads is proportional to the amount of Crx-protein precipitated from the region and represents approximately the affinity of Crx to the target DNA.

Table S3. Haplotypes of Three Apparently Unrelated German Patients with Retinitis Pigmentosa and a Homozygous p.Arg437X Mutation

SNP ID	Location	arRP173	arRP323	M09-0352
rs4270331:G>T	Exon 1	G/G	G/G	G/G
rs11125896:G>A	Intron 1	G/G	G/G	G/G
rs11125895:A>G	Exon 2	G/G	G/G	G/G
rs17513722:A>G	Exon 3	A/A	A/A	A/A
rs4672457:C>T	Exon 3	C/C	C/C	C/C
rs35746699:GCA>del	3' UTR	GCA/GCA	GCA/GCA	GCA/GCA

Table S4. Rare Sequence Variants of Uncertain Pathogenicity Detected in *FAM161*

Patient ID	Ethnic origin	Exon	Nucleotide change	Allele status	Protein variant	Evolutionary conserved aa	Control alleles (mut/wt)
M07-0321	Turkish	2	c.197 C>T	heterozygous	p.Thr66Ile	no	2/366
M08-0522	German	2	c.228 G>A	heterozygous	p.Pro76=		0/358
M09-0413	German	3	c.613 G>C	heterozygous	p.Asp205His	yes	0/400
M09-0502	German	3	c.1085 G>T	heterozygous	p.Arg362Leu	yes	1/400
M08-0312	German	3	c.1133 T>G	heterozygous	p.Leu378Arg	yes	0/400
arRP303	German	3	c.1133 T>C	heterozygous	p.Leu378Pro	yes	0/400
M08-0586	German	3	c.1153 C>G	heterozygous	p.Gln385Glu	no	2/400
M07-0320	German	3	c.1153 C>G	heterozygous	p.Gln385Glu	no	2/400
arRP305	German	3	c.1153 C>G	heterozygous	p.Gln385Glu	no	2/400

Table S5. Non-Pathogenic Sequence Variants Detected in *FAM161*

Exon / Intron	Nucleotide change*	Protein variant	SNP ID
E1	c.165G>T	p.Ala55=	rs4270331
E2	c.321A>G	p.Ile107Met	rs11125895
E3	c.706A>G	p.Ile236Val	rs17513722
E3	c.1212C>T	p.Cys404=	rs4672457
IVS3	c.1583+705G>A	none	rs17513666
IVS3/E3**	c.1583+797G>A	none/p.Leu555=	rs6545910

*Numbering relative to entry NM_032180.
**Depending on the isoform, this variant is either localized in intron 3 or in exon 3a (p.Leu555).

I.4 PAPER

FAM161A, ASSOCIATED WITH RETINITIS PIGMENTOSA, IS A COMPONENT OF THE CILIA-BASAL BODY COMPLEX AND INTERACTS WITH PROTEINS INVOLVED IN CILIOPATHIES.

Silvio Alessandro Di Gioia[1], Stef J.F Letteboer[3], Corinne Kostic[2], Dikla Bandah-Rozenfeld[4], Lissette Hetterschijt[3], Dror Sharon[4], Yvan Arsenijevic[2], Ronald Roepman[3], and Carlo Rivolta[1].

[1]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland, [2] Unit of Gene Therapy and Stem Cell Biology, Jules-Gonin Eye Hospital, University of Lausanne, Lausanne, Switzerland, [3] Department of Human Genetics and Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands and [4] Department of Ophthalmology, Hadassah-Hebrew University Medical Center, Jerusalem, Israel

Abstract

Retinitis pigmentosa (RP) is a retinal degenerative disease characterized by the progressive loss of photoreceptors. We have previously demonstrated that RP can be caused by recessive mutations in the human FAM161A gene, encoding a protein with unknown function that contains a conserved region shared only with a distant paralog, FAM161B. In this study, we show that FAM161A localizes at the base of the photoreceptor connecting cilium in human, mouse and rat. Furthermore, it is also present at the ciliary basal body in ciliated mammalian cells, both in native conditions and upon the expression of recombinant tagged proteins. Yeast two-hybrid analysis of binary interactions between FAM161A and an array of ciliary and ciliopathy-associated proteins reveals direct interaction with lebercilin, CEP290, OFD1 and SDCCAG8, all involved in hereditary retinal degeneration. These interactions are mediated by the C-terminal moiety of FAM161A, as demonstrated by pull-down experiments in cultured cell lines and in bovine retinal extracts. As other ciliary proteins, FAM161A can also interact with the microtubules and organize itself into microtubule-dependent intracellular networks. Moreover, small interfering RNA-mediated depletion of FAM161A transcripts in cultured cells causes the reduction in assembled primary cilia. Taken together, these data indicate that FAM161A-associated RP can be considered as a novel retinal ciliopathy and that its molecular pathogenesis may be related to other ciliopathies.

In this paper I performed the majority of the experiments, with the only exception being the FAM161A staining in human tissue, which was performed by Bandah-Rozenfeld from the Sharon group. I performed the yeast-two-hybrid experiments and some immunostaining experiments in Nijmegen under the supervision of Ronald Roepman and his team.

Human Molecular Genetics, 2012, Vol. 21, No. 23 **5174–5184**
doi:10.1093/hmg/dds368
Advance Access published on September 1, 2012

FAM161A, associated with retinitis pigmentosa, is a component of the cilia-basal body complex and interacts with proteins involved in ciliopathies

Silvio Alessandro Di Gioia[1], Stef J.F. Letteboer[3], Corinne Kostic[2], Dikla Bandah-Rozenfeld[4], Lisette Hetterschijt[3], Dror Sharon[4], Yvan Arsenijevic[2], Ronald Roepman[3,†] and Carlo Rivolta[1,*,†]

[1]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland, [2]Unit of Gene Therapy and Stem Cell Biology, Jules-Gonin Eye Hospital, University of Lausanne, Lausanne, Switzerland, [3]Department of Human Genetics and Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands and [4]Department of Ophthalmology, Hadassah-Hebrew University Medical Center, Jerusalem, Israel

Received July 24, 2012; Revised and Accepted August 27, 2012

Retinitis pigmentosa (RP) is a retinal degenerative disease characterized by the progressive loss of photoreceptors. We have previously demonstrated that RP can be caused by recessive mutations in the human *FAM161A* gene, encoding a protein with unknown function that contains a conserved region shared only with a distant paralog, *FAM161B*. In this study, we show that FAM161A localizes at the base of the photoreceptor connecting cilium in human, mouse and rat. Furthermore, it is also present at the ciliary basal body in ciliated mammalian cells, both in native conditions and upon the expression of recombinant tagged proteins. Yeast two-hybrid analysis of binary interactions between FAM161A and an array of ciliary and ciliopathy-associated proteins reveals direct interaction with lebercilin, CEP290, OFD1 and SDCCAG8, all involved in hereditary retinal degeneration. These interactions are mediated by the C-terminal moiety of FAM161A, as demonstrated by pull-down experiments in cultured cell lines and in bovine retinal extracts. As other ciliary proteins, FAM161A can also interact with the microtubules and organize itself into microtubule-dependent intracellular networks. Moreover, small interfering RNA-mediated depletion of FAM161A transcripts in cultured cells causes the reduction in assembled primary cilia. Taken together, these data indicate that *FAM161A*-associated RP can be considered as a novel retinal ciliopathy and that its molecular pathogenesis may be related to other ciliopathies.

INTRODUCTION

Hereditary retinal degenerations (HRDs) consist of a group of diseases that are clinically and genetically heterogeneous. Despite the many differences in progression rates and in severity characterizing each specific form, all patients with HRD lose vision because of the degeneration of rod and cone photoreceptors, and eventually of other retinal cells (1). Retinitis pigmentosa (RP) is the most common form of HRD, affecting more than one million people worldwide (2). Patients with RP usually present with night blindness due to malfunctioning rod photoreceptors that turns progressively into the impairment of cone photoreceptors and thus daytime vision as well, from the periphery to the center of the visual field (1). Later in life, many patients become legally blind. Genetically, RP can be transmitted as an autosomal dominant, autosomal recessive or X-linked trait; in rare cases, digenic inheritance has also been observed (3,4). Out of the 182 identified genes involved in HRD, 43 have been also associated with non-syndromic RP (https://sph.uth.tmc.edu/retnet/). Interestingly, RP genes display a broad range of expression: some encode proteins that are restricted to rods only, while others are expressed in other retinal cell types or are even ubiquitously active across all tissues and organs of our body (4). Functional categories

*To whom correspondence should be addressed at: Department of Medical Genetics, University of Lausanne, Rue du Bugnon 27, 1005 Lausanne, Switzerland. Tel: +41 216925451; Fax: +41 216925455; Email: carlo.rivolta@unil.ch
†Equal contribution.

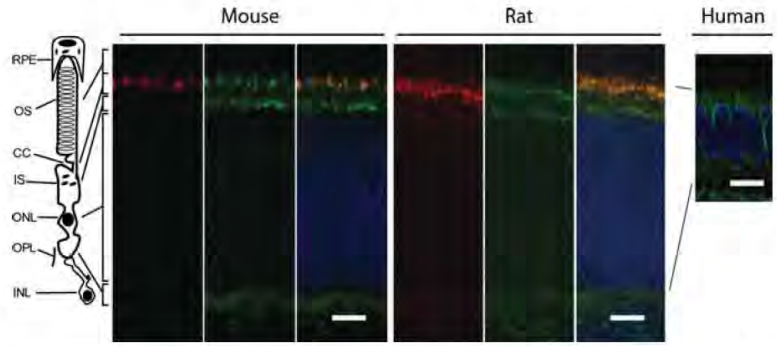


Figure 1. Localization of FAM161A in adult mouse, rat and human retina slices. Murine and rat sections were stained with anti-pan centrin (red), a ciliary marker, and FAM161A (green). The merged images show partial co-localization between the two signals. Human sections were stained with anti FAM161A antibody only (green). Nuclei are stained in DAPI (blue). Scale bars: 10 μm for mouse and rat, 50 μm for human. OS, outer segment; CC, connecting cilium; IS, inner segment; ONL, outer nuclear layer; OPL, outer plexiform layer; INL, inner nuclear layer.

of RP genes are also very diverse. However, many of them display a clear relationship with retinal biochemistry or have structural roles in retinal cells (5).

In particular, a number of HRD proteins have been found to have important roles in development, maintenance or functioning of the photoreceptor immotile cilium (6). The photoreceptor connecting cilium (CC) is a particular structure of both photoreceptor types that is located between their inner and outer segments (IS and OS). In addition to its structural role, the CC is also involved in the transport of molecules across these two compartments by intraflagellar transport (6). New evidences also demonstrate its role in cell signaling and signal transmission (7). Disruption of gene-encoding proteins that are structurally and functionally connected to primary cilia and their cellular anchoring/nucleation point, the centrosome-derived basal body (BB), causes a broad class of diseases with pleiotropic phenotypes, all classified as ciliopathies (8). Retinal ciliopathies (RCs) include all ciliopathies that cause HRD, exclusively or in conjunction with other systemic defects. For instance, typical RC genes are: the RP-1 gene (*RP1*) (9), RP GTPase regulator (*RPGR*) (10), the RPGR-interacting protein 1 gene (*RPGRIP1*), RPGRIP1-like (*RPGRIP1L*) (11), Leber congenital amaurosis 5 (*LCA5*) (12), centrosomal protein 290 (*CEP290*) (13), Orofacialdigital 1 (*OFD1*) (14,15) and serologically defined colon cancer antigen 8 (*SDCCAG8*) (16,17). *RP1* is associated with autosomal dominant and recessive RP and *RPGR* is associated with X-linked RP, whereas *RPGRIP1*, *LCA5* and *CEP290* are associated with Leber congenital amaurosis (congenital retinal blindness) among other diseases. Mutations in *OFD1* and *SDCCAG8* cause Joubert syndrome and RP and nephronophthisis with HRD or Bardet–Biedl syndrome (BBS), respectively.

Recently, we identified *FAM161A* as the causative gene for *RP28*-associated RP and one of the major causes of RP in Israel and the Palestinian territories (18,19). The *RP28* locus was first mapped on chromosome 2p in 1999, through the genetic analysis of a consanguineous family from India segregating recessive RP (20). In all patients analyzed, mutations in *FAM161A* lead to null alleles (18,19). FAM161A expression is driven by the photoreceptor-specific transcription factor CRX (18), and its presence is detected mostly in testis and retina. In this latter tissue, standard immunohistochemistry analysis revealed that FAM161A localized at the level of the IS of photoreceptors and less markedly at the level of the outer plexiform layer (OPL) of the retina (18,19), with a localization pattern similar to USH2A (21). The function of FAM161A or other members of the protein family it belongs to (family FAM161) are unknown.

In this work, we demonstrate that FAM161A is an RC protein that is part of the CC complex and interacts with other proteins associated with RCs. Furthermore, we show that its localization at the base of the cilium is not limited to the photoreceptors, but is also conserved in other ciliated cell types and tissues.

RESULTS

FAM161A is present at the base of the cilium of photoreceptors and in renal cells

To investigate the localization of FAM161A within the mammalian retina, we performed indirect immunofluorescence of fresh, unfixed cryosections of adult mouse and rat retinas, as well as paraffin-fixed human retinas with three different primary antibodies. These procedures resulted in specific staining in all species (Fig. 1). FAM161A localized at the base of the photoreceptor cilia, as well as within their IS, where the Golgi, the endoplasmic reticulum and other organelles are found. As previously shown (18), a weak positive staining was detected in the OPL of the retina. Mechanical dissociation of the murine retina, allowing photoreceptor isolation and the preservation of their OS, CC and part of the IS (22) revealed specific FAM161A localization in the region between the IS and the OS (Fig. 2). This region corresponded to the CC of photoreceptors.

To define more precisely the distribution of FAM161A within the cilium, we performed immunofluorescence microscopy by using antibodies against FAM161A and typical ciliary markers. Co-staining with antibodies against RP1, staining the ciliary axoneme (9), revealed that within the same ciliary structure, FAM161A and RP1 were contiguous but not overlapping (Fig. 3A). RPGRIP1L is a marker of the ciliary BB (11), while pan-centrin localizes within the whole CC (23). Co-staining of FAM161A and each of these two markers

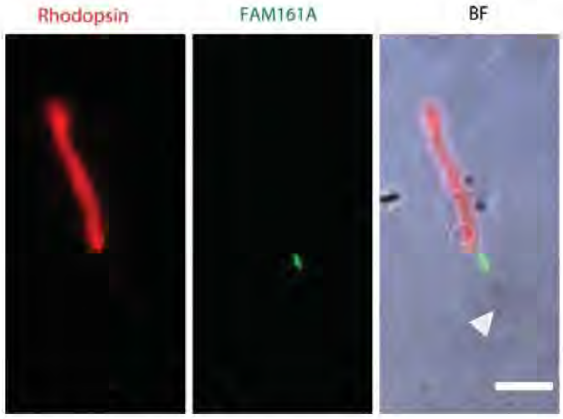


Figure 2. FAM161A localizes at the level of the CC in dissociated murine photoreceptors. Dissociated photoreceptors were co-stained with antibodies against rhodopsin (red) and FAM161A (green). During the dissociation process only part of the IS of photoreceptors is preserved (white arrowhead). BF, bright field. Scale bar: 10 μm.

revealed partial co-localization in both cases, according to a pattern indicating its presence between the BB and the CC. Specifically, immunofluorescence with anti-FAM161A and anti-RPGRIP1L antibodies revealed that FAM161A was not present just in the BB, but continued also in the lower part of the CC (Fig. 3B). Simultaneous staining of FAM161A and pan-centrin confirmed the non-complete presence of FAM161A within the whole CC and was indicative of, again, its presence within the BB (Fig. 3C).

Although FAM161A mRNA expression seems to be predominant in retina and testis, it is not restricted to these tissues (18,19). Staining of FAM161A in the rat kidney revealed again its presence at the level of cellular cilia. Similar to the results obtained in the retina, FAM161A was present at the level of the BB, as demonstrated by co-staining with RPGRIP1L (Supplementary Material, Fig. S1A), whereas it partially overlapped with the signal produced by anti-polyglutamylated tubulin antibody GT335, a marker of the CC and the BB (Supplementary Material, Fig. S1B).

FAM161A localizes at the level of the BB in ciliated cell lines

Many standard laboratory cell lines develop cilia. To identify a cellular model suitable for studying FAM161A and ciliogenesis, we assessed 10 human cell lines for *FAM161A* mRNA expression and corresponding protein detectability (data not shown). We found in retinal pigment epithelium-derived cells hTERT-RPE-1 and ARPE-19 the best models for this purpose. Co-staining with anti-FAM161A and GT335 antibodies in hTERT-RPE-1 cells showed partial overlap with the BB and the daughter centriole (Fig. 4A). Similar results were also obtained in ARPE-19 cells (not shown). In addition, co-staining between FAM161A and RPGRIP1L showed perfect co-localization (Fig. 4B). Taken together, these data indicate that FAM161A is present in the BB/centrosomes of ciliated cells of different origins. When we expressed heterologous fusion FAM161A constructs in these same cell lines, we detected it at the same location, as revealed by using an anti-FLAG antibody (Fig. 4C and D; data not shown for ARPE-19).

Recombinant expression of *FAM161A* in hTERT-RPE-1 and ARPE-19 cultures allowed some cells to significantly overexpress the FAM161A protein. In these cells, FAM161A decorated the entire microtubule network, with a pattern similar to that observed upon the overexpression of ciliary proteins such as lebercilin (12) or CEP170 (24). We therefore tested FAM161A overexpression in non-ciliated cells such as HeLa or mouse photoreceptor-derived 661W cells. Heterologous FAM161A assembled indeed in a web-like structure that partly overlapped with the cytoplasmic microtubule network, revealed by staining with anti-acetylated tubulin (Fig. 5A). Supplementation with the microtubule depolymerizing drug nocodazole in the culture medium completely abolished this FAM161A organization, indicating its microtubule dependence (Fig. 5B).

FAM161A interacts with other ciliary and centrosomal proteins

Given the specific localization of FAM161A, we decided to ascertain its possible interaction with components of the ciliary proteome that are relevant to human diseases. Based on UNIPROT *in silico* analysis, FAM161A is predicted to have three coiled-coil domains, one at the N-terminus and two located in the well-conserved domain (provisionally called UPF0564 domain) located within the C-terminal part of the protein. We performed a rapid and efficient interaction assay in yeast by using three *FAM161A* constructs as baits to simultaneously screen a library of 45 unique genes associated with various human ciliopathies or involved in ciliary functions. In addition to a FAM161A full-length form, we designed a construct that included its first 217 codons (FAM161A-N-term bait) and another construct encompassing the remainder of the protein and containing the UPF0564 domain (FAM161A-C-term bait, codons 211–660) (Supplementary Material, Fig. S2A). Several bait-prey combinations induced yeast growth, pinpointing potential interactions (Supplementary Material, Table S1). However, colonies were observed only when the full FAM161A or FAM161A-C-term sequences were used to assess a potential interaction, while FAM161A-N-term produced no detectable growth. Among all tested interactions, the strongest signals were observed between both the FAM161A full length and FAM161A-C-term preys and baits expressing SDCCAG8, OFD1, CEP290 and lebercilin, or parts thereof (Fig. 6).

To validate the interactions detected by yeast assays, we performed pull-down experiments in bovine retinal extracts, by using recombinant constructs containing the same parts of FAM161A (full form, N-term and C-term) used before but fused with the glutathione S-transferase (GST) protein (Supplementary Material, Fig. S2B). Following an overnight incubation of these chimeric proteins and retinal lysates, we could efficiently co-purify CEP290 and lebercilin, the only two proteins for which reliable antibodies were accessible to us. Again, positive signals were detected only for the full-length FAM161A protein and the construct presenting its

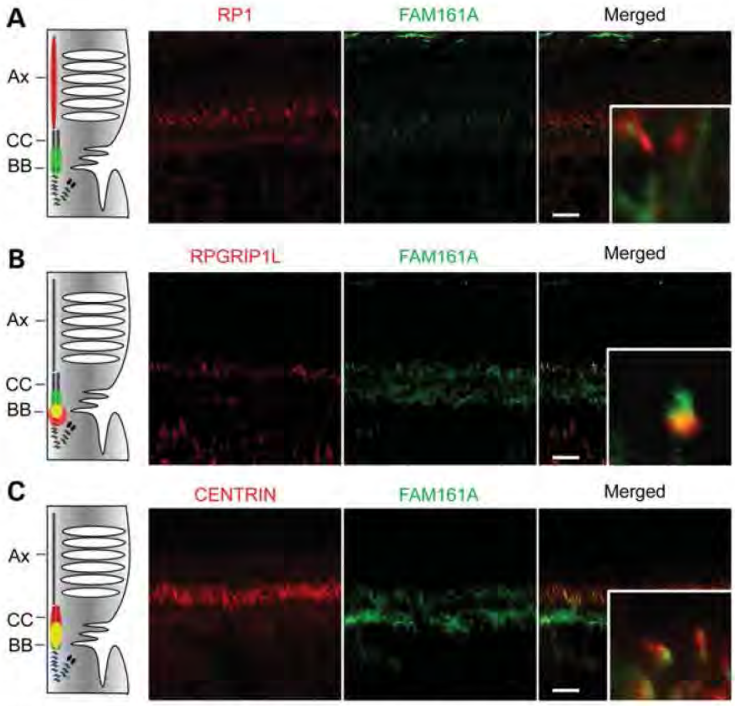


Figure 3. Co-staining of FAM161A in rat retina sections with different ciliary markers. Retinal sections were stained with an anti-FAM161A antibody (green) and antibodies against axonemal or ciliary markers (red). Co-staining with RP1, and axonemal marker, reveals no overlap (**A**). Conversely, co-localization is observed with RPGRIP1L (a marker for the BB) and centrin (a marker for the whole CC), according to specific patterns (**B** and **C**). The insets are magnifications of the merged image (scale bars: 10 μm). The cartoons at the left side of each panel summarize the localization of the signal. Ax, Axoneme; CC, connecting cilium; BB, basal body.

C-term part (Fig. 7A). It should be noted that the signal corresponding to lebercilin in co-purified extracts ran at a lower position than its counterpart in total retinal lysates. This was likely caused by specific proteolytic cleavage occurring during the overnight incubation with sepharose beads in the pull-down experiments. In support of this hypothesis, the same difference in size between lebercilin from total lysates and lebercilin after overnight incubation was consistently observed by using two other antibodies against lebercilin, targeting different epitopes (not shown).

We then overexpressed recombinant lebercilin, SDCCAG8 and OFD1 fused to a hemagglutinin (HA) tag together with GST-fused FAM161A recombinant proteins (FAM161A-GST, GST-FAM161A-N-term and GST-FAM161A-C-term) in monkey kidney-derived COS-7 cells. In GST-driven pull-down assays, FAM161A-GST and GST-FAM161A-C-term could capture lebercilin, revealed with an anti-HA antibody from the lysate of transfected COS-7 cells, whereas GST alone or GST-FAM161A-N-term could not (Fig. 7A). Furthermore, FAM161A-GST, GST-FAM161A-N-term and GST-FAM 161A-C-term could pull-down 3xHA-SDCCAG8, although the signal was significantly lower than observed with lebercilin (Fig. 7A). In the reciprocal experiments, GST-lebercilin could pull down both FLAG-tagged full-length and C-terminal FAM161A, but not its N-terminal fragment (Fig. 7B). Conversely, GST-SDCCAG8 could capture only full-length 3xFLAG-FAM161A (Fig. 7B). No signals were observed in pull-down experiments performed with OFD1 constructs, probably because of the very low level of expression of the plasmid carrying the relevant sequence in COS-7 (data not shown).

These same results were confirmed by microscopy. Following co-expression of FLAG-FAM161A and any of HA-lebercilin, HA-SDCCAG8 or HA-OFD1 in hTERT-RPE-1 cells, we observed co-localization at the level of the endogenous cilium BB (Fig. 8). Once again, no co-localization could be observed by co-expressing lebercilin and the FAM161A-N-term construct (Supplementary Material, Fig. S3).

FAM161A is involved in the assembly of the cilium

To investigate whether FAM161A had a direct role in cilium assembly, we knocked down *FAM161A* transcripts in hTERT-RPE-1 cells by a pool of three different small interfering RNAs (siRNAs) targeting its sequence. Following a >80% silencing of *FAM161A* expression [revealed by quantitative polymerase chain reaction (qPCR), Supplementary Material, Fig. S4A], we observed a significant reduction in the number of ciliated cells (Supplementary Material, Fig. S4B and C). Specifically, the percentage of cells displaying a completely formed cilium dropped from 71% in controls (cells treated with siRNA against the non-human gene luciferase) to 57%, as assessed 48 h after siRNA supplementation in

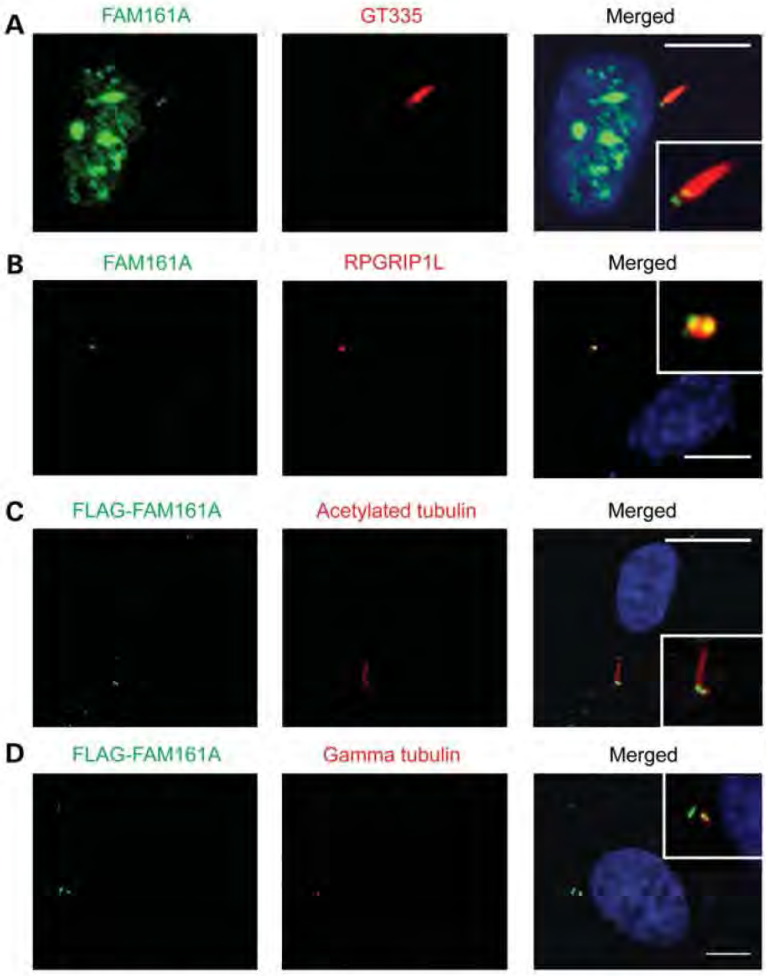


Figure 4. FAM161A in ciliated hTERT-RPE-1 cell lines. Endogenous FAM161A (green) localizes at the base of the cilium (stained with GT335, red), at the level of the BB (**A**). The BB localization is also shown by co-staining of FAM161A and RPGRIP1L (red) (**B**). These results are confirmed by co-localization of recombinant FAM161A (revealed by a FLAG antibody, green) with acetylated tubulin (red) (**C**), and gamma tubulin (red) (**D**) in the same cell lines. Insets present a magnified area of the merged picture. DAPI (blue) was used to stain the nuclei. Scale bars: 10 μm. In panel A, the patchy green staining of the nucleus represents background signal that is typical of the antibody used.

three independent experiments, each considering 300 cells per condition ($\chi^2 = 36.38$, $P < 0.0001$, Fig. 9).

DISCUSSION

More than one quarter of all genes that are mutated in HRD encode proteins that are responsible for cilium functions or development (25), highlighting the importance of this organelle for photoreceptor cell homeostasis and thus for vision in general. In this work, we demonstrate that FAM161A is a ciliary protein, interacts with other proteins involved in ciliopathies and plays a role in cilium assembly.

Previous localization experiments revealed that, within the mouse retina, FAM161A is present in photoreceptors, mostly at the level of their ISs (18,19). Co-immunostaining of FAM161A and a battery of IS markers revealed the absence of co-localization (not shown) and prompted us to investigate a smaller structure lying between the IS and the OS, the photoreceptor CC. Immunofluorescence performed with antibodies against centrin, a typical marker of the CC, revealed partial overlap with FAM161A. Additional FAM161A staining was also present in the IS and the OPL, as reported previously (18). Although we can speculate that the portion of FAM161A that is present within the IS could simply represent the nascent and immature form of this protein, at the moment we have no explanations for its localization within or in the vicinity of the OPL.

The ciliary localization of FAM161A was confirmed by the analysis of dissociated murine rod photoreceptors, providing the unique level of resolution typical of single-cell analyses. Co-staining with markers that are specifically marking

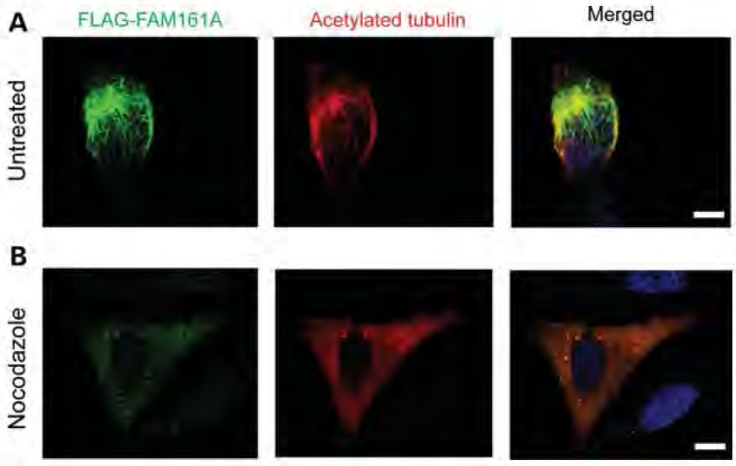


Figure 5. Overexpressed FAM161A associates with the microtubule network. When overexpressed in HeLa cells, flagged FAM161A (green) forms a web-like structure in correspondence of the cell microtubule network (red) (**A**). Treatment with 10 μM nocodazole, an inhibitor of microtubule polymerization, destroys such structure, highlighting the non-autonomous nature of this particular arrangement (**B**). Scale bars: 10 μm.

subciliary locations indicated that FAM161A exerts its function at the BB and the basal side of the CC of photoreceptors and is not detected at the photoreceptor axoneme. These data are consistent with previous work reporting the presence of FAM161A peptides in low amounts within the ciliary proteome (22).

Based on the relatively widespread mRNA expression of *FAM161A* (18,19) and on the structural conservation of the primary cilium across different cell types, we reasoned that FAM161A could be present in other tissues than the retina and in other ciliated cells than photoreceptors. Staining of rat kidney sections revealed indeed the presence of FAM161A. Remarkably, its localization within the cilium of renal cells corresponded perfectly to that observed in photoreceptors and specifically corroborated the presence of FAM161A within BB structures. The same was true when the presence of FAM161A was investigated in cultured ciliated cells. Moreover, the single-cell level of resolution that was possible by the analyses of cultured hTERT-RPE-1 and ARPE-19 allowed detecting the presence of FAM161A within centrosomal structures, as it is the case for many proteins that are component of sensory cilia (26) and in agreement with the concept that the BB is a modified centrosome of the photoreceptor cilium (27). Furthermore, several proteins involved in HRD have been shown to be directly or indirectly linked to the BB (11,13,28), the components of which are also thought to regulate the transport of molecules from the IS to the OS, across the CC (29).

In non-ciliated cell lines, overexpressed FAM161A organized in defined structures that partly overlapped with the cellular microtubule network. Such localization has also been observed for other ciliary proteins such as lebercilin (12), especially upon severe overexpression. This likely pinpoints the capacity of these centrosomal proteins to bind to acetylated microtubules and partly affect their dynamics. The effect is generally even more pronounced in non-ciliated cells that often induce significant overexpression, such as COS cells. We think that this marks the potential of ciliary/microtubule binding proteins to bundle microtubules by recruiting other microtubule-binding proteins that decorate this cytoskeletal structure (while it is not clear whether they do this natively). In support of this, treatment of cells overexpressing FAM161A with the microtubule depolymerizing drug nocodazole destroyed such FAM161A organization, indicating the need of an intact microtubule network for FAM161A to assemble in such web-like structures. Recent data by Zach *et al.* (30), available on-line a few days before the present work was submitted, described findings similar to ours on FAM161A localization. Following its overexpression in cultured cells, the authors hypothesized that FAM161A can have a stabilizing effect on microtubules in virtue of its interactions with them. Based on our collective findings, it is clear that FAM161A has the ability to interact with the cytoskeleton. However, as in our experiments, the presence of FAM161A within the microtubule network was never observed in physiological conditions, indicating that putative non-ciliary functions of FAM161A would require additional investigation.

Targeted yeast two-hybrid analysis revealed that FAM161A directly interacts with lebercilin. Similar to FAM161A, lebercilin localizes within the CC of photoreceptors and at the base of the cilium and centrosomes in ciliated cells (12). This interaction was confirmed by copurification and immunoprecipitation experiments. Very likely, the FAM161A–lebercilin interaction is achieved via the two C-terminal coiled-coil domains that are present within the UPF0564 domain, or alternatively via other undetected structures lying in the C-terminal part of FAM161A. In turn, lebercilin interacts with proteins that are important for centrosome and cilium functions, such as RPGR (12) and OFD1 (14,31). Recently, Boldt *et al.* (32) demonstrated the direct interaction of lebercilin with the intraflagellar machinery (IFT) and described a possible role of lebercilin in the protein transport across the cilium. However, since inactivation of *LCA5* did not interfere with cilium assembly or localization of other IFT proteins, lebercilin is probably not a structural component of the IFT

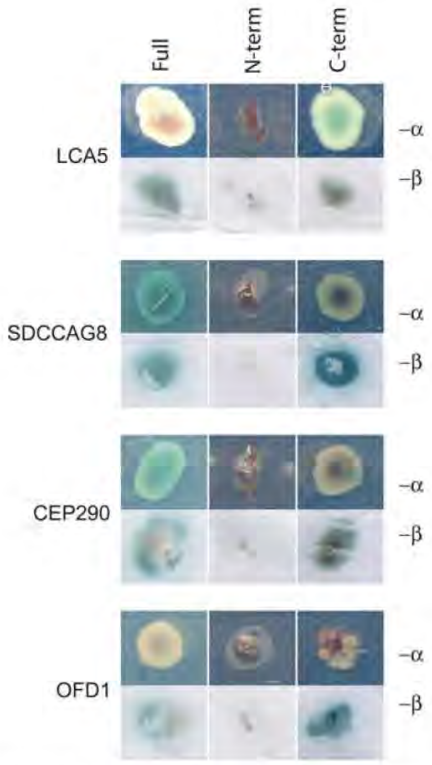


Figure 6. FAM161A interacts with different proteins involved in ciliopathies, as revealed by yeast two-hybrid direct binding assay. Constructs expressing FAM161A full form (Full), FAM161A-N-term (N-term) and FAM161A-C-term (C-term) fused to a GAL4-BD protein were co-transformed in yeast with a set of pAD vectors expressing different ciliary genes. Selective medium depleted for relevant amino acids was used to isolate the hybrids. The strength of the interaction was tested by blue staining of the colony following α-galactosidase (α) and β-galactosidase (β) activity. FAM161A showed the strongest interaction with lebercilin, OFD1, SDCCAG8 and CEP290, as assessed by both colony size and colorimetric assays. Prey constructs showed here included full-length sequences, except for OFD1 and CEP290. These were cloned as fragments and are listed in Supplementary Material, Table S1 as entries 'OFD1 328' and 'CEP290 cc4+5+6', respectively.

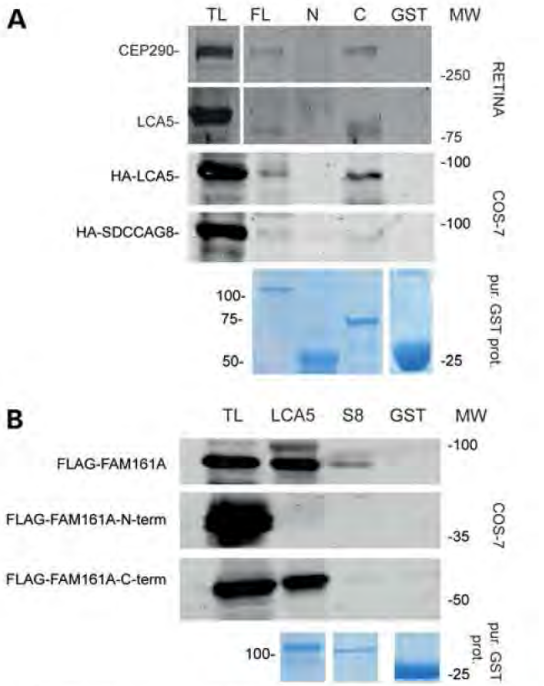


Figure 7. FAM161A interacts with RC proteins *in vivo* and in cultured cells. (**A**) Western blots of ciliopathy proteins after pull-down with GST-fused FAM161A constructs. The top two panels show pulled-down proteins from bovine retinal total lysates (TL) incubated with sepharose beads containing FAM161A-GST full length (FL), GST-FAM161A-N-term (N), GST-FAM161A-C-term (C), or GST alone (GST). GST pull-down allows FAM161A-GST and GST-FAM161A-C-term to reveal native retinal lebercilin and CEP290, whereas GST-FAM161A-N-term and GST alone do not. 2% of the total retinal lysate was loaded as input control. The third and fourth panels illustrate the same experiments performed in COS-7 cells, transfected with lebercilin (HA-LCA5) and HA-SDCCAG8. GST pull-down of FAM161A-GST and GST-FAM161A-C-term reveals binding to HA-LCA5, whereas the pull down of GST-FAM161A-N-term and GST alone does not. All FAM161A constructs can pull-down recombinant SDCCAG8. 10% of the cell lysate was loaded as input control. The fifth panel shows the GST-purified protein used for pull-down assays (pur. GST prot.). (**B**) Western blots of different FAM161A constructs after pull-down with GST-fused lebercilin (LCA5) and SDCCAG8 (S8) constructs in COS-7 cells. Signals corresponding to FLAG-FAM161A and to FLAG-FAM161A-C-term, but not to FLAG-FAM161A-N-term or GST alone, are revealed. Again, the Coomassie-stained gel image shows the GST-purified protein used for these assays. 10% of the cell lysate was loaded as input control. MW, molecular weight (kDa).

machinery. This fact could explain the absence of systemic phenotypes in patients with *LCA5* mutations, in contrast to the syndromic manifestations of IFT mutations, despite the presence of lebercilin in all ciliated tissues (32).

Interestingly, FAM161A interacts also with OFD1 and SDCCAG8, two proteins that were already observed to be part of the LCA5 complex, localizing as expected at the level of the BB. The interaction with SDCCAG8 was confirmed also in GST pull-down assays, although the level of the pulled-down protein was lower compared with lebercilin. We were unable to confirm the interaction between FAM161A and OFD1 using the same biochemical approach. Mutations in *SDCCAG8* were linked to both a retinal-renal ciliopathy (16) and to 1–2% of patients with BBS (33). In addition, mutations in *OFD1* were identified in patients with Joubert syndrome, nephronophthisis-like ciliopathies with retinal degeneration and RP (14,15,34).

Another important interactor of FAM161A is CEP290. This direct interaction, first observed in yeast binary assays, was also confirmed by pull-down experiments in retinal lysates. Mutations in *CEP290* were identified in patients with Joubert syndrome-associated ciliopathies (35,36). Additionally, mutations in *CEP290* are a major cause of Leber congenital amaurosis (13). CEP290 localizes at the CC in the murine photoreceptor, where FAM161A also partially localizes. In particular, the localization of CEP290 in the transition zone seems to be important for its function, possibly dealing with the docking and the transport of molecules inside the cilium (37). Furthermore, the depletion of CEP290 prevents cilium

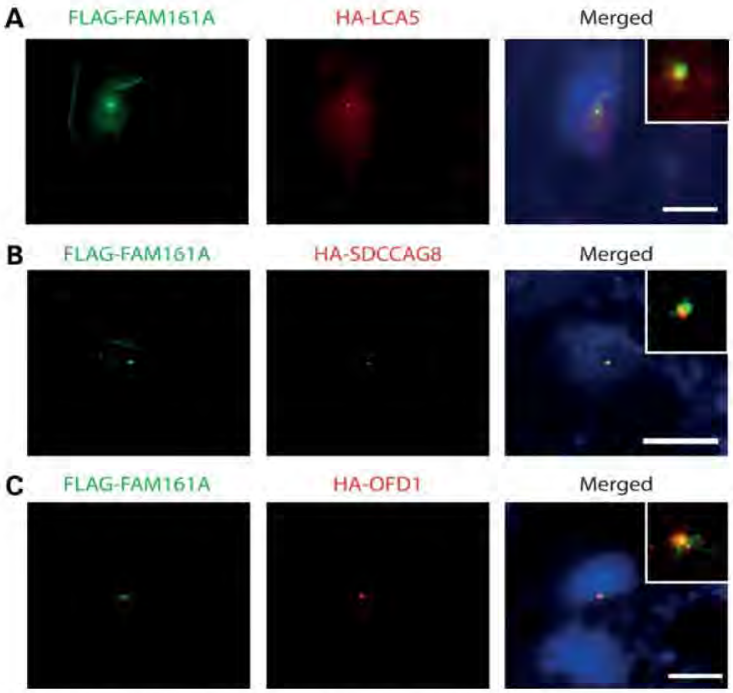


Figure 8. Recombinant lebercilin/LCA5, SDCCAG8, OFD1, and FAM161A co-localize in hTERT-RPE-1 cells after cilium induction. FLAG-FAM161A (green) co-localizes at the level of the BB with HA-LCA5 (red, **A**), SDCCAG8 (red, **B**), and OFD1 (red, **C**). Scale bars: 10 μm.

formation and assembly (38,39), as we also observed after FAM161A silencing. Because of its interactions with CEP290 and lebercilin, as well as its subciliary localization, an intriguing possibility is that FAM161A could be involved in protein transport and docking between the IS and the OS.

A strong evidence of the importance of FAM161A for the cilium comes from the negative effects of its depletion in ciliated cells, for which ciliary assembly itself is compromised. Mutations in *FAM161A* produce a retina-restricted phenotype and, compared with mutations in its interactors, a milder one. A possible explanation is that the FAM161A role could be hierarchically subordinated to that of its associated proteins and critical only for photoreceptors, where it is expressed in large amounts. Although further interaction experiments are needed to establish a solid mechanistic model, another fascinating possibility is that FAM161A, in virtue of its CRX-mediated expression, may help driving the organization of ciliary proteins having ubiquitous expression in rods and cones. In light of the phenotypic heterogeneity underlying its interactors and the current absence of an animal model, we also cannot rule out that variations in *FAM161A* are also modifying more complex syndromic ciliopathy phenotypes.

In conclusion, in this work we show that FAM161A is a functional protein of the cilium. It localizes at the base of the cilium in photoreceptors as well as in other ciliated cell types and tissues and prominently associates with a specific subset of proteins associated with RCs. Based on these findings, FAM161A-associated RP should be classified as a novel retinal ciliopathy.

MATERIALS AND METHODS

DNA constructs

Full-length *FAM161A* cDNA, obtained from human retina, was cloned into a pENTR vector (Invitrogen). Similar cloning was performed also for two deletion sequences, obtained by PCR. These generated plasmids pFAM161A-N-term, comprising codons 1–219 and containing the first of three coiled-coil domains, and pFAM161A-C-term, spanning codons 211–660 and including the UPF0564 domain. Expression plasmids (GST, FLAG and yeast clones containing the activation domain, AD or the binding domain, BD) were obtained starting from these entry clones using the Gateway cloning technology (Invitrogen). The specificity and quality of the fragments was assessed by restriction analysis and direct sequencing. Although a difference with respect to the reported reference sequence was identified, this corresponded to the annotated polymorphism rs17513722, having a 0.5 allele frequency within the European population.

Expression plasmids for *LCA5*, *SDCCAG8* and *OFD1* were described previously (12,14,16,21).

Antibodies

Custom polyclonal rabbit antibodies against FAM161A (used for the immunostaining depicted in Fig. 3A) included those described previously (18) as well as a new one (used for the staining of human retina sections), raised against a synthetic peptide corresponding to the C-terminal moiety of the murine protein (EKLQKRPMLFERVT) and used at a 1:100 dilution.

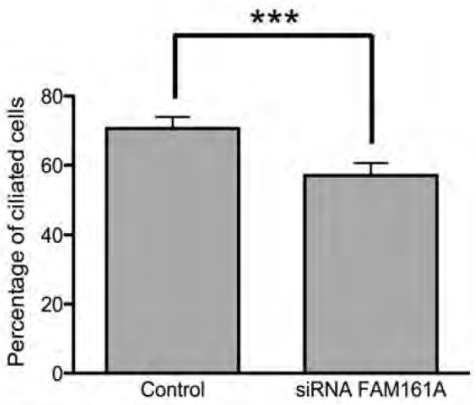


Figure 9. FAM161A knock down causes a reduction in the number of ciliated cells in hTERT-RPE-1 cell lines. The graph shows the percentage of ciliated cells after cilium induction in samples treated with FAM161A-specific siRNA and in the control group, treated with non-targeting siRNA. $N = 900$ counts per group, sampled in 3 different experiments. Error bars indicate standard errors. $P < 0.0001$, shown by the 3 asterisks.

Another commercial polyclonal rabbit anti-FAM161A antibody was purchased from Sigma-Aldrich and used to a final dilution of 1:100 in all other instances. Mouse monoclonal anti-FLAG antibody (1:1000) was obtained from Invitrogen, whereas rabbit polyclonal anti-HA antibody (1:1000) was obtained by Bethyl antibodies. Mouse and rabbit anti-γ-tubulin (1:1000) were obtained from Abcam, and mouse and rabbit anti acetylated α-tubulin (1:1000) from Sigma-Aldrich. Chicken anti-RP1 (1:2000) antibody was a kind gift of Dr E. Pierce (Harvard Medical School, Boston, USA), whereas anti-RPGRIP1L (1:1000) was obtained as reported previously (11). Mouse monoclonal anti-rhodopsin (Ret-1 clone, 1:200) antibody was purchased from Santa Cruz Biotechnology. Mouse monoclonal anti-centrin clone 20H5 (1:100) was obtained from Millipore. Mouse monoclonal antibody against polyglutamylated tubulin (GT335, 1:1000) was provided by Dr C. Janke (Institut Curie, Orsay, France). Polyclonal antibodies against lebercilin were obtained and used as reported before (12,32). Secondary goat antibodies anti-mouse, anti-rabbit, anti-chicken and anti-guinea pig were conjugated with Alexa Fluor 488 and Alexa Fluor 568 dyes (1:500). Secondary antibody anti-mouse and anti-rabbit conjugated with IRDye 680 and IRDye 800 for detection with the Odyssey system were purchased from LI-COR Biosciences and used at a 1:5000 final dilution.

Immunohistochemistry of retinal tissues and isolation of photoreceptors

Unfixed eyes of P20 Wistar rats and P20 C57BL/6J mice were isolated and frozen in melted isopentane. Cryosections of 10 μm were permeabilized in 0.01% Tween-20 in phosphate buffered saline (PBS) and then blocked in 0.1% ovoalbumin and 0.5% fish skin gelatin in PBS (blocking solution). Primary antibodies were diluted in blocking solution and then incubated overnight with the sections. After three washing steps with PBS, slides were incubated with Alexa-conjugated secondary antibodies and mounted with Prolong Gold Antifade (Invitrogen) mounting media to preserve fluorescence. Slides were then observed using direct fluorescence with the microscope Axioskop 4 (Carl Zeiss) and confocal scanning with the microscope Zeiss LSM 710 Quasar Confocal (Carl Zeiss). Images were processed using the Axiovision software (Carl Zeiss) and Adobe Photoshop CS5 (Adobe Systems). For paraffin sections (human), a standard immunohistological analysis protocol for the paraffin section was carried out as described previously (40), by using a secondary antibody the DyLight 488 goat anti-rabbit (Jackson ImmunoResearch) at a 1:200 dilution.

Isolated photoreceptors were obtained by disruption of fresh retinas from adult mice, according to the protocol by Liu *et al.* (22), modified as follows. The intact retina was resuspended in PBS and then gently vortexed for a few seconds, twice. The suspension was then spread on the surface of a microscope slide and let dry for 1 h, before proceeding to IF analysis.

Cell growth and transfection

Human immortalized retinal pigmented epithelium cells ARPE-19 and hTERT-RPE-1 were grown at 37°C in 5% CO_2 and Dulbecco's modified Eagle's medium/nutrient mixture F-12 (DMEM/F-12) (Gibco) media (1:1 ratio), supplemented with 10% fetal calf serum (FCS), 100 IU/ml penicillin and 100 μg/ml streptomycin. Human-derived HeLa, HEK293 and monkey-derived COS-7 cells were cultivated with DMEM containing 10% FCS, 100 IU/ml penicillin, 100 μg/ml streptomycin and 1% Glutamax (Gibco) under the same growth conditions. Cells were transfected using Lipofectamine 2000 (Invitrogen) for 24–48 h according the company specifications. Plasmids containing *FAM161A* sequences were used in co-localization studies on cells that were fixed 24 h after transfection.

For the microtubule disassembly assay, nocodazole (Sigma-Aldrich) was added directly to the growing medium at a 10 μM final concentration and cells were fixed in 2% paraformaldehyde (PFA) after 2 h. They were then stained according to the published protocols (41).

Immunocytochemistry

hTERT-RPE-1 and ARPE-19 cells were cultured, seeded on coverslips and serum starved for 24 h (0.2% FCS). Cells were then washed twice in PBS and fixed in 2% PFA for 20 min. HeLa cells were seeded on a coverslip and directly fixed in 2% PFA for 20 min. After fixation, they were treated with PBS plus 1% Triton X-100 for 5 min and then blocked in 2% bovine serum albumin in PBS for 1 h. Staining was achieved following overnight incubation with relevant antibodies. After incubation, fixed cells were washed twice in PBS and stained with a fluorescent secondary antibody for 1 h. Coverslips were then washed first with PBS, briefly with MilliQ water, and then mounted with Vectashield and DAPI (Vector Laboratories).

Direct yeast two-hybrid interaction assay

The direct interaction between FAM161A and other ciliary proteins was tested using a GAL4-based yeast two-hybrid

system, with yeast strain PJ69-4A. Constructs encoding full length or the deletion fragments of FAM161A described above, fused to a DNA-binding domain (GAL4-BD), were used as baits. Preys were sequences from previously described ciliopathy genes fused to an activation domain (GAL4-AD). The direct interaction between baits and preys induced the activation of the reporter genes, resulting in the growth of yeast colonies on selective media (deficient of histidine and adenine) and induction of α-galactosidase and β-galactosidase colorimetric reactions (42).

GST pull down

To produce the GST proteins, BL21(DE3) cells were transformed with FAM161A-GST, GST-FAM161A-N-term, GST-FAM161A-C-term, GST-LCA5 or GST alone constructs and induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 30°C overnight. Cells were then lysed in sodium chloride-Tris-EDTA (STE) buffer [10 mM Tris–HCl (pH 8), 1 mM ethylenediaminetetraacetic acid (EDTA), 150 mM NaCl] supplemented with 10 mg/ml lysozime, 0.5% sarkosyl, 1% Triton X-100 and Complete Protease Inhibitor Cocktail (Roche). Overnight incubation with glutathione-sepharose 4B beads (Roche) was performed to isolate the GST-fused proteins. This was followed by three washes with STE buffer and four washes with TBSDT buffer [Tris–HCl 25 mM, pH 7.4, 150 mM NaCl, 1% Triton X-100 and fresh 2 mM dithiothreitol (DTT)]. The amount of produced GST protein was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE) and blue Coomassie staining. COS-7 cells seeded in 10 cm plates were transfected using Lipofectamine 2000 with 3×FLAG-FAM161A, 3×FLAG-FAM161A-N-term, 3×FLAG-FAM161A-C-term and 3×HA-LCA5. Subsequently, proteins were extracted using the previously described lysis buffer and incubated with the isolated GST-fused protein sepharose beads from 5 h to overnight and finally washed five times with the protein lysis buffer used for Co-IP. The beads were then resuspended in loading buffer and analyzed by SDS–PAGE. FLAG or HA tags were revealed by using anti-FLAG or anti-HA antibodies.

Retinas were dissected from two fresh bovine eyes and put directly in Lysis Buffer for the Cytosolic Fraction (10 mM HEPES, pH 7.9, 10 mM NaCl, 3 mM $MgCl_2$, fresh 1 mM DTT and 1 mM sodium orthovanadate) and in lysis buffer for the membrane fraction (50 mM HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 0.5% Triton X-100, 1.5 mM $MgCl_2$, 1 mM ethylene glycol tetraacetic acid and 1 mM EDTA) supplemented with protease inhibitors. Retina extracts were then incubated overnight with FAM161A-GST, GST-FAM161A-N-term, and GST-FAM161A-C-term fusion proteins and with GST alone as a control. The beads were then washed five times with a mix of 50% cytosolic and 50% membrane lysis buffers. The beads were then incubated in loading buffer for 5 min at 95°C and the supernatant processed by SDS–PAGE. The signal of native LCA5 was revealed using a rabbit polyclonal antibody, previously described by den Hollander *et al.* (12), whereas CEP290 was revealed using the rabbit antibody obtain from Bethyl.

Knock down of *FAM161A*

To silence *FAM161A*, hTERT-RPE-1 cells at ~80% confluency were transfected with a pool of three sequence-specific siRNAs (5′-CCAACCUAGAAAAAGAGUAtt-3′, 5′-CCACAAUUACAGUACCGGAtt-3′, 5′-GCAAAAAGAAGAACGGAGAtt-3′) obtained from Ambion (Life Technologies) and with a control siRNA against the firefly luciferase gene (Qiagen). siRNAs were conjugated with a fluorescein fluorophore and transfected using the HiPerfect transfection reagent (Qiagen). Transfection efficiency was evaluated to be ~70% by counting the number of fluorescent cells transfected with control siRNA. RNA was extracted after 48 h from transfection and the relative expression of *FAM161A* was evaluated by qPCR. The number of the ciliated cells was evaluated after 2% PFA fixation of transfected cells and following staining with the anti-acetylated tubulin antibody. Statistical analysis was performed by using the on-line χ^2 calculator available at http://vassarstats.net/newcs.html.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We want to kindly acknowledge T.A. Peters for providing rat retinal slides, H. Koskiniemi for providing the FAM161A-GST plasmid, as well as D. Mans, E. van Wijk and other members of the R. Roepman/H. Kremers lab for useful advice and helpful discussions. We are grateful to A. Obolensky, E. Banin and G. Tanackovic for providing important suggestions about the experimental procedures.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Swiss National Science Foundation (310030-138346 to C.R. and Sinergia #CRSII3_141814 to C.R., D.S., and Y.A.), the Israeli Ministry of Health (3-7261 to D.S.), a grant from the Netherlands Organization for Scientific Research (NWO Vidi-91786396 to R.R.) and the European Community's Seventh Framework Programme FP7/2009 under grant agreement number 241955, SYSCILIA to R.R.

REFERENCES

1. Berson, E.L. (1993) Retinitis pigmentosa. The Friedenwald Lecture. *Invest. Ophthalmol. Vis. Sci.*, **34**, 1659–1676.
2. Hartong, D.T., Berson, E.L. and Dryja, T.P. (2006) Retinitis pigmentosa. *Lancet*, **368**, 1795–1809.
3. Rivolta, C., Sharon, D., DeAngelis, M.M. and Dryja, T.P. (2002) Retinitis pigmentosa and allied diseases: numerous diseases, genes, and inheritance patterns. *Hum. Mol. Genet.*, **11**, 1219–1227.
4. Berger, W., Kloeckener-Gruissem, B. and Neidhardt, J. (2010) The molecular basis of human retinal and vitreoretinal diseases. *Prog. Retin. Eye Res.*, **29**, 335–375.
5. den Hollander, A.I., Black, A., Bennett, J. and Cremers, F.P. (2010) Lighting a candle in the dark: advances in genetics and gene therapy of recessive retinal dystrophies. *J. Clin. Invest.*, **120**, 3042–3053.

6. Adams, N.A., Awadein, A. and Toma, H.S. (2007) The retinal ciliopathies. *Ophthalmic. Genet.*, **28**, 113–125.
7. Novarino, G., Akizu, N. and Gleeson, J.G. (2011) Modeling human disease in humans: the ciliopathies. *Cell*, **147**, 70–79.
8. Hildebrandt, F., Benzing, T. and Katsanis, N. (2011) Ciliopathies. *N. Engl. J. Med.*, **364**, 1533–1543.
9. Liu, Q., Zhou, J., Daiger, S.P., Farber, D.B., Heckenlively, J.R., Smith, J.E., Sullivan, L.S., Zuo, J., Milam, A.H. and Pierce, E.A. (2002) Identification and subcellular localization of the RP1 protein in human and mouse photoreceptors. *Invest. Ophthalmol. Vis. Sci.*, **43**, 22–32.
10. Hosch, J., Lorenz, B. and Stieger, K. (2011) RPGR: role in the photoreceptor cilium, human retinal disease, and gene therapy. *Ophthalmic. Genet.*, **32**, 1–11.
11. Arts, H.H., Doherty, D., van Beersum, S.E., Parisi, M.A., Letteboer, S.J., Gorden, N.T., Peters, T.A., Marker, T., Voesenek, K., Kartono, A. *et al.* (2007) Mutations in the gene encoding the basal body protein RPGRIP1L, a nephrocystin-4 interactor, cause Joubert syndrome. *Nat. Genet.*, **39**, 882–888.
12. den Hollander, A.I., Koenekoop, R.K., Mohamed, M.D., Arts, H.H., Boldt, K., Towns, K.V., Sedmak, T., Beer, M., Nagel-Wolfrum, K., McKibbin, M. *et al.* (2007) Mutations in LCA5, encoding the ciliary protein lebercilin, cause Leber congenital amaurosis. *Nat. Genet.*, **39**, 889–895.
13. den Hollander, A.I., Koenekoop, R.K., Yzer, S., Lopez, I., Arends, M.L., Voesenek, K.E., Zonneveld, M.N., Strom, T.M., Meitinger, T., Brunner, H.G. *et al.* (2006) Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *Am. J. Hum. Genet.*, **79**, 556–561.
14. Coene, K.L., Roepman, R., Doherty, D., Afroze, B., Kroes, H.Y., Letteboer, S.J., Ngu, L.H., Budny, B., van Wijk, E., Gorden, N.T. *et al.* (2009) OFD1 is mutated in X-linked Joubert syndrome and interacts with LCA5-encoded lebercilin. *Am. J. Hum. Genet.*, **85**, 465–481.
15. Webb, T.R., Parfitt, D.A., Gardner, J.C., Martinez, A., Bevilacqua, D., Davidson, A.E., Zito, I., Thiselton, D.L., Ressa, J.H., Apergi, M. *et al.* (2012) Deep intronic mutation in OFD1, identified by targeted genomic next-generation sequencing, causes a severe form of X-linked retinitis pigmentosa (RP23). *Hum. Mol. Genet.*, **21**, 3647–3654.
16. Otto, E.A., Hurd, T.W., Airik, R., Chaki, M., Zhou, W., Stoetzel, C., Patil, S.B., Levy, S., Ghosh, A.K., Murga-Zamalloa, C.A. *et al.* (2010) Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. *Nat. Genet.*, **42**, 840–850.
17. Billingsley, G., Vincent, A., Deveault, C. and Heon, E. (2012) Mutational Analysis of SDCCAG8 in Bardet-Biedl Syndrome Patients with Renal Involvement and Absent Polydactyly. *Ophthalmic. Genet.*, **33**, 150–154.
18. Langmann, T., Di Gioia, S.A., Rau, I., Stohr, H., Maksimovic, N.S., Corbo, J.C., Renner, A.B., Zrenner, E., Kumaramanickavel, G., Karlstetter, M. *et al.* (2010) Nonsense mutations in FAM161A cause RP28-associated recessive retinitis pigmentosa. *Am. J. Hum. Genet.*, **87**, 376–381.
19. Bandah-Rozenfeld, D., Mizrahi-Meissonnier, L., Farhy, C., Obolensky, A., Chowers, I., Pe'er, J., Merin, S., Ben-Yosef, T., Ashery-Padan, R., Banin, E. *et al.* (2010) Homozygosity mapping reveals null mutations in FAM161A as a cause of autosomal-recessive retinitis pigmentosa. *Am. J. Hum. Genet.*, **87**, 382–391.
20. Gu, S., Kumaramanickavel, G., Srikumari, C.R., Denton, M.J. and Gal, A. (1999) Autosomal recessive retinitis pigmentosa locus RP28 maps between D2S1337 and D2S286 on chromosome 2p11-p15 in an Indian family. *J. Med. Genet.*, **36**, 705–707.
21. van Wijk, E., van der Zwaag, B., Peters, T., Zimmermann, U., Te Brinke, H., Kersten, F.F., Marker, T., Aller, E., Hoefsloot, L.H., Cremers, C.W. *et al.* (2006) The DFNB31 gene product whirlin connects to the Usher protein network in the cochlea and retina by direct association with USH2A and VLGR1. *Hum. Mol. Genet.*, **15**, 751–765.
22. Liu, Q., Tan, G., Levenkova, N., Li, T., Pugh, E.N. Jr., Rux, J.J., Speicher, D.W. and Pierce, E.A. (2007) The proteome of the mouse photoreceptor sensory cilium complex. *Mol. Cell Proteomics.*, **6**, 1299–1317.
23. Wolfrum, U. (1995) Centrin in the photoreceptor cells of mammalian retinae. *Cell Motil. Cytoskeleton*, **32**, 55–64.
24. Guarguaglini, G., Duncan, P.I., Stierhof, Y.D., Holmstrom, T., Duensing, S. and Nigg, E.A. (2005) The forkhead-associated domain protein Cep170 interacts with Polo-like kinase 1 and serves as a marker for mature centrioles. *Mol. Biol. Cell*, **16**, 1095–1107.
25. Wright, A.F., Chakarova, C.F., Abd El-Aziz, M.M. and Bhattacharya, S.S. (2010) Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nat. Rev. Genet.*, **11**, 273–284.
26. Jakobsen, L., Vanselow, K., Skogs, M., Toyoda, Y., Lundberg, E., Poser, I., Falkenby, L.G., Bennetzen, M., Westendorf, J., Nigg, E.A. *et al.* (2011) Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. *EMBO J.*, **30**, 1520–1535.
27. Moser, J.J., Fritzler, M.J., Ou, Y. and Rattner, J.B. (2010) The PCM-basal body/primary cilium coalition. *Semin. Cell Dev. Biol.*, **21**, 148–155.
28. Chakarova, C.F., Khanna, H., Shah, A.Z., Patil, S.B., Sedmak, T., Murga-Zamalloa, C.A., Papaioannou, M.G., Nagel-Wolfrum, K., Lopez, I., Munro, P. *et al.* (2011) TOPORS, implicated in retinal degeneration, is a cilia-centrosomal protein. *Hum. Mol. Genet.*, **20**, 975–987.
29. Roepman, R. and Wolfrum, U. (2007) Protein networks and complexes in photoreceptor cilia. *Subcell Biochem.*, **43**, 209–235.
30. Zach, F., Grassmann, F., Langmann, T., Sorusch, N., Wolfrum, U. and Stohr, H. (2012) The retinitis pigmentosa 28 protein FAM161A is a novel ciliary protein involved in intermolecular protein interaction and microtubule association. *Hum. Mol. Genet.*, in press.
31. Singla, V., Romaguera-Ros, M., Garcia-Verdugo, J.M. and Reiter, J.F. (2010) Ofd1, a human disease gene, regulates the length and distal structure of centrioles. *Dev. Cell*, **18**, 410–424.
32. Boldt, K., Mans, D.A., Won, J., van Reeuwijk, J., Vogt, A., Kinkl, N., Letteboer, S.J., Hicks, W.L., Hurd, R.E., Naggert, J.K. *et al.* (2011) Disruption of intraflagellar protein transport in photoreceptor cilia causes Leber congenital amaurosis in humans and mice. *J. Clin. Invest.*, **121**, 2169–2180.
33. Schaefer, E., Zaloszyc, A., Lauer, J., Durand, M., Stutzmann, F., Perdomo-Trujillo, Y., Redin, C., Bennouna Greene, V., Toutain, A., Perrin, L. *et al.* (2011) Mutations in SDCCAG8/NPHP10 Cause Bardet-Biedl Syndrome and Are Associated with Penetrant Renal Disease and Absent Polydactyly. *Molecular syndromology*, **1**, 273–281.
34. Zullo, A., Iaconis, D., Barra, A., Cantone, A., Messaddeq, N., Capasso, G., Dolle, P., Igarashi, P. and Franco, B. (2010) Kidney-specific inactivation of Ofd1 leads to renal cystic disease associated with upregulation of the mTOR pathway. *Hum. Mol. Genet.*, **19**, 2792–2803.
35. Sayer, J.A., Otto, E.A., O'Toole, J.F., Nurnberg, G., Kennedy, M.A., Becker, C., Hennies, H.C., Helou, J., Attanasio, M., Fausett, B.V. *et al.* (2006) The centrosomal protein nephrocystin-6 is mutated in Joubert syndrome and activates transcription factor ATF4. *Nat. Genet.*, **38**, 674–681.
36. Valente, E.M., Silhavy, J.L., Brancati, F., Barrano, G., Krishnaswami, S.R., Castori, M., Lancaster, M.A., Boltshauser, E., Boccone, L., Al-Gazali, L. *et al.* (2006) Mutations in CEP290, which encodes a centrosomal protein, cause pleiotropic forms of Joubert syndrome. *Nat. Genet.*, **38**, 623–625.
37. Garcia-Gonzalo, F.R., Corbit, K.C., Sirerol-Piquer, M.S., Ramaswami, G., Otto, E.A., Noriega, T.R., Seol, A.D., Robinson, J.F., Bennett, C.L., Josifova, D.J. *et al.* (2011) A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. *Nat. Genet.*, **43**, 776–784.
38. Tsang, W.Y., Bossard, C., Khanna, H., Peranen, J., Swaroop, A., Malhotra, V. and Dynlacht, B.D. (2008) CP110 suppresses primary cilia formation through its interaction with CEP290, a protein deficient in human ciliary disease. *Dev. Cell*, **15**, 187–197.
39. Kim, J., Krishnaswami, S.R. and Gleeson, J.G. (2008) CEP290 interacts with the centriolar satellite component PCM-1 and is required for Rab8 localization to the primary cilium. *Hum. Mol. Genet.*, **17**, 3796–3805.
40. Banin, E., Obolensky, A., Idelson, M., Hemo, I., Reinhardtz, E., Pikarsky, E., Ben-Hur, T. and Reubinoff, B. (2006) Retinal incorporation and differentiation of neural precursors derived from human embryonic stem cells. *Stem Cells*, **24**, 246–257.
41. Liu, Q., Zuo, J. and Pierce, E.A. (2004) The retinitis pigmentosa 1 protein is a photoreceptor microtubule-associated protein. *J. Neurosci.*, **24**, 6427–6436.
42. Letteboer, S.J. and Roepman, R. (2008) Versatile screening for binary protein-protein interactions by yeast two-hybrid mating. *Methods Mol. Biol.*, **484**, 145–159.

I.4 SUPPLEMENTARY MATERIAL

Supplementary Table S1 List of tested pAD ciliopathy constructs used for the binary interaction assay

0: no blue color after the colorimetric assay or no hybrid colony formed
1: faint blue signal or not well-formed colony
2: medium blu signal and clearly-formed colony
3: strong blue signal and very well formed colony

Abbreviations: amino acids (aa); fragments (fr); C2 domain of RPGR and RPGRIP1L (C2); RPGR interacting domain (RID); Pleckstrin-like homology binding domain 1 (PH-B1); coiled coil (cc); N terminal (N-term); C terminal (C-term)
FAM161A full construct (FAM-full); FAM161A-N-term construct (FAM-Nt); FAM161A-C-term (FAM-Ct)
α gal (alpha galactosidase assay); β gal (beta galactosidase assay)

Protein	Gene symbol	Entrez ID	FAM-full α gal	FAM-full β gal	FAM-Nt α gal	FAM-Nt β gal	FAM-Ct α gal	FAM-Ct β gal
1	MAEA	10296	0	0	0	0	0	
2	C20orf11	54994	0	0	0	0	0	
3	YPEL5	51646	0	0	0	0	0	
4	MKLN1	4289	0	0	0	0	0	
5	SMU1	55234	0	0	0	0	0	
6	RANBP10	57610	1	0	0	0	2	
7	RMND5B	64777	0	0	0	0	0	
8	RMND5A	64795	0	0	0	0	0	
9	MAPRE1	22919	0	0	0	0	0	
10	EPS8 *aa 551-822*	2059	0	0	0	0	0	
	EPS8 *aa 551-616*		0	0	0	0	0	
	EPS8 *aa 182-248*		0	0	0	0	0	
	EPS8 *aa 1-211*		0	0	0	0	0	
	EPS8 *L329P*		0	0	0	0	0	
	EPS8		1	1	0	0	0	
11	ESPN	83715	1	1	0	0	0	
12	IFT20	90410	1	2	0	0	0	
13	IFT27	11020	0	0	0	0	0	
14	IFT52	51098	0	0	0	0	0	
15	TRAF3IP1	26146	0	0	0	0	0	
16	IFT57	55081	0	0	0	0	0	
17	IFT74	80173	0	0	0	0	1	
18	IFT81	28981	0	0	0	0	0	
19	IFT88	8100	2	2	0	0	0	
20	TTC30A	92104	0	0	0	0	0	
21	IFT140	9742	0	0	0	0	0	
22	WDR19	57728	0	0	0	0	0	
23	IFT43	112752	0	0	0	0	0	
24	NPHP4 fr1 *aa 484-784*	261734	0	0	0	0	0	
	NPHP4 fr2 *aa 484-981*		0	0	0	0	0	
	NPHP4 fr3 *aa 683-784*		0	0	0	0	0	
	NPHP4 fr4 *aa 683-981*		0	0	0	0	0	
	NPHP4 fr5 *aa 591-960*		0	0	0	0	0	
	NPHP4 fr6 *aa 591-1093*		0	0	0	0	0	
	NPHP4 fr7 *aa 961-1340*		0	0	0	0	0	
	NPHP4 fr8 *aa 411-590*		0	0	0	0	0	
	NPHP4 fr9 *aa 41-683*		0	0	0	0	0	
	NPHP4		0	0	0	0	0	
25	BBS5 fr1 *aa 2-150*	129880	0	0	0	0	0	
	BBS5 fr2 *aa 111-194*		0	0	0	0	0	
	BBS5 fr3 *aa 151-341*		0	0	0	0	0	
	BBS5 PH-B2 *aa 124-341*		0	0	0	0	0	
	BBS5		0	0	0	0	0	
26	SDCCAG8	10806	2	2	0	0	0	
	SDCCAG8 fr 1 *aa 2-294*		3	3	0	0	3	
	SDCCAG8 fr 2 *aa 286-541*		0	0	0	0	0	
	SDCCAG8 fr 3 *aa 533-713*		0	0	0	0	0	
27	C8orf37	157657	0	0	0	0	0	
28	RPGRIP1	57096	0	0	0	0	0	
	RPGRIP1 C2 short *aa 397-551*		0	0	0	0	0	
	RPGRIP1 C2 long *aa 204-551*		0	0	0	0	0	
	RPGRIP1 C2-end *aa 584-1259*		0	0	0	0	0	
29	RPGRIP1L C2 C-term+C2 N-term *aa 509-914*	23322	0	0	0	0	0	
	RPGRIP1L C2 end *aa 591-1235*		0	0	0	0	0	
	RPGRIP1L RID *aa 897-1235*		0	0	0	0	0	
30	CC2D2A	57545	0	0	0	0	0	
	CC2D2A cc *aa 433-637*		0	0	0	0	0	
	CC2D2A C2 *aa 992-1177*		0	0	0	0	0	
	CC2D2A N-term *aa 1-446*		0	0	0	0	0	
	CC2D2A C2-end *aa 992-1561*		0	0	0	0	0	
31	CRB1 human	23418	0	0	0	0	0	
	CRB1 bovine	520406	0	0	0	0	0	
32	IFT88 N-term *aa 1-449*	1800	0	0	0	0	0	
	IFT88 C-term *aa 442-833*		0	0	0	0	0	
33	WDR35	57539	0	0	0	0	0	
34	OFD1 328 *aa 240-1012*	8481	2	2	0	0	1	
	OFD1 369 *aa 356-1012*		2	2	0	0	2	
35	LCA5L fr1 *aa 2-376*	150082	0	0	0	0	0	
	LCA5L fr2 *aa 371-670*		0	0	0	0	0	
	LCA5L		0	0	0	0	0	
36	RPGR	6103	0	0	0	0	0	
	RPGR N-term *aa 1-401*		0	0	0	0	0	
	RPGR C-term *aa 404-815*		0	0	0	0	0	
37	NEK4 fr1 *aa 2-300*	6787	0	0	0	0	0	
	NEK4 fr2 *aa 293-566*		0	0	0	0	0	
	NEK4 fr3 *aa 560-841*		0	0	0	0	0	

	NEK4		0	0	0	0	0	
38	CEP290 cc4+5+6 *aa 703-1130*	80184	3	3	0	0	3	
39	NEK8	284086	0	0	0	0	0	
40	USH2A	7399	0	0	0	0	0	
41	GPR98	84059	0	0	0	0	0	
42	DFNB31	25861	0	0	0	0	0	
43	USH1C	10083	0	0	0	0	0	
44	LCA5	167691	1	0	0	0	0	
	LCA5 fr1 *aa 1-95*		0	0	0	0	0	
	LCA5 fr2 *aa 96-305*		2	3	0	0	3	
	LCA5 fr3 *aa 306-490*		0	0	0	0	0	
	LCA5 fr5 *aa 1-305*		2	3	0	0	2	
	LCA5 fr6 *aa 306-697*		0	0	0	0	0	
45	USG1G	124590	1	0	0	0	0	

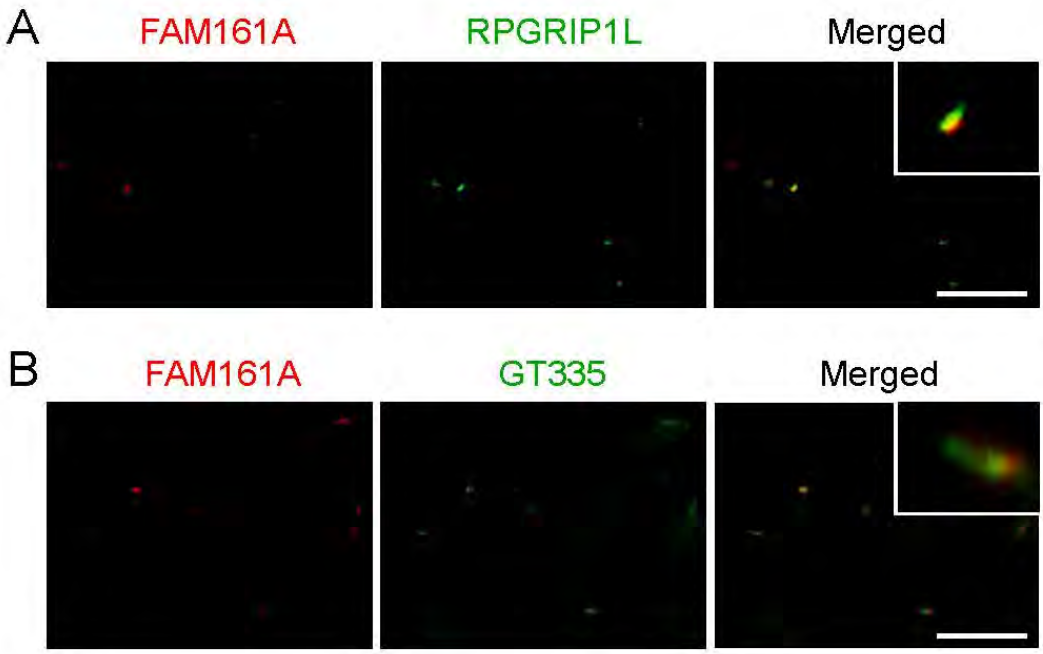


Figure S1. FAM161A localizes at the level of the basal body in cells from rat kidney sections. The overlay between FAM161A (red) and RPGRIP1L (green), a marker of the basal body, is almost complete (A). Co-staining between FAM161A (red) and GT335 (green), a marker of the cilium, shows partial co-localization within the basal body only. Scale bars: 10 µm.

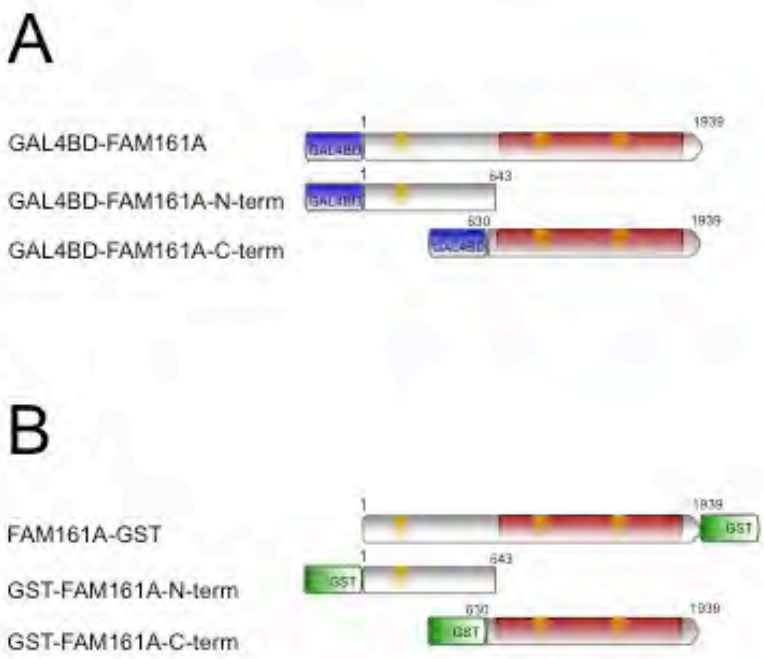


Figure S2. Schematic structure of FAM161A constructs used for yeast two-hybrid and pull-down assays. Yellow bars indicate the position of the predicted coiled-coil structures, whereas the UPF0564 domain is shown in red. Numbers indicate reference amino acid residues of the major FAM161A protein form. (A) Yeast two-hybrid; the GAL4 binding domain (GAL4-BD), fused to FAM161A sequences, is depicted in blue. (B) Pull-down experiments; the GST-tag is shown as a green box.

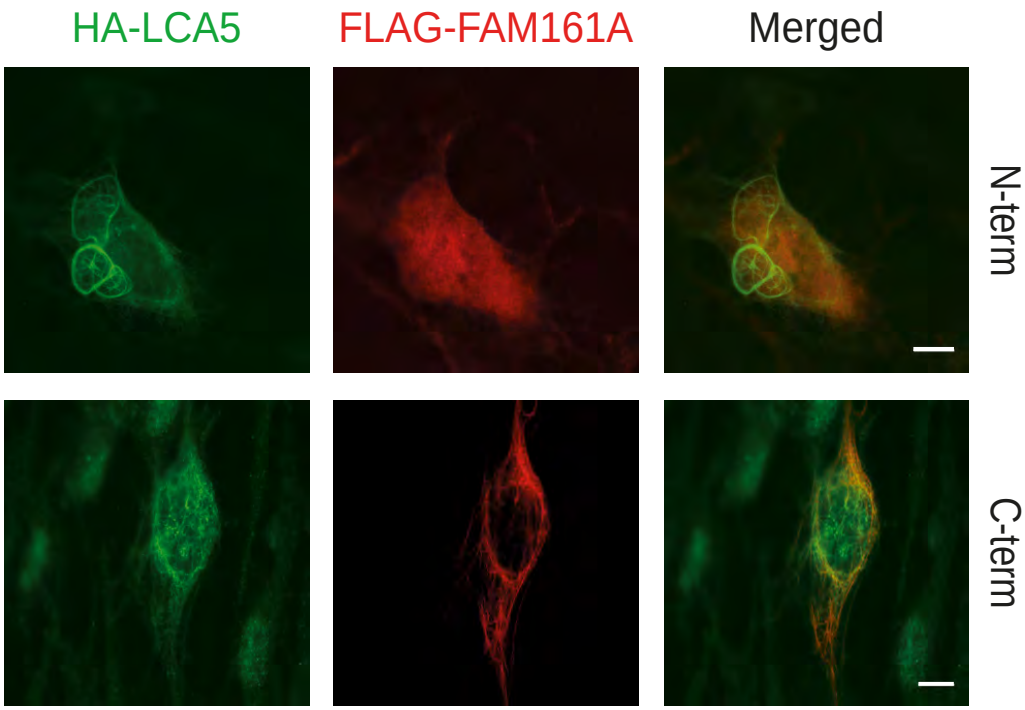


Figure S3. The interaction between FAM161A and lebercilin/LCA5 is driven by the C-terminal domain of FAM161A. The construct expressing the N-terminal (upper panels - N-term) moiety of FLAG-FAM161A (red) localizes in the cytoplasm, whereas HA-LCA5 is present at the level of the microtubule network. Conversely, the C-terminal (bottom panels – C-term) part of FLAG-FAM161A (red) decorates microtubules and partly overlaps with HA-LCA5 (green). Scale bars: 10 µm.

I.5 ADDENDUM

YEAST-2-HYBRID SCREENING OF HUMAN AND BOVINE RETINAL LIBRARIES AND TANDEM AFFINITY PURIFICATION ANALISYS TO IDENTIFY FAM161A SPECIFIC INTERACTORS

Silvio Alessandro Di Gioia[1], Stef J.F Letteboer[2], Lisette Hetterschijt[2], Sylvia van Beersum[2] , Jeroen van Reeuwijk[2], Ronald Roepman[3], and Carlo Rivolta[1].

[1]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland, [2] Department of Human Genetics and Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Abstract

FAM161A is a new ciliary protein involved in retinitis pigmentosa. The role of this protein in the ciliary compartment is still not well characterized. To better elucidate the function of this gene we performed two different high-throughput screening techniques; yeast two-hybrid screening of human and bovine retina cDNA library and tandem affinity purification analysis from cell extracts overexpressing FAM161A. We obtained several positive hits using these two techniques with enrichment of proteins from basal body/centrosome and transport proteins. The identification of several proteins involved in intracellular and ciliary trafficking as interactors of FAM161A could suggest a possible role of this protein in the molecular transport either to or on the cilium and therefore clarify the possible pathogenic mechanism linked to mutation in this gene. Confirmation of these results using biochemical analysis is necessary to reduce the false positive rate that is often observed for these techniques.

I performed this work mainly in Nijmegen under the supervision of Ronal Roepman and his team. These data are not published.

INTRODUCTION

Retinal degenerations are one of the main symptoms of ciliopathies. These are a class of diseases in which the functionality of the cilium, an important organelle of mammalian cells, is disturbed causing an impairment in cellular trafficking or development [109]. Interestingly several papers demonstrated that proteins for which the gene is mutated in ciliopathies often interact together, forming molecular networks that also interconnect different classes of ciliopathies [67, 89, 110-112]. We recently identified FAM161A, a gene mutated in an autosomal recessive retinitis pigmentosa, as a new ciliary gene, based on localization studies as well as some functional evidence [113]. In line with our findings, a different group confirmed this ciliary localization. Furthermore, they investigated the microtubule binding activity and interaction of FAM161A with its mammalian paralog FAM161B [114] and they suggested a possible role of this family in microtubule network stabilization. Based on these researches we used two different high-throughput screening techniques, yeast-2-hybrid (Y2H) of human and bovine retina library and tandem affinity purification (TAP) associated to mass spectrometry, to investigate "neighbors" of FAM161A, in order to elucidate the function of FAM161A itself. These techniques have previously been extensively used to investigate the interactome of diverse ciliary proteins [85, 88, 111, 115, 116].

METHODS

DNA constructs.

All DNA constructs and plasmids for FAM161A were generated using the gateway system from Invitrogen (Life technology, USA), using the plasmid described previously as a donor [113]. Destination plasmids for Y2H assay were also described previously [113]. For the N-terminal TAP experiment we cloned in a TAP destination vector presenting a Steptavidin/FLAG (SF) at the N-terminal of the protein.

Cell culture and transfection.

HEK293T (human embryonic kidney) cells were grown under normal conditions using DMEM medium with Glutamax from Sigma-Aldrich (Switzerland) supplemented with bovine serum (10%) and antibiotics. The transfection of the SF-N-TAP-FAM161A was obtained using Effectene transfecting reagent from Qiagen (Venlo, The Netherlands). Cells were collected 48 hours after transfection.

Tandem affinity purification

HEK293T cells were transfected with SF-N-TAP-FAM161A and harvested for 20 min at 4°C in a lysis buffer [30 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.5% Nonidet-P40] with the addition of protease inhibitor cocktails and phosphatase inhibitor cocktail 1 and 2 (from Roche Diagnostic, Manheim, Germany). Debris were pelleted with 10K g centrifugation and the supernatant incubated for 2 h at 4°C with Strep-Tactin superflow (IBA, Göttingen, Germany). After 3 washes with wash buffer [30 mM Tris HCl (pH 7.4), 150 mM NaCl, 0.1% Nonidet-NP40 supplemented with phosphatase inhibitors], protein complexes were eluted with 2mM desthiobiotin (IBA) in Tris–Phosphate buffer (TBS). The second purification, that is the specific step of TAP, was performed using anti FLAG-M2 agarose (purchased from Sigma-Aldrich) by incubating them with the eluate of the first purification for 2 h at 4°C. After 3 washes with wash buffer, the complexes were eluted with FLAG peptide (200 μg/ml, Sigma-Aldrich) in TBS. A part of the eluate was then run on SDS-PAGE gel to separate the different bands. The gel was then stained using silver staining according to classical protocols. The remaining sample was then precipitated with chloroform and methanol and stored at -80°C

Mass spectrometry

For the LC-MS/MS analysis, protein precipitates were subjected to tryptic proteolysis [117] and then loaded in the UltiMate 3000 nano HPLC system. The elution was obtained by an acetronitrile gradient. The eluate was then analyzed by mass spectrometry with LTQ Orbitrap XL (Thermo Fisher Scientific, Waltham, USA). The spectra obtained from this test were extracted using Bioworks browser (Thermo Fisher Scientific). The software output was then analyzed using Sequest (vers SRF v.5, ThermoFinningan, San Jose, CA). The setup for this experiment was the following: search in Uniref100 database; trypsin digestion; fragment ion mass tolerance of 1 Da; parent ion tolerance 1 PPM. Results were then loaded in Scaffold (vers. 2.02.01, Proteome Software Inc., Portland, OR) for validation purposes. Criteria for exclusion/inclusion of identified peptides were based on previous publications [118]

Yeast two-hybrid

For Y2H analysis we used a GAL4-based assay HybriZAP from Stratagene (La Jolla, CA, USA). Full-length FAM161A and partial constructs corresponding to N-term and to C-term of FAM161A were fused to a DNA-binding domain (GAL4-BD) and used as

a bait to screen human and bovine retinal cDNA library, retro-transcribed by fresh tissue extracted RNA using oligo-dT primers. Both bait and prey were cloned in the yeast strain Pj69-4A, carrying the *HIS3* (histidine), *ADE2* (adenine), *MEL1* (α-galactosidase) and *LacZ* (β-galactosidase) reporter genes. Interactions were assessed by the activation of reporter genes, such as the ability to grow in a selective media (histidine and adenine deprived media), and colorimetric assays (α-galactoside "on plate" assay and β-galactoside filter lift assay) [119]. A tip was used to collect traces of the colony and a colony PCR was performed using primers specific for the prey plasmid. PCR product was then purified by Millipore PCRclean membrane applying necessary vacuum, and then sequenced using BigDye terminator reagent v1.1 (Applied Biosystems). Capillary electrophoresis was performed using ABI 3130XL from Applied Biosystems, and electropherograms were analyzed using CLC Bio workbench v.5.0. The corresponding protein was identified by blast screening.

Bioinformatic analysis

Gene ontology (GO) annotations were performed by using Database for Annotation, Visualization and Integrated Discovery (DAVID) web interface from National Institute of Allergy and infectious Disease, NIH (www.david.abcc.ncifcrf.gov) [120]. The gene interaction network was generated using GENEMANIA web interface from the university of Toronto (www.genemania.org). Statistical analysis was performed using the Keisan on-line calculator (http://keisan.casio.com)

RESULTS

TAP analysis reveals enrichment in ribosomal protein, with a small portion of cytoskeletal-associated proteins.

We performed the TAP analysis using FAM161A conjugate to a N-term SF tag. The efficiency of transfection was tested by SDS-PAGE, which confirmed the correct expression of the recombinant FAM161A in HEK293T cells (not shown). Peptides obtained by this analysis were reported with associated function and Entrez ID (Table 1). A gene ontology analysis of the positive hits revealed an enrichment of translation proteins (64.2% of the total unique peptides), especially ribosomal proteins, both cellular and mitochondrial ones, and other proteins with diverse functions. Network interaction analysis revealed how the majority of these positive hits were grouped together in the center of the network, whereas a few outliers were identified and they were not linked with any other protein in the network (Fig. 1). These outliers were

DCAF8, EML6, POC1A, CEP78, ASPM, and, connected to the network but located far from the main group, LRRK2, MYO5A and MYO5C. Interestingly, a gene ontology analysis of these outliers revealed an enrichment of proteins involved in cytoskeleton or microtubule organization (Fig. 2 and Table 2).

Yeast two-hybrid screening analysis identifies several interactors of FAM161A involved in protein transport along the microtubule network and proteins that localize at the level of the cilium or basal body.

We performed Y2H screening using all previously generated BD-FAM161A constructs (FAM161A, FAM161A-N-term and FAM161A-C-term) to investigate a human retina and bovine cDNA library. Interestingly the screening of human retina using N-term construct of FAM161A revealed a very low number of positive colonies (13) and many of them did not pass the colorimetric assay (Fig. 3). Among the positive ones, not all passed the PCR screening, probably resulting in false positive results. For this reason, the N-terminal bait was not used for the bovine retina library screening. Conversely, several positive colonies were obtained using an entire FAM161A construct or C-term FAM161A (more than 200 blue colonies per plates, with more colonies for full form compared to the C-term construct). Due to a technical limitation, notably with problems in the quality of the PCR sequences, not all colonies were successfully screened with a higher success rate for the bovine retina screening (Table 3). Positive clones were sequenced with results listed here (Table 4). All the listed genes were positive in the colorimetric assays. Among these genes more than 40% were also present in the ciliary proteome database [121]. This over-representation of ciliary genes is statistically significant [cumulative hypergeometic probabability: 6.3×10^{-6} considering the total number of proteins listed in Swissprot entry mapped to an Entrez gene identifier (18523) and all proteins present in ciliary database–see table 3]. GO analysis for the cellular component of the identified genes showed an enrichment of proteins present in the microtubule cytoskeleton, centrosome and cilium, with significative p-values (Fig. 4a). Additionally, GO analysis for molecular function indicated a possible role in motor activity or structural molecule activity (Fig 4b). Network analysis of the interactions showed a clear interconnection between many of the identified members as well as a few of the outliers. The lacking of functional information in the literature for some of these interactors could explain their location outside the main network (Fig. 4). Three branches of genes were not connected to the main nodes. These isolated small branches were composed of different couples of genes: PLEKHM3 and MPRIP; GAS8 and MAPK8IP1; CCDC155 and C15orf26; NIN and RAB11FIP3.

DISCUSSION

We recently demonstrated that FAM161A is a new ciliary gene by showing the interaction with other ciliopathy-associated proteins. In the previous paper we already performed a binary interaction assay in yeast, although it was a direct assay in which we tested our baits with yeasts expressing a specific prey corresponding to one of the already known ciliary proteins. This technique, even if it is a powerful method to identify transients or weak interactions, does not allow the investigation of the total proteome of the cell. We therefore decided to perform two more comprehensive proteomic techniques to investigate the function of FAM161A.

The first technique was TAP analysis. With this system we overexpressed our protein of interest in standard laboratory cell lines and isolated all the interactors by a simple immunoprecipitation. Here, in contrast to Y2H screening in which direct binary interactions are tested, large molecular complexes can be immunoprecipitated, allowing the identification of non-direct interactions. This technique uses two consecutive purification steps with two different tags in order to obtain cleaner results and fewer proteins attached to the complexes, so they can easily be identified by mass spectrometry [122]. TAP is very efficient, although it can be somewhat biased; firstly, by using the overexpression machinery, the expression rate of the protein is far greater than that normally observed, which can lead to mislocalization results in the immunoprecipitate. Secondly, many false positive results are connected to the stickiness of the TAG or of the expressed protein, which could lead to the formation of complexes with endogenous proteins that are much more abundant in cells compare to our specific proteins of interest. Thirdly, this system is inefficient for testing transient interactions.

In our experiments we observed an enrichment of ribosomal proteins. Although some of these were also identified in other ciliary proteomic studies [123], this result is unexpected and does not agree with previously observed experimental data. A possible explanation for this is that ribosomal complexes necessary for the translation of the protein remained attached to the growing protein and they were immunoprecipitated as complex. Even though we cannot exclude that these interactions are real, the possibility that they were due to a technical issue is quite large. Another analysis, using a C-terminal TAG, could help to clarify this point. Nonetheless, with the exception of ribosomal complexes, only a few hits were completely outside the large network map and these seemed to be the most interesting for further investigation. They were mostly associated to microtubule binding and centrosome/basal body activity, reinforcing the

previously observed data on FAM161A function. For example POC1A, a gene important for centriole and basal body structure [124], has been recently associated to a new ciliopathy, with dwarfism as a main symptom [125]. A confirmation of these data with further biochemical studies is required to finally elucidate the function of FAM161A.

The second technique we performed was the Y2H screening. With this system it is possible to identified direct binary interactions between two proteins. The strength of this technique is to use a specific library of cDNA from the tissue of interest; in our case we used the retina extracted from bovine and from human. For the analysis we used previously generated constructs as baits. Interestingly no positive colonies were obtained using the N-terminal portion of FAM161A, confirming the importance of the UPF0564 domain for interactions. Results obtained from this analysis revealed an enrichment of proteins for the cytoskeleton network or in the cilia basal body. In addition to this we saw a statistically significant 3 fold enrichment of ciliary proteins, based on data extracted from the ciliary proteomic database, relative to the entire proteome. These data confirmed the observation that FAM161A is a ciliary protein. Furthermore, a gene ontology analysis revealed that a significant number of identified hits correspond to proteins involved in molecular transport through the microtubule network. These results could shed new light on the function of FAM161A. In fact due to the large number of interactions already observed with proteins present in a different compartment of photoreceptor cilium it is possible to suppose that FAM161A has a role in intraflagellar transport (IFT) or vesicle transport. In addition, the overall picture from the interaction network generated from Y2H positive hits, showed an unexpectedly high degree of interconnections among the observed hits. Several, if not all, of the identified FAM161A interactors localize to the connecting cilium, or at centrosomes in non-ciliated cells. Biochemical studies, as well as *in vivo* studies on mouse models for some of these genes, could help to clarify the real nature of these interactions. Another interesting interactor identified here is *KIF3C*. It is the third important member of the Kinesin-II complex involved in ciliogenesis. Although its role in mouse photoreceptor cilia is not clear, since a complete KO mouse showed a perfectly healthy phenotype[126], it has been demonstrated to interact with KIF3A for membrane vesicle transport [127]. Moreover KIF3C was identified in several ciliary proteomes and a recent paper showed the importance of this protein for zebrafish ciliogenesis [128]. Whereas a single KO fish for *Kif3b* or *Kif3c* seems not to show any defects in photoreceptor cilium, a double KO for both kinesin proteins displays complete loss of photoreceptor cilia [128]. In addition in this work it was demonstrated the redundancy of kif3b and kif3c functions, since the

recovery of only one of the proteins is enough to revert the phenotype. Interestingly it was also observed a different pattern of expression of kif proteins during cilium development. While kif3b is necessary for early stages of ciliogenesis, the presence of kif3b and kif3c is needed in adult stages, probably indicating a change of cargo loading in different stages of photoreceptor life. It is interesting to observe that FAM161A, a protein that is expressed mainly during at the adult stage, interacts with KIF3C which also has a role in adulthood.

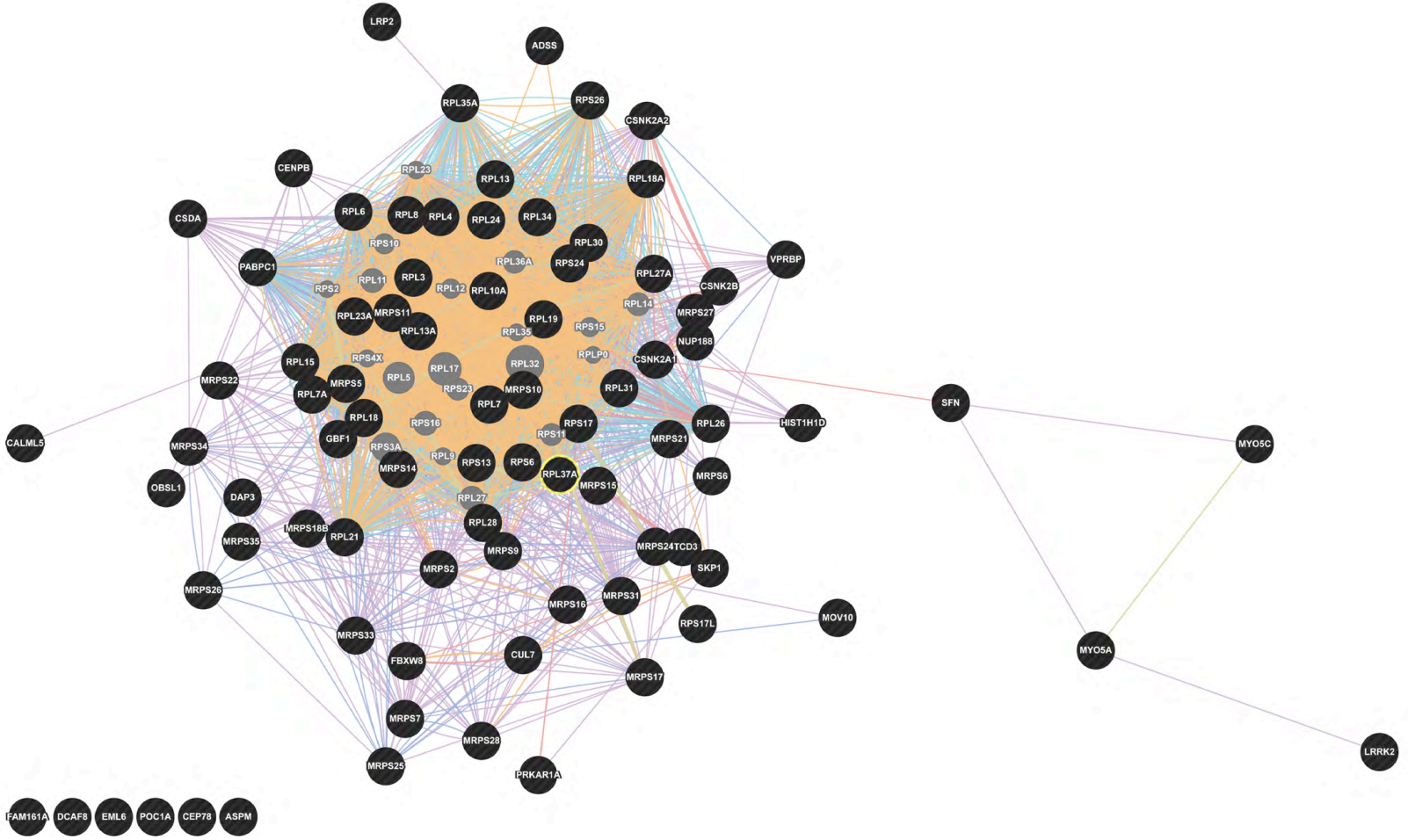


Figure 1. Network of interactions of TAP positive hits.
Graphical view of all the interactions between proteins immunoprecipited with FAM161A from TAP experiments. It is possible to identify the outliers that are indicated in the bottom left of the picture. Other outliers are present at the right side of the map. Purple lines indicate when they are found to be affected together in co-expression studies, light green indicates that are found in the same pathway, yellow indicates that they are sharing the same protein domains.

Table 1. TAP analysis results. Here are reported only proteins already filtered for typical false positive results, found in TAP analysis using our internal database as control set. In green are enlighten all the genes found also in the ciliary proteome (ciliaproteome.org, v3).

Entrez ID	Gene Symbol	Gene full name	Tap bait count	Localization
28957	MRPS28	mitochondrial ribosomal protein S28	10	Mitochondrion.
7818	DAP3	death associated protein 3	9	Mitochondrion.
51081	MRPS7	mitochondrial ribosomal protein S7	6	Mitochondrion (By similarity).
56945	MRPS22	mitochondrial ribosomal protein S22	8	Mitochondrion.
28973	MRPS18B	mitochondrial ribosomal protein S18B	5	Mitochondrion.
55173	MRPS10	mitochondrial ribosomal protein S10	4	Mitochondrion.
60488	MRPS35	mitochondrial ribosomal protein S35	5	Mitochondrion (By similarity).
23107	MRPS27	mitochondrial ribosomal protein S27	6	Mitochondrion.
51116	MRPS2	mitochondrial ribosomal protein S2	5	Mitochondrion.
10240	MRPS31	mitochondrial ribosomal protein S31	8	Mitochondrion.
64965	MRPS9	mitochondrial ribosomal protein S9	7	Mitochondrion.
64969	MRPS5	mitochondrial ribosomal protein S5	4	Mitochondrion.
64949	MRPS26	mitochondrial ribosomal protein S26	7	Mitochondrion.
55037	PTCD3	pentatricopeptide repeat domain 3	6	Mitochondrion (Potential).
6160	RPL31	ribosomal protein L31	5	
6194	RPS6	ribosomal protein S6	11	
51373	MRPS17	mitochondrial ribosomal protein S17	6	Mitochondrion.
6128	RPL6	ribosomal protein L6	7	
64960	MRPS15	mitochondrial ribosomal protein S15	4	Mitochondrion.
6143	RPL19	ribosomal protein L19	8	
1459	CSNK2A2	casein kinase 2, alpha prime polypeptide	9	
1457	CSNK2A1	casein kinase 2, alpha 1 polypeptide	11	
1460	CSNK2B	casein kinase 2, beta polypeptide	11	
64968	MRPS6	mitochondrial ribosomal protein S6	2	Mitochondrion.
51806	CALML5	calmodulin-like 5	10	
6154	RPL26	ribosomal protein L26	8	
51021	MRPS16	mitochondrial ribosomal protein S16	2	Mitochondrion.
65993	MRPS34	mitochondrial ribosomal protein S34	5	Mitochondrion.
6137	RPL13	ribosomal protein L13	8	
6138	RPL15	ribosomal protein L15	7	
6142	RPL18A	ribosomal protein L18a	6	
6207	RPS13	ribosomal protein S13	7	
6231	RPS26	ribosomal protein S26	10	
6229	RPS24	ribosomal protein S24	4	
4736	RPL10A	ribosomal protein L10a	6	
6141	RPL18	ribosomal protein L18	9	Cytoplasm.
64432	MRPS25	mitochondrial ribosomal protein S25	5	Mitochondrion.
6129	RPL7	ribosomal protein L7	8	
6158	RPL28	ribosomal protein L28	7	
6165	RPL35A	ribosomal protein L35a	4	
50717	DCAF8	DDB1 and CUL4 associated factor 8	8	
5573	PRKAR1A	protein kinase, cAMP-dependent, regulatory, type I, alpha (tissue specific extinguisher 1)	6	
6147	RPL23A	ribosomal protein L23a	5	
220717	RPLP0P6	ribosomal protein, large, P0 pseudogene 6	10	
6132	RPL8	ribosomal protein L8	4	Cytoplasm.
23521	RPL13A	ribosomal protein L13a	5	
26986	PABPC1	poly(A) binding protein, cytoplasmic 1	5	Cytoplasm. Nucleus.
6122	RPL3	ribosomal protein L3	5	Nucleus, nucleolus. Cytoplasm.
9730	VPRBP	Vpr (HIV-1) binding protein	7	Cytoplasm. Nucleus.
4644	MYO5A	myosin VA (heavy chain 12, myoxin)	4	
259266	ASPM	asp (abnormal spindle) homolog, microcephaly associated (Drosophila)	6	Cytoplasm (By similarity). Nucleus (By similarity)
84140	FAM161A	family with sequence similarity 161, member A	2	
54460	MRPS21	mitochondrial ribosomal protein S21	3	Mitochondrion.
6156	RPL30	ribosomal protein L30	4	
63931	MRPS14	mitochondrial ribosomal protein S14	2	Mitochondrion.
6168	RPL37A	ribosomal protein L37a	2	
64963	MRPS11	mitochondrial ribosomal protein S11	1	Mitochondrion.
6152	RPL24	ribosomal protein L24	11	
6144	RPL21	ribosomal protein L21	1	
6500	SKP1	S-phase kinase-associated protein 1	6	
6164	RPL34	ribosomal protein L34	2	

6218	RPS17	ribosomal protein S17	5	
1E+08	RPS17L	ribosomal protein S17-like	5	
				Cytoplasm. Nucleus (By similarity).
2810	SFN	stratifin	5	Secreted.
3007	HIST1H1D	histone cluster 1, H1d	8	Nucleus. Chromosome.
4343	MOV10	Mov10, Moloney leukemia virus 10, homolog (mouse)	1	Cytoplasm, P-body.
51650	MRPS33	mitochondrial ribosomal protein S33	3	Mitochondrion.
64951	MRPS24	mitochondrial ribosomal protein S24	1	Mitochondrion (By similarity).
6157	RPL27A	ribosomal protein L27a	5	
6130	RPL7A	ribosomal protein L7a	5	
9820	CUL7	cullin 7	2	Cytoplasm.
8531	CSDA	cold shock domain protein A	3	Cytoplasm. Nucleus.
26259	FBXW8	F-box and WD repeat domain containing 8	1	
				Cytoplasm, cytoskeleton, centrosome, centriole. Cytoplasm, cytoskeleton,
25886	POC1A	POC1 centriolar protein homolog A (Chlamydomonas)	1	cilium basal body.
1059	CENPB	centromere protein B, 80kDa	1	Nucleus. Chromosome, centromere.
84131	CEP78	centrosomal protein 78kDa	1	Cytoplasm, cytoskeleton, centrosome.
6124	RPL4	ribosomal protein L4	5	
55930	MYO5C	myosin VC	4	
				Nucleus, nuclear pore complex
23511	NUP188	nucleoporin 188kDa	3	(Probable).
				Cytoplasm. Note=Partially associated
159	ADSS	adenylosuccinate synthase	2	with particulate fractions.
23363	OBSL1	obscurin-like 1	2	
		golgi brefeldin A resistant guanine nucleotide exchange		
8729	GBF1	factor 1	3	
				Cytoplasm. Membrane; Peripheral
120892	LRRK2	leucine-rich repeat kinase 2	3	membrane protein. Mitochondrion.
400954	EML6	echinoderm microtubule associated protein like 6	5	Cytoplasm, cytoskeleton (Potential).
				Membrane; Single-pass type I membrane protein. Membrane,
4036	LRP2	low density lipoprotein receptor-related protein 2	1	coated pit.

Not present in ciliary database
Present also in ciliary database

Table 2. Outliers found in TAP analysis with cellular localization and function

Gene name	Localization	Function (based on Uniprot)	Protein Domain
DCAF8	peroxisome	Could be involved in CUL4-DDB1 E3 ubiquitin protein ligase complex	WD40
EML6	Cytoplasm/microtubule	May modify the assembly dinamics of microtubule (by similarity)	WD40
POC1A	Cytoplasm/Basal body	Deficiency cause severe ciliogenesis defect	WD40
CEP78	Cytoplams/Centrosome	Not known	LRR
ASPM	Cytoplasm	Regulation of the mitotic spindle (By similarity)	CH
LRRK2	Cytoplasm/trans Golgi	May play a role in the phosphorilation of proteins central to Parkinson Disease	LRR, DFG, RAS, WD40
MYO5A	Cytoplasm/trans Golgi	Actin- based motor. Involved in dendrite formation	-
MYO5C	Cytoplasm/trans Golgi	Actin- based motor. Involved in dendrite formation	-

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
	GOTERM_CC_FAT	cytoskeletal part	RT		4	57.1	1.6E-3	7.6E-2
	GOTERM_CC_FAT	cytoskeleton	RT		4	57.1	4.6E-3	1.1E-1
	GOTERM_CC_FAT	microtubule cytoskeleton	RT		3	42.9	1.0E-2	1.6E-1
	GOTERM_CC_FAT	intracellular non-membrane-bounded organelle	RT		4	57.1	2.8E-2	3.1E-1
	GOTERM_CC_FAT	non-membrane-bounded organelle	RT		4	57.1	2.8E-2	3.1E-1

Figure 2. GO enrichment of genes from table 2.
Enrichment analysis performed using DAVID bioinformatics software. It is shown enrichment for protein involved in cytoskeleton structure and microtubule associated proteins.

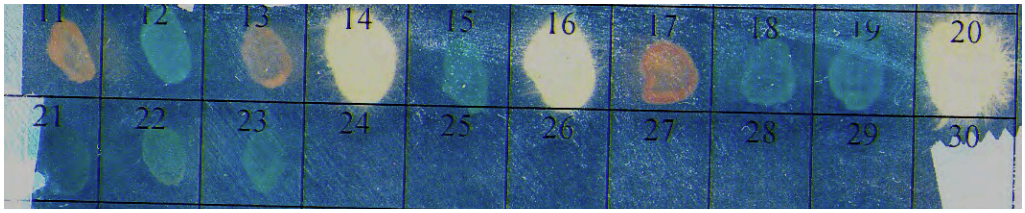


Figure 3. Colorimetric assay for N-term FAM161A Y2H screening.
In this diagram colonies obtained by Y2H screening of the human retina library using a FAM161A-N-term construct are shown. It is possible to appreciate that only a few colonies passed the colorimetric assay by α-galactosidase reaction.

Table 3. Total number of screened colonies.

	Total number of selected colonies	Successfully sequenced colonies	Percentage of succesfully sequenced
HUMAN RETINA-FAM161A_FULL	247	132	53.44%
HUMAN RETINA-C-term_FAM161A	37	23	62.16%
BOVINE RETINA-FAM161A_FULL	140	94	67.14%
BOVINE RETINA-C-term_FAM161A	55	33	60.00%

a.

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
	GOTERM_CC_FAT	microtubule cytoskeleton	RT		9	19.6	5.9E-5	4.6E-3
	GOTERM_CC_FAT	cytoskeleton	RT		13	28.3	9.8E-5	3.9E-3
	GOTERM_CC_FAT	cell projection	RT		9	19.6	3.1E-4	8.1E-3
	GOTERM_CC_FAT	microtubule organizing center	RT		6	13.0	4.4E-4	8.7E-3
	GOTERM_CC_FAT	cytoskeletal part	RT		10	21.7	5.1E-4	8.0E-3
	GOTERM_CC_FAT	kinesin complex	RT		3	6.5	9.6E-4	1.3E-2
	GOTERM_CC_FAT	centrosome	RT		5	10.9	2.5E-3	2.8E-2
	GOTERM_CC_FAT	non-membrane-bounded organelle	RT		15	32.6	3.2E-3	3.1E-2
	GOTERM_CC_FAT	intracellular non-membrane-bounded organelle	RT		15	32.6	3.2E-3	3.1E-2
	GOTERM_CC_FAT	cilium	RT		4	8.7	4.4E-3	3.8E-2
	GOTERM_CC_FAT	microtubule	RT		5	10.9	5.2E-3	4.0E-2
	GOTERM_CC_FAT	Golgi apparatus	RT		8	17.4	6.0E-3	4.2E-2
	GOTERM_CC_FAT	microtubule organizing center part	RT		3	6.5	9.1E-3	5.9E-2
	GOTERM_CC_FAT	neuron projection	RT		5	10.9	1.1E-2	6.6E-2
	GOTERM_CC_FAT	microtubule associated complex	RT		3	6.5	2.8E-2	1.5E-1
	GOTERM_CC_FAT	proteinaceous extracellular matrix	RT		4	8.7	4.9E-2	2.3E-1
	GOTERM_CC_FAT	extracellular matrix	RT		4	8.7	5.9E-2	2.6E-1
	GOTERM_CC_FAT	microtubule basal body	RT		2	4.3	8.9E-2	3.5E-1

b.

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
	GOTERM_MF_FAT	microtubule motor activity	RT		3	6.5	1.5E-2	7.8E-1
	GOTERM_MF_FAT	structural molecule activity	RT		6	13.0	1.8E-2	5.9E-1
	GOTERM_MF_FAT	enzyme binding	RT		5	10.9	3.8E-2	7.2E-1
	GOTERM_MF_FAT	motor activity	RT		3	6.5	4.8E-2	6.9E-1
	GOTERM_MF_FAT	protein kinase binding	RT		3	6.5	5.1E-2	6.3E-1
	GOTERM_MF_FAT	kinase binding	RT		3	6.5	7.2E-2	7.0E-1

Figure 4. GO enrichment analysis of Y2H positive hits.
a) Analysis of cellular component reveals an enrichment of cytoskeleton and centrosomal compartment. The numbers of genes indicated for each subcategory is reported, as well the P-value by a EASE score using a modified Fisher test exact P-value. **b)** Analysis with molecular function by gene ontology terms, revealed a possible function in microtubule motor activity.

Table 4. List of positive hits found by Y2H screening of human and bovine retina cDNA library.

Gene name, EntrezID, cellular localization and function based on Uniprot database is reported. A cross indicates in each column if the hit was found in human retina (Human column), bovine retina (Bovine column), and if it was obtained using BD-FAM161A (Full) or FAM161A-C-term (C-term) as bait. In green are the genes found to be present in the ciliary proteome. At the bottom the statistical calculation for the ciliary protein enrichment is displayed.

Entrez ID	Gene	Human	Bovine	Full	C-term	Cellular localization	Function _based on Uniprot
2202	EFEMP1	x	x	x	x	Extracellular matrix	Extracellular matrix composition and Epidermal growth factor receptor
22981	NINL	x		x		Cytosol, microtubules, MOC, centrosome	Microtubule organization in interphased cells, interphere with mitotic spindle assembly.
1039	CDR2	x	x	x	x	Cytoplasm, nucleus	Immunological activation
9240	PNMA1	x	x	x		Nucleus, nucleulus, cytoplasm, focal adesions	
6744	SSFA2	x		x		Cytoplasm,cytoscheleton, membrane	Ephitelial cell polarity
51199	NIN	x		x		Cytoplasm, centrosome, cytoskeleton, centriole	Centrosome formation
4217	MAP3K5	x		x	x	Endoplasmic reticulum	MAPkinase involved in signal transduction
4753	NELL2	x	x		x	Cytoplasm, secreted	Glycoprotein. Involved in neural cell growth
2896	GRN	x	x	x	x	Extracellular matrix, mithocondrion, intracellular membrane organelle	Cell growht control
3913	LAMB2	x		x	x	Secreted, extracellular space, extracellular matrix, basement membrane	Laminins, a family of extracellular matrix glycoproteins, are the major noncollagenous constituent of basement membranes. They have been implicated in a wide variety of biological processes including cell adhesion, differentiation, migration, signaling, neurite outgrowth and metastasi
64753	CCDC136	x	x	x	x	Membrane; Single-pass membrane protein (Potential)	Found to be upregulated in cone only rd17 mouse
9782	MATR3	x		x		Nucleus matrix	may play a role in transcription or may interact with other nuclear matrix proteins to form the internal fibrogranous network
10241	CALCOCO2	x		x		Cytoplasm, perinuclear region. Golgi apparatus. Cytoplasm, cytoskeleton	May play a role in ruffle formation and actin cytoskeleton organization. Maybe important for transport?
84140	FAM161A	x	x	x	x		
374650	GOLGA6L5	x		x			
	FLJ94640	x		x		Cytoplasm, perinuclear region. Nucleus, nucleolus	Similar to KRT18 an intermediate filament protein
9922	IQSEC1	x		x		Cytoplasm. Nucleus	Actine cytoskeleton riorganization .in addition to accelerate GTP gamma S binding by ARFs of all three classes, it appears to function preferentially as a guanine nucleotide exchange protein for ARF6, mediating internalisation of beta-1 integrin
54550	NECAB2	x		x		Cytoplasm	EF-calcium binding protein. Function unknown
25957	PNISR	x		x		Nucleus speckle	SR protein. Function unknown
9321	TRIP11	x		x		Associates with the ends of centrosome-nucleated microtubules	Probably involved in maintaining cis-Golgi structure around centrosomes: overexpression of this protein block retrograde and anterograde transport between golgi and ER
9659	PDE4DIP		x	x	x	Golgi apparatus (By similarity). Cytoplasm, cytoskeleton, centrosome	The protein encoded by this gene serves to anchor phosphodiesterase 4D to the Golgi/centrosome region of the cell. It works as an IS cargo between golgi and centrosome
80184	CEP290		x	x		Cytoplasm, cytoskeleton, centrosome. Nucleus. Cell projection, cilium. Cytoplasm, cytoskeleton,cilium basal body	may play a role in ciliary transport processes
9479	MAPK8IP1		x	x		Cytoplasm (By similarity). Cytoplasm, perinuclear region	May function as a regulator of vesicle transport, through interactions with the JNK-signaling components and motor proteins
7431	VIM		x	x		Cytoplasm	Vimentins are class-III intermediate filaments found in various non-epithelial cells, mesenchymal cells
84034	EMILIN2		x	x		Secreted, extracellular space, extracellular matrix	May be responsible for anchoring smooth muscle cells to elastic fibers
7399	USH2A		x	x	x	Cell projection, stereocilium membrane	may be important in development and homeostasis of the inner ear and retina
22891	ZNF365		x	x		localizes to the centrosome at all stages of the cell cycle	Overexpression of ZNF365A caused abnormal mitosis, and a mutant form of ZNF365A lacking the C-terminal region disrupted localization of gamma-tubulin to the centrosome
56254	RNF20		x	x		Nucleus (Probable)	E3 ubiquitin ligase that regulates chromosome structure by monoubiquitinating histone H2B.

140710	SOGA1		x	x		Secreted (By similarity)	Regulates autophagy by playing a role in the reduction of glucose production in an adiponectin- and insulin-dependent manner
59274	MESDC1		x	x		Not known	Unknown
285527	FRYL		x	x		Not known (nucleus)	Plays a key role in maintaining the integrity of polarized cell extensions during morphogenesis (by similarity)
4163	MCC		x	x	x	Cell membrane. Nucleus. Cytoplasm	a phosphoprotein associated with the plasma membrane and membrane organelles, regulates also lamellipodia
11190	CEP250		x	x		Cytoplasm, perinuclear region. Cytoplasm, cytoskeleton, centrosome, centriole. Cytoplasm,cytoskeleton, cilium basal body	Probably plays an important role in centrosome cohesion during interphase
66008	TRAK2		x	x	x	Cytoplasm (By similarity). Early endosome (By similarity). Mitochondrion (	May regulate endosome-to-lysosome trafficking of membrane cargo, including EGFR
23338	PHF15		x	x		Not known (nucleus)	Histone acetylase complex (not known)
84437	MSANTD4		x	x		Not known	Unknown
9727	RAB11FIP3		x	x		Recycling endosome membrane; Peripheral membrane protein. Cytoplasm, cytoskeleton, centrosome.Cleavage furrow. Midbody.	Acts as a regulator of endocytic traffic by participating in membrane delivery
10142	AKAP9		x	x	x	Cytoplasm. Cytoplasm, cytoskeleton, centrosome. Golgi apparatus	Scaffolding protein that assembles several protein kinases and phosphatases on the centrosome and Golgi apparatus
2802	GOLGA3		x	x		Cytoplasm. Golgi apparatus, Golgi stack membrane; Peripheral membrane protein	play a role in transport and Golgi apparatus localization, Important for vescicular transport.
89953	KLC4		x	x		Cytoplasm, cytoskeleton	Kinesin is a microtubule-associated force-producing protein that may play a role in organelle transport
23164	MPRIP	x	x	x		Cytoplasm, cytoskeleton	Targets myosin phosphatase to the actin cytoskeleton
7109	TRAPPC10		x	x		Golgi apparatus, cis-Golgi network	Multisubunit transport protein particle (TRAPP) complex and may be involved in vesicular transport from theendoplasmic reticulum to the Golgi
2622	GAS8		x	x		Cytoplasm, cytoskeleton (By similarity). Cytoplasm, cytoskeleton, flagellum basal body (By similarity). Golgi apparatus (By similarity).	Cytoskeletal linker which binds microtubules and probably functions in axonemal and non-axonemal dynein regulation
389072	PLEKHM3		x	x		Not known	Unknown
7110	TMF1		x	x		Cytoplasm. Nucleus. Golgi apparatus membrane	May play critical roles in two RAB6-dependent retrograde transport processes: one from endosomes to the Golgi and the other from the Golgi to the ER
161502	C15ORF26		x	x		Not known	Unknown
147872	CCDC155		x	x		Not known (membrane)	Unknown
3797	KIF3C		x	x		Cytoplasm, cytoskeleton (Probable)	Microtubule-based anterograde translocator for membranous organelles
64837	KLC2		x	x		Cytoplasm, cytoskeleton (Probable)	Kinesin is a microtubule-associated force-producing protein that may play a role in organelle transport.
57698	KIAA1598		x	x		Cell projection, axon (By similarity)	in a single neurite before polarization, leading to axon induction for polarization
257194	NEGR1		x		x	Cytoplams	Maybe involved in neuronal cell adhesion
57568	SIPA1L2	x		x		Not known	Unknown
3320	HSP90AA1	x		x		Cytosol	Heat shock protein

Not present in ciliary proteome database
Present in ciliary proteome database

Total protein in Swissprot database	Total protein in the ciliary database	Total Hits in Y2H assay	Total hits in Y2H present in the ciliary database	Fold enrichment	Cumulative Hypergeometric probability
18523	2726	52	21	2.6	6.3×10^{-6}

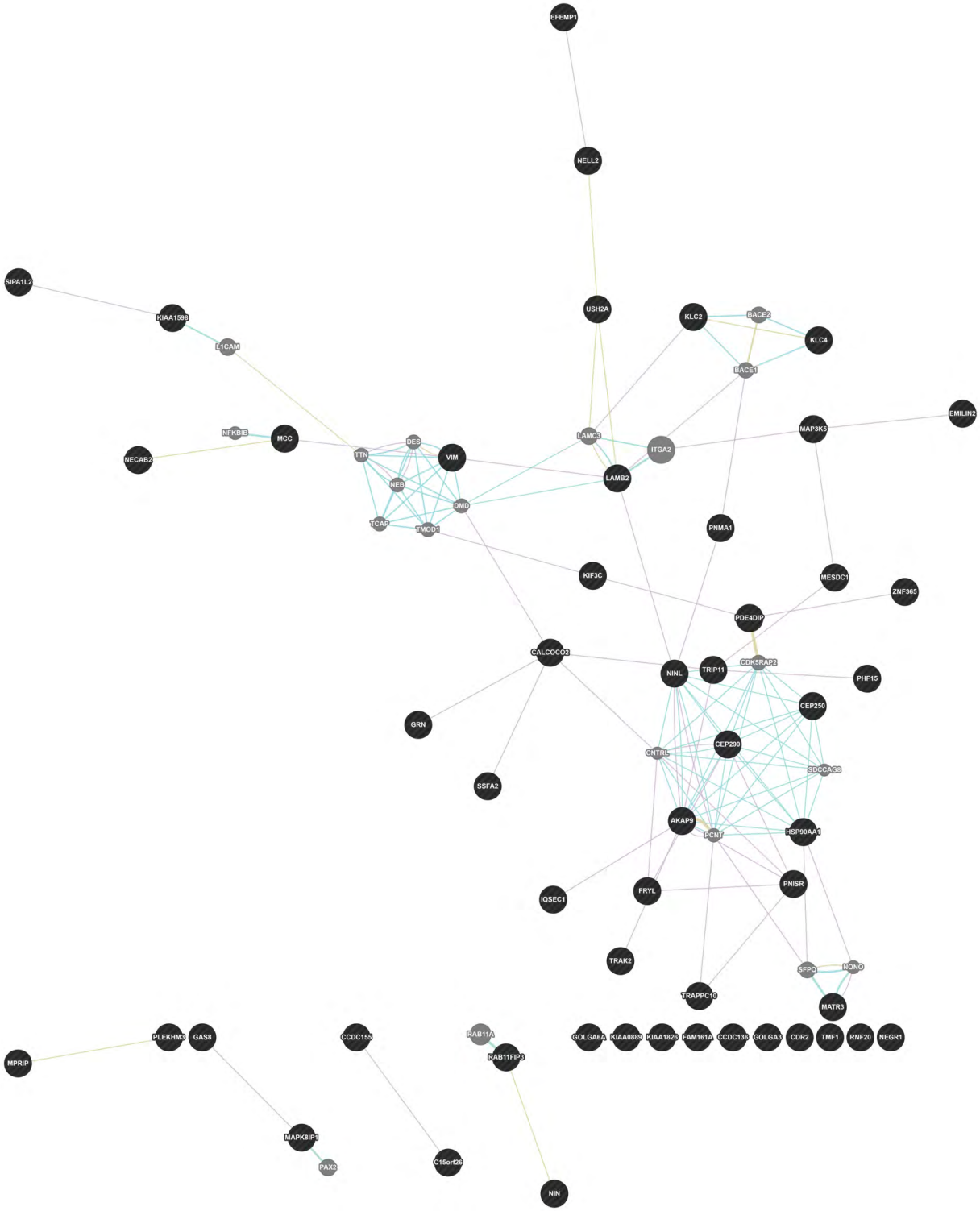


Figure 5. Interaction network of Y2H positive hits.
Purple lines indicate when two proteins are found together in co-expression studies, light green when are found in the same pathway, yellow when they are sharing the same protein domains. Grey circles are the connecting genes not directly identified by this screening.

I.6 DISCUSSION AND PERSPECTIVES

In this first part of the dissertation we demonstrate that non-sense mutations in *FAM161A* cause retinitis pigmentosa in humans, and we discovered that the protein encoded by this gene is a new ciliary protein, with a possible role in cellular and ciliary trafficking.

FAM161A is a member, alongside FAM161B, of the FAM161 family found in mammals. They share a specific and unique, large and uncharacterized domain at the C-terminal of the protein, called UPF0564. Multiple alignments for this domain show a widespread conservation among different species and phyla, from complexes organism, to parasites, unicellular eukaryotes or prokaryotes, such as yeasts and simple bacterial cells or unicellular algae. Interestingly, a search for protein containing UPF0564 in different species revealed that this is rarely present in association with other domains in these proteins, suggesting an unconventional specialization for this domain. Furthermore, it covers, on average, more than 60% of the entire protein and, a search for the UPF0564 domain across multiple species showed enrichment in cilia or flagellated organisms (among the kingdom Plantae, for example, it is only conserved in the flagellate algae Chlamydomonas, the model organism for IFT) confirming the possible role of this protein family in cilium or flagellum function. In our work we clearly show how the UPF0564 domain is necessary for interactions with FAM161A partners and for linking with the microtubule network, demonstrating the importance of this domain in protein function.

We identified mutations in *FAM161A* as the cause of *RP28*-associated retinitis pigmentosa. The original *RP28* family had a homozygous p.Arg229X in exon 3, whereas we identified p.Arg437X in a homozygous state in affected patients from three families from Germany. Although these families were unrelated and came from different regions, they presented the same haplotype in the *FAM161A* gene, suggesting a founder effect for this mutation. Relevant to our paper, a publication from the Sharon group in Israel [129] demonstrated that mutations in *FAM161A* is one of the most common causes for autosomal recessive retinal degeneration among patients from Israel and the Palestinian territories, with a frequency of almost 12% among affected individuals. They identified several families with a frameshift mutation, p.Thr452SerfsX3, on exon 3, and two other non-sense variants specific for these populations, p.Arg523X and p.Arg596X,

in exon 3 and 5, respectively. Interestingly, these mutations were found in each of the three different ethnic groups from Israel and the Palestinian territories, demonstrating a relatively ancient origin of these mutations. We identified the same mutation, p.Thr452SerfsX3, in a homozygous state in two different patients of our cohort from United States. It is likely that these two patients had an Ashkenazy Jewish ancestry, demonstrating the founder effect of these changes. Interestingly, all the mutations reported so far are frameshift or non-sense located in the middle of the gene. *In silico* analysis revealed that they should cause the complete knock-out of the transcript, due to mRNA degradation induced by the Non-sense mediated mRNA decay (NMD) machinery. We confirmed this hypothesis showing that by inhibiting NMD machinery with emetine or cycloexamide in lymphoblastoid cell lines derived from patients carrying the p.Thr452SerfsX3 mutation we induced the recovery of the *FAM161A* transcript to the level of healthy controls (not shown). This demonstrated that this mutation generated a null allele, as with other similar mutations that induce a premature stop in the transcript. So, the complete depletion of *FAM161A* seems necessary to cause disease.

In addition to the genetic study, we characterized the expression profile of *FAM161A*. We found this gene to be expressed in different tissues both in human and mouse, although the highest expression was observed in the retina and the testis. The testis expresses more than 70% of all human transcripts at a high level [130], therefore for our first analysis we considered it as background noise. However, according to our new findings of the role of FAM161A in cilium, maybe the importance of this gene in testis and other ciliated cells should be reconsidered. The retina specificity of *FAM161A* was also confirmed by the identification a two specific *Cone-Rod homeoboX (CRX)* binding domains at in the promoter and first intron of the gene. *CRX* is a photoreceptor specific transcription factor active during development as well as for maintaining adult retinal cell function[131]. Mutations in *CRX* cause different retinal diseases, such as cone-rod dystrophy and LCA [132, 133]. Several genes involved in retina function and which are mutated in diseases of retinal degeneration are under the control of *CRX* [134], demonstrating the importance of this gene in maintaining the expression of retinal related genes during adult photoreceptor life. Although several ciliopathy genes (*CEP290, RP1, RPGRIP1…)* have a CBR in their promoter sequences as identified by *CRX*-specific ChIP-seq studies [134], the importance of this transcription factor in cilium biology has not been well studied up to now. Recently, the discovery that the expression of another cilia associated gene *MAK* [135] is driven by *CRX,* suggests an active role of this transcription factor in maintaining and expressing ciliary genes in photorceptors.

In our studies we also performed a co-localization analysis to elucidate the function of *FAM161A* in the retina structure. Our first analysis identified the protein signal in the IS and within the OPL in adult murine retina, whereas the signal was absent in mice younger than P10. This pattern matches the mRNA expression of *fam161a* during mouse development, starting at P10 when the OS is developing. Subsequently we refined the localization of FAM161A to the BB, proximal CC and partially in the IS. The increase of resolution was achieved using freshly dissected and unfixed murine retinas giving a better preservation of the intracellular structures, and in particular of the ciliary region, compared to paraformaldehyde or methanol fixed tissues. FAM161A co-localized with specific basal body markers, such as RPGRIP1L and partially with centrin, a marker for the whole connecting cilium. The lack of co-localization with RP1, a marker for the ciliary axoneme, demonstrated that the protein has a restricted expression to the IS part of the cilium. Clear staining was also observed in the IS, in particular at the interface between the IS and the ONL. This area corresponds roughly to the location of the Golgi apparatus, where all protein involved in photoreceptor function are assembled, and then transported to the other cellular compartments through microtubule-based vesicle transport. Our finding of several different proteins involved in Golgi-centrosome/BB transport by yeast-2-hybrid screening using FAM161A as bait for bovine and human retina library could somehow explain this peculiar localization. An intriguing hypothesis is that FAM161A could act as a scaffold molecule for the transport of specific complexes between these two different compartments of the cell and then partially across the cilium with the IFT complex. Although data we currently have are insufficient to confirm this hypothesis, they give some further indication. For example, we clearly demonstrated that FAM161A interacts with lebercilin, a ciliary protein encoded by the gene LCA5, and associated with LCA. Recently it was demonstrated a direct interaction between lebercilin and the IFT complex [136]. Moreover mutations in LCA5 induce the disruption of the IFT complex with subsequently problems in OS formation [136]. In our direct 1-on-1 screening assays we observed the interaction between FAM161A and two very distinct IFTs, IFT20 and IFT88. In contrast to other IFT proteins, IFT20 is localized in the Golgi complex, as well as in the BB [137], and a direct role of this IFT in vesicle formation between the Golgi and BB has been shown, since the depletion of photoreceptor cells causes an accumulation of opsin in the IS [60]. Conversely, IFT88 is a typical IFT protein found in the BB and it has been shown to interact with many other ciliary proteins involved in retinal degeneration, such as RPGR [138]. The depletion of IFT88 causes defects in cilium elongation as well as the accumulation of opsin molecules in the IS [59]. Although these interactions have not yet been confirmed by biochemical

studies, an intriguing hypothesis is that FAM161A, and consequently the UPF0564 protein complex could be a new member of the IFT machinery. To support this hypothesis, the identification of several hits by yeast-2-hybrid screening for proteins involved in transport either from the Golgi or the cilium, such as kinesin-2, and in particular the interaction with KIF3C, suggest that investigations should continue in this direction. This hypothesis, could also explain the observed interactions between FAM161A and other ciliary protein such as CEP290, CEP250 or SDCCAG8. The correct localization of these proteins in the BB or CC is fundamental to understand their function, as it has been well demonstrated for CEP290. Specifically, the knockdown of CEP290 destroys the subcellular localization of different molecular complexes involved in cilia regulation, such as PCM1 and CP110 [77]. In addition to this, CEP290 has been shown to directly interact with several proteins that we identified in the FAM161A interactome, such as NIN [74]. From our data, however, we cannot conclude whether CEP290 is necessary for FAM161A localization and function, or vice-versa. The analysis of FAM161A localization in the *rd16* mouse, carrying a large in-frame deletion in the CEP290 gene, could help to elucidate this point. Moreover, the IFT process is necessary for the elongation and formation of a new cilium during ciliogenesis, since the reduction or knockout of IFT proteins is often associated with defective cilium elongation [139, 140]. For FAM161A, we observed, although with less intensity, a reduction in the number of ciliated cells, probably indicating an impairment of ciliogenesis. This milder effect could be related to the specific cell line in which siRNA silencing was performed, which are not ciliated photoreceptor cells. A knockout mouse model of FAM161A would definitely help to clarify the pathogenic processes caused by the lack of this protein.

Relevant to our paper, another publication investigated the role of FAM161A in both cell lines and photoreceptors [114]. Using electron microscopy, the authors observed the localization of FAM161A in the BB and connecting cilium, which is consistent with our data. In addition they emphasized a possible role of FAM161A as a microtubule binding protein, and in particular as a microtubule stabilizing protein. These results were obtained by the ectopic expression of recombinant FAM161A in ciliated cells, where they observed a staining corresponding to FAM161A all over the microtubule network. They showed a direct binding with tubulin, although they were not able to demonstrate the direct involvement of FAM161A in tubulin modulation, either *in-vitro* or *in-vivo.* In addition, we never observed a native FAM161A signal outside the BB or centrosome in ciliated cells, and both EM data and IF data did not reveal any signal in the microtubule network of photoreceptors. We think that the

hypothesis that FAM161A as a new microtubule associated protein (MAP), with a function in regulating the microtubule stability (such as RP1 in the axoneme [50]) should be reconsidered unless further research proves otherwise. However, the microtubule binding activity demonstrated in this research fits perfectly with our proposed model of FAM161A as a new IFT component. In fact IFT proteins move on microtubule network by interacting with motor molecules, and were localized primarily at the level of the BB or CC, with an expression pattern similar to FAM161A. In addition other centrosomal or kinesin proteins have been localized to the microtubule network when expressed ectopically. This has also been observed for LCA5, which is, as previously mentioned, implied in IFT transport or regulation.

To conclude, although our data are not conclusive enough to clearly establish the function of FAM161A, we have nonetheless taken further steps in the understanding of this protein. In early 2010, FAM161A was simply one of the many unknown genes, with unknown function and an enigmatic meaningless name. Today we know that this gene is involved in retinal degeneration, that is one of the most common causes of autosomal recessive RP in Israel and the Palestinian territories, that is involved in cilium function and that it interacts with other ciliary protein. But this is just the beginning of an interesting story where the best has yet to come.

"Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning". Winston Churchill, English statesman, 1965.

II

INVESTIGATING THE GENETIC BASIS OF PERIODIC FEVER WITH APHTHOUS STOMATITIS, PHARYNGITIS AND CERVICAL ADENITIS (PFAPA) SYNDROME

II.1 INTRODUCTION

1.1 INNATE IMMUNE SYSTEM AND INFLAMMATION

"Heat cannot be separated from fire, or beauty from the Eternal", Dante Alighieri, Italian poet, 1321

The immune system is the watcher of our body. It is responsible to identify, block and destroy any known pathological agent, such as viruses, bacteria, parasites and even own cells that are damaging for the system. The immune system can be divided in two components: the innate immune system (IIS) and the acquired (or adaptative) immune system (AIS). Whereas the latter is highly specific and efficient to recognize and destroy a specific agent, the former is less selective though not less important. These two components work together for the host defense.

The innate immune system represents the first protection of our body. The first defensive elements of the IIS are physical barriers that hinder the entering of a pathogenic agent inside the organism. They evolved to be almost inaccessible to any organism using different system of safeguard (for example tears for the eye, or keratinization for the skin). But in case that an agent manages to enter the body, it can encounter other defense mechanisms of the IIS, composed of specific cells. Some of these, such as macrophages, dendritic cells and mastocytes, are located at the level of entering tissues, and they have the capability of internalizing the pathogen and activate a response known as inflammation.

Inflammation comprises a series of macroscopic and molecular changes aimed to create a setting in which the pathogenicity of the agent is reduced, whereas the capability of the host to respond the attack is increased. Macroscopic changes are:

1. Increase of the vascularization at the level of the injury that causes the typical redness of the infected area.
2. Increase of body temperature, since some pathogens are sensible to high inner body temperature [141]. This event is known as fever.
3. Pain, that works as an alarm for the body. It is caused by the release of pain molecules that stimulates nerves.

4. Swelling, due to the accumulation of fluid in the injured tissue and it is due to the increased blood flow and to the cellular infiltration.

All these changes are driven by the release of different inflammatory molecules, among them the inflammatory cytokines, which are synthetized by the cells in response to the pathological stress.

Among inflammatory cytokines, some are main actors in starting the acute inflammation reaction, especially in case of a bacterial infection. These are the members of the interleukin-1 superfamily, IL-1β and IL-18 [142]. They mediate various effects of the inflammation and participate in lymphocytes activation. The expression of these cytokines is initiated by class of pattern recognizer receptors called toll-like receptors (TLRs), which recognize a specific pathogen-associated molecular pattern (PAMPs), like the lipopolysaccharide (LPS) of external bacterial membrane [143]. This pathway leads to the release of NF-κB and the expression of pro-IL-1β. However, the pro-IL-1β is not active as pro-protein, but it needs a cleavage step to be released in an active state. This cleavage is mediate by the Caspase-1 (CASP-1). It is interesting to notice that, whereas the activation of all the other cytokines is controlled at the transcriptional level, this additional control step was observed only for IL-1β and IL-18 [144] . Similarly, the activation of the CASP-1 itself is strictly controlled. This molecule is produced as pro-CASP-1 and an important intracellular complex, called inflammasome, drives its maturation in response to a pathogenic agent. This large protein complex (700 kDa) is sufficient *in vitro* to activate CASP-1 [145]. It is mainly composed by a group of intracellular protein known as NOD-Like Receptors (NLRs) that usually presents a specific and well conserved molecular structure (Fig. 1a): a CARD or pyrin effector domain (PYD) at the N-terminal or C-terminal of the protein, responsible for the interaction with other effector proteins; a central NACHT domain that binds ATP and it is necessary for the self-oligomerization of these proteins; at C-terminal a variable numbers of leucine-rich repeats (LRRs). Similar to the extracellular TLRs, NLRs work as intracellular pattern recognition receptors and can recognize specifically intracellular pathogenic elements such as LPS or bacterial flagellin thanks to their LRRs domains. It is the CARD (Caspase activator and recruitment domain) or the PYD (with the participation of other accessory CARD containing protein, such as ASC) that directly recruits, by the interaction with CASP-1 CARD domain, the pro-CASP-1 and all other mediators of the proteolysis. Each pathogenic agent can recognize and activate the specific inflammasome that takes its name from the most represented NLRs (fig 1b). For example, in human several different bacterial PAMPs can activate the IPAF

inflammasome (e.g: Salmonella [146] and Pseudomonas aeuruginosa [147]), whereas NALP1 inflammasome is activated by the lethal toxin of *Bacillus anthracis* [148]. Another important complex is the NALP3/cryopyrin inflammasome. It is activated by different PAMPS such as the LPS, peptidoglycans and also nucleotide sequences [149, 150]. Mutation in NALP3 and in proteins that interact directly with this complex have been associated with a very well defined class of diseases in which the inflammation system is impaired, the autoinflammatory syndromes. The main role of all other known NLRs in the context of the inflammasome is not well described, but their identity at the structural level suggests a similar or overlapping functions.

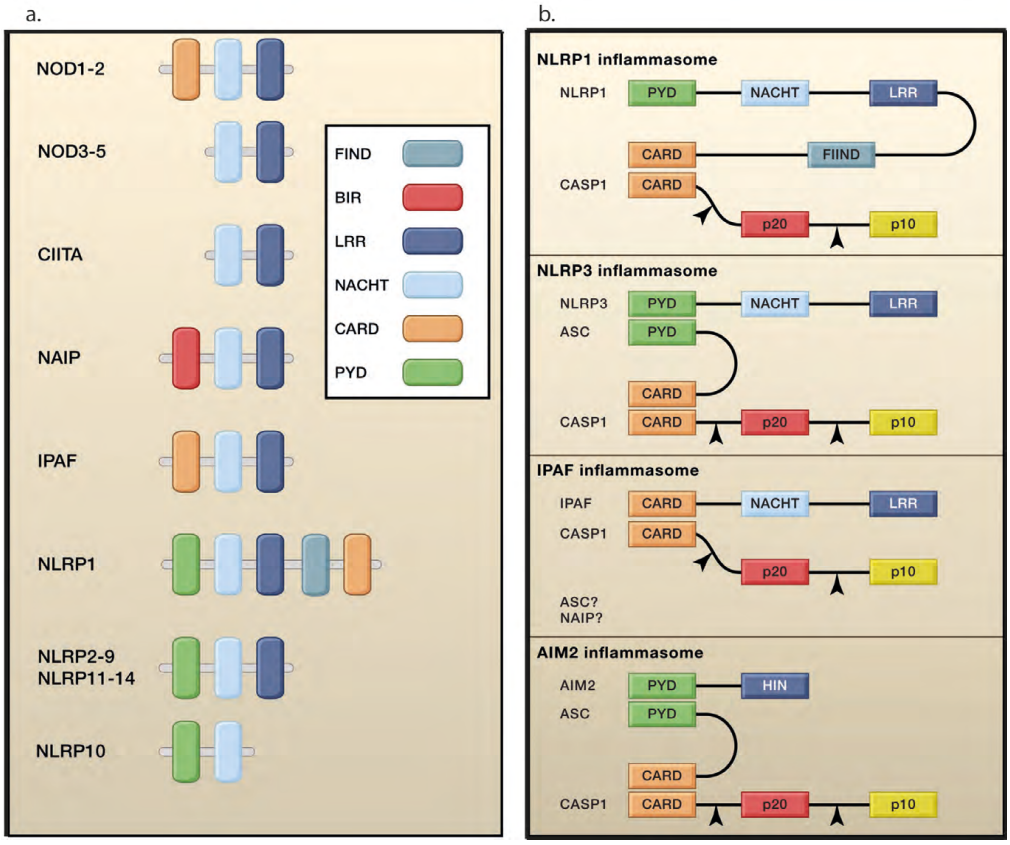


Figure 1. Schematic representation of NLR proteins structure and of relative inflammasomes.
a) The picture shows a schematization of protein domains found in NLRs. Each domain is indicated by a square shape and by a different color as reported in the legend. **b)** A schematic cartoon representing some of most studied inflammasomes, each activated by different stimuli. It is possible to observe that the NLRP3 inflammasome requires the interaction with another CARD domain containing protein ASC, whereas other inflammasomes such as NLRP1 already have a CASP domain in its structure. The interaction of ASC and NAIP with the IPAF inflammasome is not clear. (Abbreviations: FIND, function to find; BIR, baculovirus IAP repeat; LRR, leucine-rich repeats; NACHT, conserved in NAIP, CIITA, HET-E and TP-1 domain; CARD, caspase recruitment domain; PYD, pyrin domain; HIN, HIN-200 domain.). (Adapted by Schroder and Tschopp, 2010).

1.2 AUTOINFLAMMATORY SYNDROMES (AIS)

"If you do not conquer self, you will be conquered by self", Napoleon Hill, American writer, 1883

Autoinflammatory diseases are a group of disorders characterized by precise attacks of apparently unprovoked inflammation, without the identification of autoreactive lymphocytes [151]. Differently from autoimmune diseases, such as lupus erythematous and rheumatoid arthritis, where adaptive immunity is the main target, it is the innate immune system that is affected in autoinflammatory diseases. It is worth bearing in mind, however, that new evidence reinforces the idea of a clinical continuity between these two systems [152]. From a clinical point of view the symptoms of an autoinflammatory disease include classical symptoms of systemic inflammation such as fever, cutaneous rash or joint swelling, as well as other co-occurring symptoms, such as headache, diffuse pain, ulceration and aphthous stomatitis. Regarding genetics, which is the basis of most of these syndromes, the majorities are inherited as Mendelian traits, although a strong environmental influence has also been suggested for some of them (e.g., gout in which susceptibility seems connected to the NLRP3 inflammasome [153]). Furthermore, the number of individuals with a clear clinical classification, but for whom the genetic causes have not been found, is recently increasing in number [152]. From a molecular point of view, AIS can be categorized based on which pathway of innate immunity is affected [154]: 1) IL-1β activation; 2) NF-κB activation; 3) problems with molecular folding of some protein involved in the innate immune systems; 4) complement disorders; 5) cytokine signaling; 6) macrophage activation. The most common forms of autoinflammatory disease belong to the first three groups and they will be here discussed in details (Fig 2).

As previously described, the key molecule in the activation of the innate immunity response is IL-1β. This molecule alone, in small amounts, can promote fever *in vivo* [142] along with other systemic effects. The production and release of IL-1β is strictly regulated, and so it is clear how the effect of a molecular disruption in a component of this pathway could cause an autoinflammatory disease (Fig. 2a). For example a mutation in NLRP3, the key molecule of the cryopyrin inflammasome causes a pleiotropic class of autosomal dominant autoinflammatory syndromes known as Cryopyrin-associated periodic syndromes (CAPS or cryopyrinopathies). The milder form of CAPS is called familial cold autoinflammatory urticaria (FCAS) and it leads to a few symptoms such as

cold-induced fever and urticaria. A more severe form of CAPS is Muckel-Wells syndrome, characterized by fevers, hives, sensoneural hearing loss and cold-induced arthritis. Finally, the most severe form is the neonatal-onset multisystem inflammatory disease (NOMID) with fever, urticarial, overgrowth of long bones and chronic aseptic meningitis [155]. Mutations in *NLRP3* have a gain of function effect leading to a continuous inflammasome activation with overproduction of active IL-1β, demonstrated by overexpression of mutant NLRP3 in monocytes cells [156].

Another gene involved in the IL-1β activation pathway is *MEVF,* which encodes for the protein pyrin or marenostrin (Fig. 2a). Mutations in this gene cause another common autoinflammatory syndrome known as Familiar Mediterranean Fever (FMF). This disease is characterized by recurrent fever attacks with peritonitis, pleural inflammation and arthritis. The name is derived from the high rate of mutated alleles found in populations from the Mediterranean basin and Middle East [157], probably caused by a selective advantage of the heterozygous population for these alleles to a specific endemic infectious diseases. FMF is transmitted as a recessive trait, although this seems somewhat contradictory considering the functional data published on pyrin. In fact, pyrin seems to have both inflammatory and anti-inflammatory effects, shown by several functional experiments using siRNA to inhibit *MEFV* [158, 159]. Also other *in vivo* experiments on lymphocytes derived from healthy donors or affected controls confirmed this dual function of pyrin [158, 159]. Recently, a new mouse model for FMF seems to support the anti-inflammatory hypothesis, showing an increase in IL-1β release by KO mutant macrophages in response to inflammatory stimuli [160]. One possible explanation for this dual role of pyrin is linked to its function. Pyrin presents a PYR domain at N-terminal, like NLRP3. This domain is responsible for the interaction with ASC, one of the key molecules of the inflammasome. It has been postulated that pyrin could work either to inhibit IL-1β activation as antagonist of Caspase-1 binding with ASC [161], or by being directly part of the NLRP3-inflammasome [162]. It is therefore possible that these two different pathways could be activated at different moments and by different stimuli. To confirm this hypothesis, an ASC dependent, but NLRP3- independent, activation of the inflammasome has been reported [163]. This dual role of pyrin function could explain why in certain cases *MEFV* mutations are associated with autosomal dominant inheritance with a possible gain of function effect. Interestingly healthy carriers with a recessive mutation may as well have some evidence of perturbed inflammation [164], or just show a limited phenotype for a short time

[165], demonstrating a non-linear effect of these mutations on the general inflammation pathway.

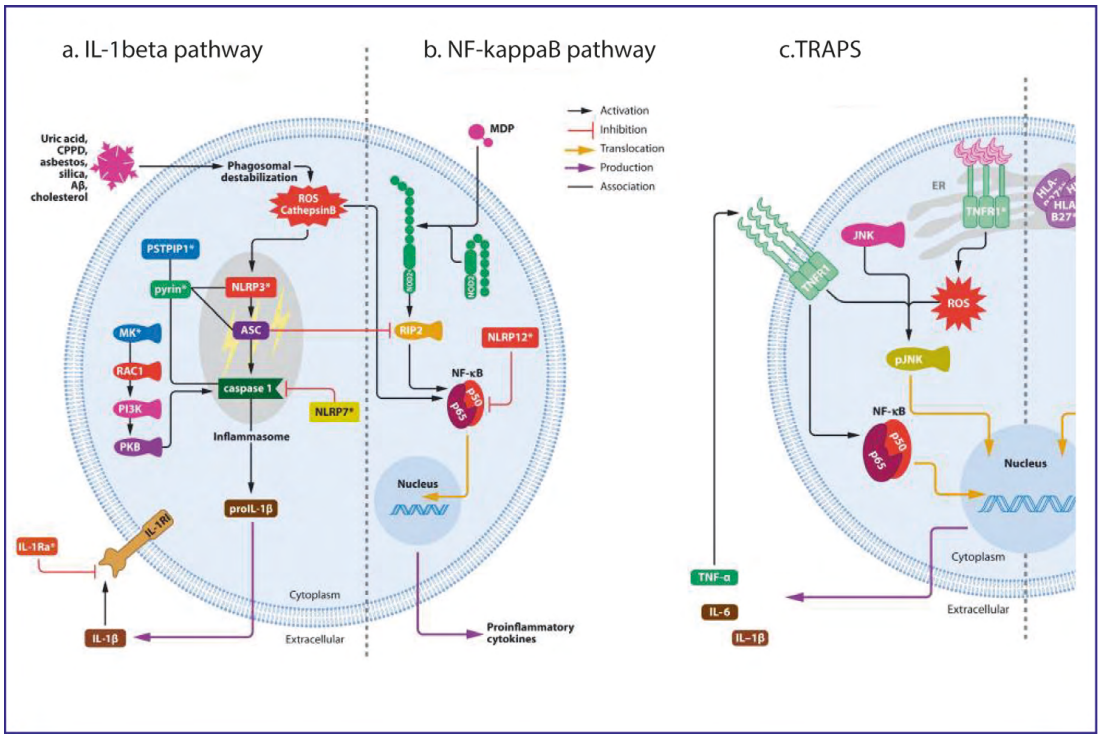


Figure 2. Pathways involved in autoinflammatory diseases.
a) Schematic representation of a cell and some of the molecules involved in active IL1-Beta release. The diagram shows the involvement and interaction between these molecules. NLRP3 interacts with ASC, forming the NLRP3 inflammasome to activate IL-1Beta. The involvement of Pyrin and Mevalonate Kinase (MK) is shown. Other molecules involved in other autoinflammatory diseases, such as PSTPIP1 (causing Pyogenic Arthritis with Pyoderma gangrenosum and Acne –PAPA syndrome) and IL1-RA (causing Deficiency of the Interleukin-1 receptor syndrome - DIRA) are drawn. Different stimuli activating this pathway are indicated at the top of the figure. **b)** NF-κB activation in recurrent fever. Here the components of NF-kb activation are reported. As described already in the text NLRP12 has a negative effect on NF-κB activation that drives to the production of inflammatory cytokines. NOD2 interacts with the cell wall component of bacteria muramyl dipeptide (MDP) and indirectly activates NF-κB. **c)** TRAPs associated disease mechanism. In case of TRAPS, the disease mechanism is connected to an increase of unfolded proteins in the cytoplasm that leads to cellular stress and cell death. Moreover the presence of this stress increases the number of stress molecules such as JNK and MAPK making the cell much more sensitive to inflammatory stimuli. All proteins indicated with an asterisk are found to be associated with an autoinflammatory syndrome. (Adapted from Masters et al., 2009).

Among the autoinflammatory recurrent fevers, hyperimmunoglomulinemia D with periodic fever syndrome (HIDS) deserves a mention. This disease is due to autosomal dominant mutations in the gene *MVK,* for which the product is the ubiquitously expressed enzyme mevalonate kinase (Fig. 2a) [166]. The functional link between the

inflammasome and this protein was not overly clear, considering that homozygous recessive mutations in *MVK* cause a severe form of mevalonic aciduria, an early deadly disease [167]. In fact it seems that mutations in only one allele of this gene is enough to protect the body from mevalonic aciduria symptoms, but not enough to save it from HIDS. It has been clearly demonstrated that the missing link between the inflammasome and MVK is found in the mevolanate kinase pathway, which is important for the production of protein kinase B, one of the key factors in caspase-1 activation [168].

Moving onto the class of the NF-κB disorders (Fig. 2b) is a syndrome called familial cold associated periodic syndrome 2 (FCAS2), better known as Guadalupe variant periodic fever syndrome. This disease is associated with non-sense and splice site mutations in the *NLRP12* gene and presents some symptoms that are common to CAPS, like recurrent episodes of fevers triggered by cold exposure, arthralgia and myalgia [169]. This protein seems to be directly involved in the inhibition of NF-κB activation [170]. Therefore, dominant non-sense and frameshift mutations cause haploinsufficiency with consequent constant activation of the NF-κB inflammatory pathway. Recently, a recurrent heterozygous missense mutation in *NLRP12* has been connected to a gain-of-function phenotype and to a functional defect in the inflammasome [171], revealing a new important role of this protein in autoinflammatory disease.

The last AIS discussed here belongs to the class of protein folding disorders and is known as TNF receptor-associated periodic syndrome (TRAPS) (Fig. 2c). This disease is caused by a mutation in the TNF receptor-1 gene (*TNFRSF1A*) [172]. It is characterized by periodic fever with severe abdominal pains, arthritis, migratory skin rash and periorbital edema, with amyloidosis. Also TRAPS is transmitted as an autosomal dominant disease without geographic preference and with a clear mutational hotspot located mainly to the extracellular domain of the receptor [154]. It has been shown that many of the mutations causing TRAPS have an effect on protein folding and cause the aggregation of the receptor in the ER as well as the inhibition of trafficking to the membrane [173]. Furthermore, a reduction in the binding capacity of TNF has been demonstrated [174]. Another mechanism involved in the disease is linked to the decrease of receptor shedding and consequential accumulation in the ER. For example, it has been observed that mutations affecting the proteolitic sites on the protein sequence causes a reduced receptor shedding with not correct folding of the same [175]. Moreover, it has been also observed that the intracellular accumulation of TNFRSF1A induces the intrinsic activation of several death pathways and MAPK kinases making

cells that express a mutant TNF receptor more sensitive to the inflammatory response [175]. Recently a new role of defective autophagy due to a mutation in *TNFRSF1A* has been proposed for TRAPS induced inflammation as an important mechanism in TRAPS pathogenesis. It is likely that the sum of these different defects cause the different spectrum of the TRAPS phenotypes.

We have just briefly described some AIS for which a specific genetic cause has been identified. For other AIS however, the genetic cause has not been found or even postulated, and the etiology is still unclear. Among these there is the periodic fever with aphtous stomatitis and cervical adenitis (PFAPA) syndrome.

1.3 PFAPA SYNDROME

"She gives me hot and cold fever, then she leaves me in a cool, cool sweet...", from song Crazy Little Thing Called Love by Freddie Mercury, British rocker , 1987.

Firstly described by Marshall in 1987 [176], this autoinflammatory syndrome is characterized by two or three days of high fever (higher than 39°C) that reoccurs regularly every month, alongside at least one of these other symptoms: aphthous stomatitis, pharyngitis and cervical adenitis [177]. Other clinical manifestations can be associated with PFAPA, such as headache, nausea, abdominal pain, and cutaneous rashes, even if these other symptoms are quite rare. It is an early onset disease, with a median age of 5 years old, and the syndrome usually shows a complete resolution of attacks after adolescence. However, cases of PFAPA not resolved after adolescence have been reported in the literature [178-181]. In most cases, patients are otherwise asymptomatic during the episodes, without any problem in development or growth. A molecular profile based on inflammation molecules expression does not show any difference between healthy controls and affected individuals during asymptomatic stages between two PFAPA flare-ups [182]. Diagnosis of PFAPA is based mainly upon symptoms, but since they mostly overlap with other autoinflammatory fever diseases, a genetic screening is often recommended in order to exclude other recurrent fevers like FMF, HIDS, TRAPS and CAPS.

Recently new molecular and clinical aspects of PFAPA have been revealed, showing a completely different molecular profile compared with other autoinflammatory syndromes. In fact, during flares, PFAPA patients present classical symptoms of inflammation with non-specific leukocytosis, neutrophilia and elevated levels of C-reactive protein and fibrinogen [183]. In addition to this, the expression profile of inflammatory genes during a flare seems to be very specific for PFAPA, with a general increase in inflammatory genes, both interferon γ and interleukin-1 related. Moreover, a general increase in the activation of the adaptive immune system has been observed [182]. The involvement of IL-1β pathway in PFAPA also seems to be confirmed by the relevant increase of IL-1β, IL-1 receptor antagonist and other cytokines during the febrile attack in the affected patients compared to healthy controls, and compared to levels in the same patient during the different stages of the disease [184].

The most suitable treatment for PFAPA is still mostly based on treating the symptoms. The use of corticosteroids, such as prednisone, is the golden standard for treatment, even though a possible side effect is a decrease in the time window between two flares [177]. Another drug used in the treatment of FMF, the microtubule destabilizer drug cimetidine, has been tested also on PFAPA patients, although it was only effective in a portion of them (almost 30%)[185]. Additionally, despite being controversial due to the invasiveness of the technique, the surgical removal of the tonsils has been proposed, and it has led to the complete remission of symptoms in a high percentage of cases [186-188]. Recently, a new innovative treatment has been investigated based on the finding of inflammasome involvement in the disease. The validity of IL-1 inhibitors, such as Anakinra (IL-1RA), has been shown to lead to a remission of the attacks in a small cohort of patients and, due to the low side effects of this treatment, could became the standard for treating this disease [182].

The genetic origin of PFAPA is one of the most debated issues of this syndrome. Several different theories on its etiology have been proposed, although the official definition is to consider it as sporadic disease [189, 190]. Some evidence supports this hypothesis: 1) it is present in people with different ethnic backgroups; 2) it seems to resolve spontaneously and with tonsil removal; 3) it has a very early onset, when the immune system is not completely efficient, suggesting environmental factors. On the other hand several clues lean towards a hereditary hypothesis. Among these is the fact that no trace of any pathological agent has been found in the throat of affected children during the attacks as well as the responsiveness to steroids but not to antimicrobial agents and the periodicity of the phases between flares, which can often be accurately

predicted [190]. In addition to this, the discovery of siblings with PFAPA seems to reinforce the idea of an inherited disease [191, 192]. Our collaborators Dr. Hofer and Dr. Cochard, from the pediatric department of the University of Lausanne Hospital, showed a clear positivity in family history (e.g., via a questionnaire) for episodes of recurrent fevers in almost 45% of families in the total analyzed PFAPA cohort (composed of 85 patients). When the same survey was administrated to comparable families of non-PFAPA controls only 12% these gave positive results [193]. Moreover, a recent paper showed a high percentage of family history for recurrent fevers in PFAPA patients, in which the disease persisted through to adulthood, probably indicating the presence of a milder and a more severe form of this disease [194]. Although these studies seem to indicate a genetic predisposition, it cannot be excluded that this is due to specific variants present in known inflammatory genes. For this reason several works investigate the involvement of FMF, TRAPS, MVK and NLRP3 mutations in PFAPA cohorts. For example in Dagan et al. [195], common mutations of FMF, TRAPS, CAPS and of the gene involved in Blau Syndrome, *CARD15*, were screened in a cohort of 57 PFAPA patients. In this cohort, 27% of the patients carried a variant in *MEFV*, though this can be expected as this population had a high number of patients from the Mediterranean basin, where the carrier frequencies of MEFV variant is about 1 in every 4. Furthermore, 5.3% carried a mutation in *CARD15*, around 2% in *TNFRSF1A* and in NLRP3. The others were genetically negative for each screened mutation. Of course clear limitations are evident in this analysis and in similar studies. Firstly, not the whole gene is screened, but only predominant mutations, so more specific, new variants, or even functional polymorphisms, could cause the PFAPA phenotype. Secondly, the inclusion criteria based on clinical data are not always enough to distinguish PFAPA patients from other milder forms of autoinflammatory fevers. Another paper proposed a possible link between MVK mutations and PFAPA, based simply on the high level of IgD in some PFAPA patients, but it failed to find pathogenic changes in the MVK gene [196]. A more comprehensive study revealed that among 199 patients that completely fulfilled the PFAPA clinical criteria, the majority of them were found negative for any variants in inflammatory genes. Of the remaining part, 32 patients were found to have mutations already known to cause another inflammatory disease (they were probably misdiagnosed PFAPA subjects), whereas more than 37 carried either heterozygous or rare polymorphism with not clear pathological effect in one of the known recurrent fever associated gene [197]. Another more recent study, performed by the group of our collaborator Dr. Hofer here in Lausanne, based on 57 patients affected by PFAPA, found variants in other inflammatory syndrome genes in 25% of their cohort, with 12 patients

with a NLRP3 variant, 4 with MEFV variants and 1 with MVK and TNFRSF1A variants [184]. Interestingly all variants were present in a heterozygous state, including MEFV, and were mainly low penetrance variants, or rare polymorphisms with unknown phenotype. The presence of rare polymorphisms in these genes with frequencies of around 2% in healthy controls raises the question whether these variants should really be considered as a cause of autoinflammatory diseases and how they are linked to the PFAPA phenotype.

II.2 METHODS

2.1 PATIENT SELECTION, DNA COLLECTION AND PEDIGREE ANALYSIS

The families participating in this study were from a large collection of PFAPA patients whom were examined and selected by Dr. Michaël Hofer, at the Lausanne University Hospital. The familiarity and inheritance pattern was investigated by administrating a specific survey to family members of the affected patients. PFAPA phenotype was defined according to published clinical criteria [198], and for the cases where the phenotype was unclear or a specific ethnicity was ascertained, a genetic screening for the most common recurrent fever genes was performed. Informed consent was provided according to the human rights ethical directives. Blood from all available members of 15 small and unrelated Swiss families was collected and high quality ultra pure DNA was extracted from the blood using a NucleonBaCC3 DNA extraction kit (GE Healthcare Life Science). Some of the blood was kept to isolate the peripheral blood mononucleate cells (PBLs) for EBV-driven lymphoblastoid transformation. Pedigrees of collected families were drawn according clinical records, and the inheritance pattern was directly derived from pedigrees analysis. As controls for the screening of rare variants, data obtained by exome sequencing of 416 healthy anonymous individuals from a large population based study done in Lausanne (the CoLaus study) were used [199].

2.2 PBLS TRANSFORMATION BY EPSTEIN-BARR VIRUS (EBV) INFECTION

PBLs from affected and non-affected members of each family were collected from whole blood using Ficoll-Paque Plus (GE Healthcare, Lifescience), by density gradient centrifugation separation. Once isolated, cells were treated for EBV immortalization using classical protocols [200], optimized for our requirements. After three washes with PBS supplemented with 2% FCS, $1x10^6$ PBLs cells were put in a 10 ml Falcon tube, diluted in 1 ml of LCC medium (RPMI 1640 medium, 20% serum, 2 µg/ml Cyclosporine A) and put in direct contact with EBV virus suspension for 1h. Cells were then plated into a well of a 12 well plate and medium changed when necessary. After about one month in culture, cells were completely immortalized and then frozen in RPMI 1640 with 20% serum and 10% DMSO to cryopreserve them.

2.3 GENOTYPING AND LINKAGE ANALYSIS

Six out of 15 families were selected, based on the availability of clinical information and the size of the family trees. All members of each of these families were whole genome genotyped by Illumina Human Linkage12 panel (Illumina, CA), containing more than 6000 highly heterozygous SNPs, with an average distance between each SNPs of about 441Kb. Output was then analyzed using the software Illumina Genotype Studio (Illumina Inc., CA) and the SNPs calls were filtered out based on quality and the presence of Mendelian inconsistencies. Genotype data were handled using the commercial software Progeny 7(Progeny Software LLC) to generate appropriates files for linkage analysis software.

Linkage analysis was carried out with Merlin software using the Vital-IT high performance computer structure (Swiss Institute of Bioinformatic, Lausanne, www.vital-it.ch). Both parametric and not-parametric analysis models were tested. Because of the lack of data on genetic background for this disease, we could not perform a classical computational approach to estimate the penetrance factor. Therefore, penetrance factor for parametric analysis was empirically calculated based on the observed families as a ratio between the number of affected patients and the total number of subjects predicted to be affected based on the model.

2.4 NGS EXPERIMENTS

We performed two NGS experiments. In the first we captured all exons and boundary introns of all the annotated genes (based on NCBI build 3.1 annotations) present in the identified linkage interval for only three affected individuals. The homemade chip was designed in collaboration with Genotypic Ltd (Bangalore, India) and was performed using SureSelect on-array chip capturing (Agilent, US). Captured DNA was sequenced by Fasteris (Geneva) facilities in Switzerland, with the Genome Analyzer II from Illumina. Reads were aligned to the whole genome using CLC bio Genetic Workbench software (CLC bio, Denmark).

In the second NGS experiment we performed the whole-exome capturing of two affected members from four unrelated families of our cohort, plus four individual patients. This was done at the Lausanne Genomic Facility (GTF) in collaboration with Dr. Andrea Superti Furga. Exons were captured using SureSelect Exome kit V5 (Agilent, US), and sequenced by Illumina GAII with paired ends. Reads were aligned to the reference

genome hg19 release taken by UCSC website. We used for our analysis alignment data obtained from the sequencing service, and from our homemade pipeline for alignment specifically developed for this purpose for comparison reasons.

2.5 BIOINFORMATICS ANALYSIS

A specific home made pipeline was developed for the alignment of sequenced reads to the whole genome. This was necessary to reduce the abnormal false positive rates present in the variant file provided by the sequencing facility. This high false positive rate call was probably due to a suboptimal quality of some of the DNA samples. The homemade pipeline implemented in the analysis used the best combination available of aligner software (Novoalign- from Novosoft) and genotype calling tools (UnifiedGenotyper tool from GATK) as already demonstrated in other publications [201]. A schematic diagram showing all steps and software used for the pipeline is presented here (Fig. 3). Specific parameters used for the analysis are indicated. VCF files were annotated and variants analyzed using Ingenuity Variant Analyzer software (Ingenuity System, CA).

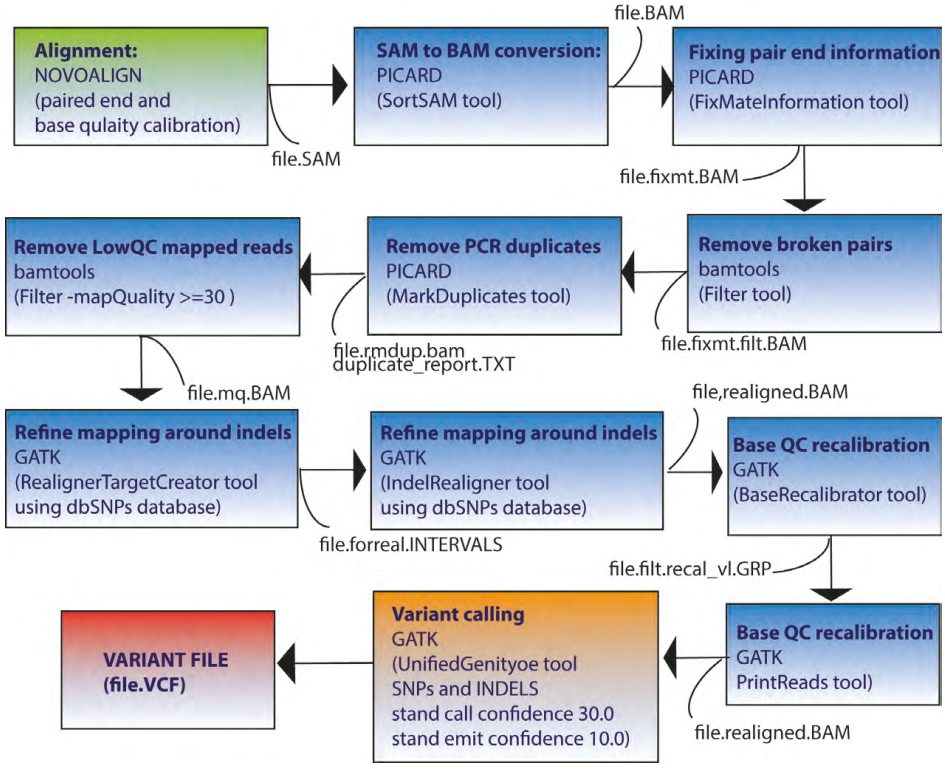


Figure 3. Scheme of bioinformatics analysis performed with the homemade pipeline to obtain a variant calling file. The alignment part is indicated in green and requires almost 12 hours. Preparing steps for genotype calling are in blue and require almost 6 hrs. Variant calling is in orange and it requires about 4-5 hours to be completed.

II.3 RESULTS

3.1 PFAPA SYNDROME PRESENTS AN AUTOSOMAL DOMINANT INHERITANCE WITH INCOMPLETE PENETRANCE

Pedigree analysis (Fig. 4) showed a clear pattern of inheritance. It must be noted however, that affected pathogenic status was clinically determined only for the most recent generation, whereas for previous generations it was inferred from questionnaires completed by family members. Based on the pedigree analysis, we propose an autosomal dominant with incomplete penetrance inheritance for the majority of families. Because of the lack of genetic data for this disease to infer the penetrance factor from the literature, we empirically determined it (around 30%) by simple pedigree analysis.

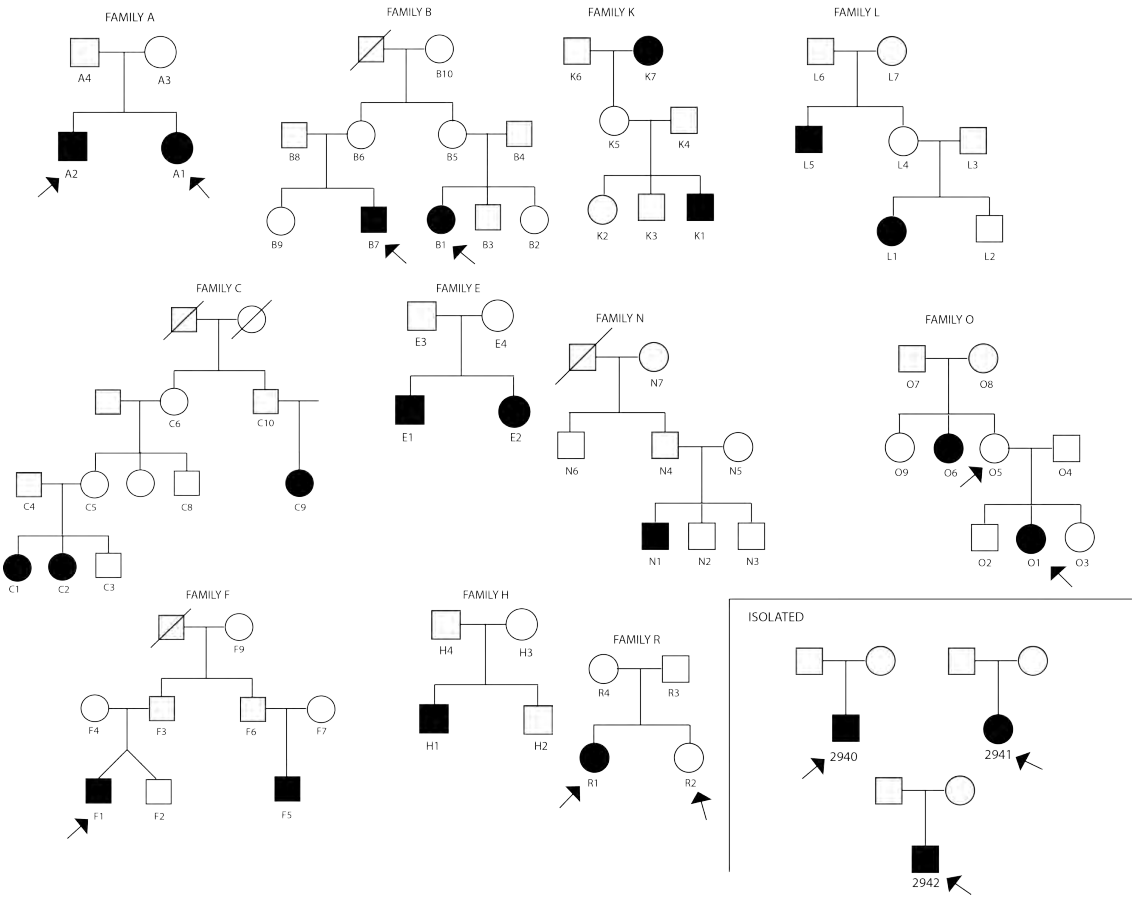


Figure 4. Pedigrees of selected families. An arrow indicates subjects that were full exome sequenced.

3.2 LINKAGE ANALYSIS REVEALS A SHARED REGION ON CHROMOSOME 8

In collaboration with Dr. Hofer, we selected all members of some of the most informative families (families A, B, C, H, K, L, O), to perform a genome wide genotyping using Linkage12 Chip from Illumina. We obtained an extremely high quality call rate (>98%), and low Mendelian inconsistency (less than 0.2%). Multifamily whole genome analysis revealed a single peak with a LOD score of 3.4 on the short arm of chromosome 8 considering a fully penetrant model (in which the obliged carriers are considered as affected) (Fig. 5b). The LOD score is reduced to 2.4 when applying the reduced penetrance model, although the unique peak on chromosome 8 is still confirmed (Fig 5c). The same result was obtained when using a non-parametric model, where any postulation on the genetic inheritance pattern is contemplated. For this model no penetrance was considered, and we only included individuals that presented clinical symptoms in the analysis. In this case we obtained a maximum LOD score of 2.9, but it was again the only positive peak obtained with this analysis (Fig 5a). We performed a simulation study to test the strength of the observed linkage, and this excluded that the positive peak was obtained by chance. The identified region was localized between 8q24.1 and 8q24.3 band of the long arm of chromosome 8, and was 11 Mb large (Fig 5d). At the time of the linkage we used Hg17 release of the human genome, and that region contained only 40 RefSeq genes (table 1). Among these genes (Fig 5e), we decided to screen one candidate of interest, the gene *Src-like adaptor protein 1 (SLA),* located in the intronic region of the thyroglobulin gene. The screening of all exons and intron boundaries, in all affected members, as well as the screening of the glucocorticoid binding domain did not revealed any pathological variant. To complete the screening of all other genes in the interval we decided to perform a custom gene capture of all exons and intron boundaries for each of these 40 genes. We performed the screening over three affected individuals from three different families (B1, L1, C2). Unfortunately, the capture procedure gave some unexpected technical issues, with unwanted noise, that spoiled a correct alignment. This could be related to the high number of pseudogenes and repetitive sequences present in this region that impaired the capture process.

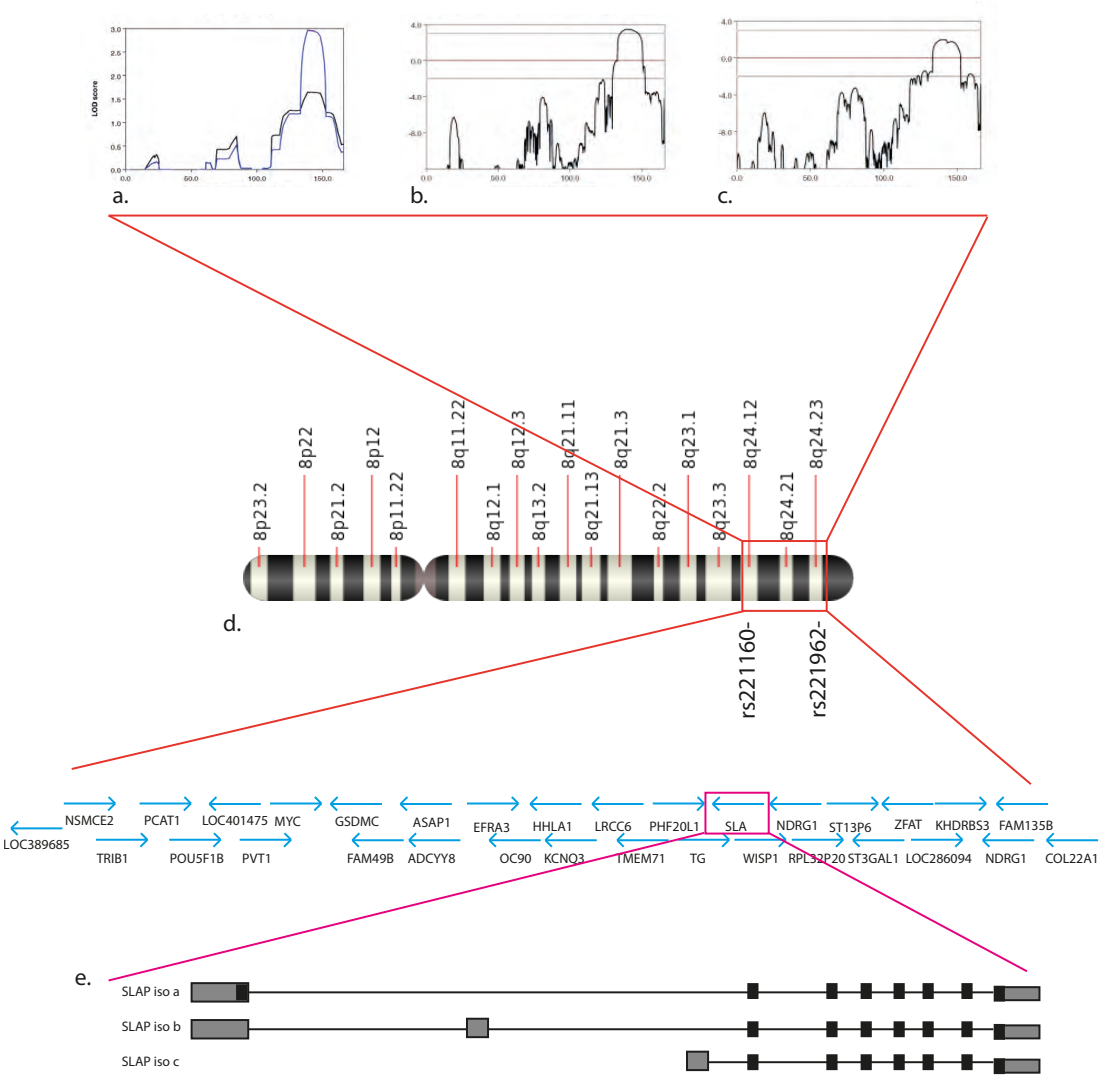


Figure 5. Linkage analysis.
a) Output of Merlin linkage analysis for the non-parametric model. **b)** Output for parametric model with complete penetrance. **c)** Reduced penetrance model. **d)** Linkage analysis revealed a peak on the long arm of chromosome 8 between the two indicated markers. Genes present in the interval are reported. **e)** Among these genes, we performed a complete sequencing of all exons of the gene SLA.

Table 1. Genes present in the linkage interval.

Gene	Function
FAM84B	Unknown
PCAT1	lncRNA
POU5F1B	Probably transcription factor
LOC727677	lncRNA
MYC	Transcription factot
MIR1204	miRNA
PVT1	lncRNA
MIR1205	miRNA
MIR1206	miRNA
MIR1207	miRNA
MIR1208	miRNA
LOC728724	lncRNA
GSDMC	Leucin zip. Not known
FAM49B	Unknown
MIR5194	miRNA
ASAP1	Membrane trafficking
LOC1005071	lncRNA
ASAP1-IT1	lncRNA
ADCY8	Adenilate cyclase brain specific
EFR3A	Membrane protein involved in hearing

Gene	Function
OC90	Subunit of otoconin
HHLA1	Human endogenous retro\
KCNQ3	Potassium channel subuni
HPYR1	lncRNA
LRRC6	Ciliary protein
TMEM71	Unknown
PHF20L1	Unknown
TG	thyroglobulin
SLA	Regulator of T cells
WISP1	lncRNA
NDRG1	Mitotic spindle protein
ST3GAL1	Glxcosilating protein
ZFAT	Transcription factot
ZFAT-AS1	lncRNA
MIR30B	miRNA
MIR30D	miRNA
LOC286094	lncRNA
KHDRBS3	RNA binding protein
FAM135B	Unknown
COL22A1	Collagen protein

3.3 EXOME SEQUENCING SHOWS NO CLEAR COMMON VARIANT ASSOCIATED WITH PFAPA PHENOTYPE.

Since custom capture sequencing failed, we decided to directly perform whole exome sequencing for some affected individuals from our families, as well as for a few independent samples (samples A1, A2, B1, B7, O1, O5, R1, R2, F1, 2940, 2941, 2942 – 12 samples sequenced in total) in collaboration with Dr. Andrea Superti-Furga. The sequencing procedure worked well, although the percentage of correctly aligned reads was not equal for all samples (in particular for sample A2, B7, O1 and O5 the percentage of aligned sequence was around the 75%, compared to the others that had an average of 95% of correctly aligned sequences). To improve this output we performed a new alignment protocol using a homemade pipeline. In this way we reduced the error rate linked to computational mistakes for low quality samples. Variants called with both alignments were considered to be highly consistent.

To begin with, we searched for rare variants on the isolated region of chromosome 8. We could not identify any shared gene carrying one or more possible pathogenic variants in this region, nor on the whole of chromosome 8. Then we investigated the whole exome for genes presenting rare variants that were shared between all affected patients. In Table 2 are reported genes where rare variants (MAF <2%) that were found to be present in more than 90% of the analyzed samples (this value was chosen because we could not asses the clinical status of sample R2). Many of the identified genes belonged to the *MUC* family. These were excluded from subsequent analysis because of their high sequence identity that usually impairs the correct alignment of the sequences. Similarly, all the other common genes were excluded, either because the changes were not found in both alignments, or because we observed many novel and rare variants also in totally unrelated controls (sequenced with the same machine, and aligned with the same pipeline), suggesting problems at the level of repetitive sequences, or a badly annotated reference. Indeed, by this analysis, we could not identify any common mutated gene(s) involved in PFAPA.

Since the genetic background for this disease is not very well defined, we decided the to investigate the possibility of genetic heterogeneity, by considering only rare variants present in some, but not all, of the samples but present in all the affected members of the same family. These genes are listed in Table 3. It is possible to observe that some of these variants were shared by only two or three different families, and

therefore we initially excluded them in this first analysis. Among the remaining genes, an interesting candidate was *SARM1,* a gene involved in the innate immunity pathway. We identified a novel variant p.G49A both in heterozygous and homozygous states in 6 different samples out of 12. Although both alignment pipelines positively called this variant, it was not confirmed by Sanger sequencing. Another interesting gene was *C16orf11*. We identified two novel missense changes p.R3Q and p.R29Q in patients B1 and O1 respectively, whereas a very rare variant p.R106W (rs4551533) was identified in samples A1, A2 and 2940. Two other less rare variants were identified in sample F1 and R1 (p.G379D – rs200682205 and p.G502S- rs74459225). The function of this gene is unknown, but the fact that it was only found in one of the two affected family members of family B did not support a role for this gene in the disease. Sanger sequencing confirmation is in progress. Moreover we identified several novel and non-annotated variants, which were unique for each family. The role of these variants in the frame of PFAPA is not clear and it is very difficult to ascertain. All identified novel variants divided per family are reported in Table 6.

Table 2. Genes in common between all samples

GENE NAME	Number of variants	Percentage of cases	Exclusion Criteria	Found in both alignment
WNK1	54	100%	Alignment mistake	No
MUC6	40	100%	Mucin gene	yes
MUC2	19	100%	Mucin gene	Yes
NBPF15	11	100%	Many novel variants presents also in controls	Yes
ZNF384	1	100%	Bad reference	Yes
OR4C45	1	100%	Many novel variants presents also in controls	Yes
MUC17	91	90%	Mucin gene	Yes
MUC16	26	90%	Mucin gene	Yes
TTN	18	90%	Low coverage regions	Yes
ZNF492/ZN	16	90%	Many novel variants presents also	Yes

F98			in controls	
GOGAL6L1	14	90%	Not well annotated gene	Yes
PCMTD1	8	90%	Many novel variants presents also in controls	Yes
C11orf40	3	90%	Suspected False positive	No
KIAA1751	2	90%	Suspected False positive	No

3.4 SCREENING OF AUTOINFLAMMATORY FEVER GENES DOES NOT SUPPORT THEIR INVOLVEMENT PFAPA ETIOLOGY

Since the presence of rare variants in other inflammatory genes could have an affect on PFAFA phenotype, we decided to search specifically for variants in all known autoinflammatory genes in our cohort. We identified some rare variants in *MEFV* for samples O1 (p.I591T, rs11466045, minor allele frequency – MAF - in Lausanne controls 0.07), R1 (p.F425Y, rs104895169, not present in Lausanne cohort), A1 and A2 (p.E148Q, rs3732930, MAF 0.06). Another more common variant (p.R202Q, rs224222, MAF 0.18) was identified in a heterozygous state in samples 2940, 2942, F1, R2 and in a homozygous state for sample B7. Interestingly, patients A1 and A2 also presented the heterozygous variant in *NLRP3* p.Q703K (rs35829419, MAF 0.04), whereas a rare variant often associated with a milder form of CAPS p.V198M (rs121908147, MAF 0.05) was found in a heterozygous state only in the obligated carrier O5 and not in the affected member of the family O1. The rare functional polymorphism p.R121Q (rs4149584, MAF 0.018) of *TNFRSF1A,* usually associated with milder forms of TRAPS, was only identified in family B and it co-segregates with all affected members of the family. Interestingly, we also identified a novel change in a heterozygous state in *NLRP12* associated with FCAS2 in patients R1 (p.R211C, novel), but this was also present in an clinically unaffected sister. This change was found in a well conserved region, in the NACHT domain of this protein. No variants were found in *MVK* with the exception of the frequent polymorphism p.S52N (rs7957619, MAF 0.14) present in A1, R1 and 2940 (Table 4).

Table 3. Genes in common in a few families. Genes under investigation are marked in orange.

Gene name	Families	Nrs. of variants	Percentage	Exclusion criteria	Found in both alignment	Sanger Sequenced
MLL3	A-B-2940	3	100%	Two different variants in B1 and B7	No	No
CASP8AP2	A-B	3	100%		No	No
AGRN	A-O-2940	3			Yes	No
CRB1	A-O	2	100%	Predicted tolerated	No	No
DCC	A-O	2	100%		Yes	No
FLJ43860	A-O-F-R	2	100%		Yes	No
KALRN	A-O-2941	2	100%	One is 3'UTR variant	Yes	No
PCLO	A-O	2	100%	Bad alignment	Yes	No
C6orf170	A-O-F-2941	1	100%	0.05 MAF	Yes	No
PHC3	A-O	1	100%		No	No
PLEC	A-O	2	100%	Many missense variants in controls	Yes	No
LAMB1	B-O-R	2	100%		Yes	No
ZNF160	B-O	2	100%	Frequent variants	Yes	No
FAM83H	B-O	2	100%	Suspected false positive	No	No
KIF20B	B-O	2	100%	Suspected false positive	No	No
TBP	B-O	1	100%	In-frame indel	No	No
SCN7A	A-2940	2	100%		Yes	No
SFI1	A-2940	2	100%		Yes	In progress

C16orf11	A-2940-B-F-O	4	100%		Yes	No
COL4A4	A-2940-2941	1	100%		Yes	No
RIMS2	B1,A1, 294, O1-O5	3	100%		yes	No
SARM1	A-B-F-O-2940,2941,2942	1	100%	Low coverage, False positive	Yes	yes (not confirmed

3.5 Several rare variants were present in inflammasome-related genes

Since patients affected by PFAPA syndrome showed an impairment in IL-1β and NF-κB signaling and production, we decided to look closely at all known genes belonging to the inflammasome complex, with particular attention to genes involved in the production of mature IL-1β. All identified variants in these genes are listed (Table 5). We identified two non-annotated changes in *NLRP5,* for patients 2940 and F1 (p.R274Q, and p.N255S respectively). Again, patient 2940 presented another rare variant in the gene *NLRP2,* p.L607P (rs182098487, MAF 0.001). A novel change was found in this same gene, p.G504D (novel), in patient B7, although it was not found in the affected cousin. Moreover, *NLRP2* was also the NLR gene that showed the highest number of variants among PFAPA subjects, even if these were quite common in the control Lausanne cohort (p.S4L, rs142463014, for 2492, MAF 0.09; p.I331V, rs61735077, for R1 and R2, MAF 0.05). Another common variant in *NLRP4* is p.T162M (rs117212164, MAF 0.03), co-segregating in family O. It caused a change in a very conserved residue in the NACHT domain of *NLRP4.* Among other NLRP family member genes we identified a rare change in *NLRP10*, p.I384T (rs150112481, MAF 0.01) although only in patient O1, and the change p.S1025L (rs11671248, MAF 0.08) in *NLRP11* for patient 2940. Even if *NLRP1* has a major role in IL-1β activation, not many rare variants were identified in the coding region of this gene, with the only exception being p.V939M (rs61754791, MAF 0.01) found only in patient 2942. Interestingly, however both patient 2941 and family R shared a common rare haplotype composed by different missense polymorphisms located at the C-terminal of the protein (Table 6). This haplotype has been recently associated with an impairment of IL-1β processing by the inflammasome. Common polymorphisms in *CARD8* were found in a heterozygous state in our cohort and these have been previously associated with other inflammatory system defects. These

polymorphisms were p.C10X (rs2043211, MAF 0.30) found in patients A1, B7, F1, 2940 and p.V148fs (rs140826611, MAF 0.06) in A2, F1, 2941 and 2942. Among other inflammasome genes, we discovered a heterozygous missense change in *NLRC4*, better known as IPAF, p.G786V (rs149451729, MAF 0.03) in patient 2941; a novel change in *NOD1* (p.C241G) in patient 2942, again in the NACTH domain of this protein. A table (Table 4) summarizes all identified changes in inflammasome genes for each gene family.

Although we found several rare changes in these affected patients, we wanted to investigate if these observations were statistically significant. To perform this test we took advantage of high throughput sequencing data obtained by the collection of CoLaus samples, that correspond to 400 fully exome sequenced, healthy individuals from Lausanne. These individuals were sequenced using the same facility as our PFAPA cohort, reducing background noise linked to the platform and to the machine. We compared the groups using a standard χ^2 test, the frequency of finding a rare variant (MAF<2%) in the inflammasome gene in our cohort compared to the frequency of finding a rare variant in the same genes in the 400 healthy controls. A graph (Fig. 5) summarizes the results, and we showed that none of the genes were statistically significant in our cohort compared to the control population. Values close to significance were obtained for *NLRP2, NLRP10*, and *MEFV* (respectively p<0.09, p<0.06, and p<0.08) but the small size of our cohort gives us low statistical power for this type of analysis. A summary of all analyzed variants identified in inflammasome genes per single individual is reported in Table 5.

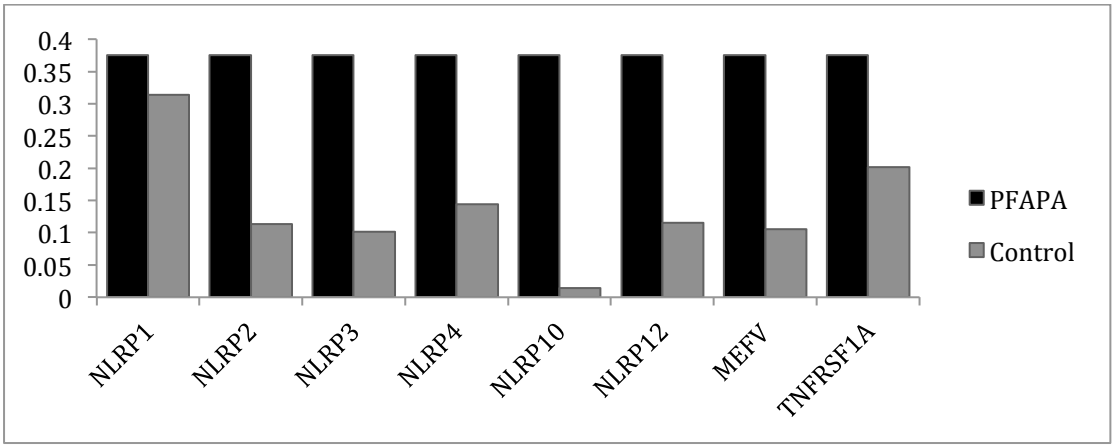


Figure 5. Frequencies of rare variants of inflammasome related genes compared to their frequencies in controls. In this graph frequencies of rare variants (MAF <2%) found at least once in a PFAPA patients are compared with the frequencies of rare variants in the same gene extracted from data obtained by 416 controls exomes. For any of these genes the statistic analysis was significant (χ^2 test analysis)

Table 4. Variants found in NLR genes and the position in the functional domains

IPAF genes					
GENE	**rs number**	**Patients**	**AA change**	**MAF**	**Domain**
IPAF (NLRC4)	149451729	2941	p.G786V	0.004	LRR6
NAIP	61757629	B7	p.A161T	0.02	BIR2
NLRP genes					
GENE	**rs number**	**Patients**	**AA change**	**MAF**	**Domain**
NLRP1	61754791	2942	p.V939M	0.02	LLR5
NLRP2	142463014	2942	p.S4L	0.01	DAPIN
	novel	B7	p.G504D	0	NACHT
	182098487	2940	p.L607P	0.005	between NACHT and LRR1
NLRP3	121908147	O5	p.V200M	0.01	between DAPIN and NACHT
	35829419	A1,A2	p.Q703K	0.05	between NACHT and LRR1
NLRP4	117212164	O1,O5	p.T162M	0.004	NACHT
NLRP5	novel	F1	p.N255S	0.001	between DAPIN and NACHT
	novel	2940	p.R274Q	0.001	between DAPIN and NACHT
NLRP10	150112481	O1	p.I384T	0.001	NACHT
NLRP11	11671248	2940	p.S1025L	0.01	C-terminal
NLRP12	novel	R1,R2	p.R211C	0.001	NACHT
NOD genes					
GENE	**rs number**	**Patients**	**AA change**	**MAF**	**Domain**
NOD1 (CARD4)	novel	2942	p.C241G	0	NACHT

NOD3 (NLRC3)	novel	2940	p.R1005Q	0	LRR15
NOD4	16965150	2941, F1	p.S210L	0.03	before NACHT domain
NOD5 (NLRX1)	150153921	B7	p.L193V	0.007	NACHT
	145779362	O1,O5	p.R547W	0.007	between NACHT and LRR1
Other NACTH domain					
GENE	**rs number**	**Patients**	**AA change**	**MAF**	**Domain**
NWD1	149694092	F1	p.S306L	0.006	Between WD and NACTH

Table 5. Variants in NLR genes per family.

FAMILY A								
Gene	**rs Number**	**NT change**	**AA change**	**MAF**	**A1**	**A2**	**SIFT**	**Polyphen**
MEFV	11466045	c.442G>C	p.E148Q	0.007	+/-	+/-	Damaging	Probably damaging
NLRP3	35829419	c.2113C>A	p.Q705K	0.04	+/-	+/-	Tolerated	Benign
CARD8	2043211	c.304T>A	p.F52I	0.3	+/-	+/-	Damaging	Damaging
	140826611	c.440_441dupAA	p.Val148LysfsX26	0.04	+/+	+/-	Damaging	Damaging
FAMILY B								
Gene	**rs Number**	**NT change**	**AA change**	**MAF**	**B1**	**B7**	**SIFT**	**Polyphen**
TNFRSF1A	4149584	c.362G>A	p.R121Q	0.018	+/-	+/-	Tolerated	Probably damaging
TNFRSF8	2230624	c.818G>A	p.C273Y	0.013	+/-	+/-	Damaging	Probably damaging
CARD8	2043211	c.304T>A	p.F52I	0.3	+/-	+/-	Damaging	Damaging

NLRP2	novel	c.1511G>A	p.G504D	-	+/+	+/-	Tolerated	Probably damaging
NAIP	61757629	c.481G>A	p.A161T	0.03	+/+	+/-	Tolerated	Probably damaging
NOD1	150153921	c.577C>G	p.L193V	0.007	+/+	+/-	Damaging	Probably damaging
FAMILY O								
Gene	**rs Number**	**NT change**	**AA change**	**MAF**	**O1**	**O5**	**SIFT**	**Polyphen**
NLRP4	117212164	c.485C>T	p.T162M	0.036	+/-	+/-	Damaging	Damaging
NLRP10	150112481	c.1151T>C	p.I384T	0	+/-	+/+	Damaging	Damaging
NOD5	145779362	c.1639C>T	p.R547W	0.007	+/-	+/-	Damaging	Probably damaging
NLRP3	121908147	c.592G>A	p.V200M	0.007	+/+	+/-	Tolerated	Benign
FAMILY R								
Gene	**rs Number**	**NT change**	**AA change**	**MAF**	**R1**	**R2**	**SIFT**	**Polyphen**
NLRP12	NOVEL		p.R211C	.	+/-	+/-	Damaging	Damaging
MEFV	104895169	c.1274A>T; c.1760-28T>A	p.F425Y	0.0012	+/-	+/+	Tolerated	Probably damaging
LPIN3	NOVEL	c.1168G>T	p.D390Y	.	+/-	+/+	Damaging	Probably damaging
	201671223	c.1807C>T	p.R603C	0.001	+/-	+/-	Tolerated	Bening
Patient F								
Gene	**rs Number**	**NT change**	**AA change**	**MAF**	**F1**	**SIFT**	**Polyphen**	
NLRP5	NOVEL	c.764A>G	p.N255S	-	+/-	Tolerated	Benign	

NOD4	16965150	c.629C>T	p.S210L	0.03	+/-	Tolerated	Benign
NWD1	149694092	c.917C>T	p.S305L	0.007	+/-	Tolerated	Benign

Patient 2940

Gene	rs Number	NT change	AA change	MAF	2940	SIFT	Polyphen
NLRP2	182098487	c.1889T>C	p.L608P	0.005	+/-	Damag.	Damag.
NLRP5	NOVEL	c.821G>A	p.R274Q	.	+/-	Activate	Benign
NLRP11	11671248	c.3074C>T	p.S1025L	0.01	+/-	Damag.	Benign
NOD3	NOVEL	c.3014G>A	p.R1005Q	.	+/-	Tolerated	-

Patient 2941

Gene	rs Number	NT change	AA change	MAF	2940	SIFT	Polyphen
IPAF	149451729	c.2357G>T	p.G786V	0.004	+/-	Tolerated	Possibly damaging
NOD4	16965150	c.629C>T	p.S210L	0.03	+/-	Tolerated	Benign
NLRP1	35596958	c.3265A>G	p.M1089V	0.05	+/-	Activate	Benign
	34733791	c.2994C>T	p.T965I	0.05	+/-	Tolerated	Benign
	11657747	c.2633C>T	p.T878M	0.05	+/-	Tolerated	Benign
	52795654	c.2345C>G	p.T782S	0.05	+/-	Tolerated	Benign
	11651595	c.737C>G	p.T246S	0.05	+/-	Tolerated	Benign
CARD8	140826611	c.440_441dupAA	p.Val148LysfsX26	0.04	+/+	Damag.	Damaging

Patient 2942

Gene	rs Number	NT change	AA change	MAF	2940	SIFT	Polyphen

NLRP1	61754791	c.2815G>A	p.V939M	0.02	+/-	Tolerated	Possibly damaging
NLRP2	142463014	c.11C>T	p.S4L	0.01	+		
NOD1	NOVEL	c.721T>G	p.C241G	.	+/-	Tolerated	Benign
CARD8	140826611	c.440_441dupAA	p.Val148LysfsX26	0.04	+/+	Damag.	Damag

II.4 DISCUSSION AND PERSPECTIVES

PFAPA syndrome is one of most common autoinflammatory periodic fever syndromes in young children [154]. The impact on quality of life of the child is not overly dramatic, with no adverse effect on growth or cognitive development. However, it creates undesired suffering to families and to patients, especially during attacks. The syndrome usually resolves after adolescence, but several adult-onset or non-resolving cases of the disease have been recently reported [178]. Due to the previously-described clinical features, the diagnosis of this disease has always been problematic, and guidelines for the correct diagnosis have rapidly changed many times during the past few years [177, 183, 197]. Clearly the etiology of the disease is linked to an impairment in the immune system, but no clear genetic base has been demonstrated up to now. Since we had a large cohort of families with PFAPA we therefore decided to perform a genetic study on these subjects.

All patients were diagnosed with PFAPA based solely on clinical observation. For a few special cases (e.g., origin of the ancestors, responsiveness to some specific treatment...) a genetic analysis of other inflammatory genes was performed (FMF, TRAPS and CAPS). Pedigrees and affected status were inferred by submitting questionnaires to other family members of the patients, and the pedigree drawing was based upon their responses. For a few cases, where another young affected subject was available, a clinical screening was performed. This approach showed some clear limitations. In fact, even though often used by physicians, a questionnaire-based diagnosis does not have a high level of specificity or sensitivity, and the effectiveness of this method mostly relies on the trustworthiness of the answer [202]. Notably, the selection of good families is crucial for a correct genetic analysis, and self-reported clinical status for family members could hinder the subsequent genetic analysis.

Based on pedigrees, we inferred an autosomal dominant model with incomplete penetrance, with a penetrance factor of almost 30%. Furthermore some families showed a fully penetrant model (e.g. family K), whereas for others a recessive status could be inferred (e.g. family F). However, we could not exclude a low penetrance model, or a multigenic or non-Mendelian inheritance pattern because of the few number of affected members and the size of each family pedigree.

We selected certain families (according to some quality values: family size and availability of samples, clinical features and presence of affected siblings or cousins), for which we performed whole genome SNPs genotyping from collected blood. The subsequent linkage analysis revealed a clear peak on the long arm of chromosome 8, notably between 8q24.1 and 8q24.3. This region contained few genes and was very close to the telomeres of chromosome 8. The validity of the identified area was confirmed also using a non-model-restricted analysis, non-parametric analysis, and by comparing genotypes of affected individuals against non-affected ones. Interestingly the identified region contained several lncRNAs and many miRNAs, as well as other unknown genes or genes without clear function. The screening of the most interesting candidate, *SLA*, did not reveal any possible pathological variant. This gene was chosen based on its function in B and T cell development [203] and, moreover, for its involvement in antigen-activated mast cells regulation [204]. Interestingly, *SLA* presents a strong responsive site to corticosteroids (corticosteroid binding site - CBS) and it is one of the most responsive genes to glucocorticoid action [205]. Unfortunately, the screening of CBS in all affected patients did not reveal any variants either. Therefore we decided to perform a custom sequence capture of all exons of all known genes in the interval. The custom capture chip was designed based on annotation of build Hg35.1 of the human genome. This implies that we missed some recently discovered genes, and all the hypothetical genes and lncRNAs that corresponded to the majority of genes present in the locus. In addition to this, some technical issues, mostly linked to the quality of the capture, hindered a correct alignment to the linkage region and did not reveal any positive hits.

As a consequence of this failure, we decided to perform a complete exome sequencing for some of the affected members of these families, plus a few extra isolated cases. We chose them based on the quality and availability of DNA. Unfortunately, the quality of the alignment was not equal for all samples, with a consequent increase of false positive rate calls for some of these samples. To decrease this rate we decided to remap the raw reads using our homemade pipeline, which had more stringent and controlled quality filters, compared to the mapping performed by the sequencing facility. We decided that all variants called by both pipelines were more likely to be truly positive, giving to them higher scores compared to variants found in only one of the two alignments. Careful screening of the whole of chromosome 8 did not reveal any particular rare or pathological variant shared by all of the affected subjects, suggesting a possible erroneous result for the chromosome 8 linkage. Conversely, if we consider the

whole exome, we identified some novel sequence variants in common genes (table 1), but a more accurate analysis of these genes revealed that they mainly represented experimental noise. As a confirmation of this, these variants were found in very high number in the same genes, and they were different between the two alignments, confirming that they are simply mistakes of alignment, maybe due to the high similarity of these gene families (e.g. mucin genes) or due to the bad annotation of the reference genome (e.g LOC or ZNF genes). Not surprisingly, the error rate was far higher in samples that presented the lowest percentage of correctly aligned reads.

Since we could not completely exclude the possibility of a misdiagnosis, we also decided to analyze rare variants that were in common in a subset of families. In particular we selected genes that had the highest number of different rare variants and with the highest quality scores. In our first analysis we excluded all genes shared only between two or three independent families and concentrated our attention on ones that were shared by more than three independent families and that were confirmed in both alignments. Among these we identified the novel variant p.G49A in the gene *Sterile alpha and TIR motif containing 1 (SARM1)*. The Protein encoded by this gene is involved in the negative regulation of the TLR pathway by the activation of NF-κB [206]. Although it was present in almost all families, in a homozygous state for some members, Sanger sequencing confirmed it as being an artifact. The second candidate was *C16orf11,* a gene with an unknown function. We identified two novel changes and a very rare variant present in some of the families (A, O, B and 2940), whereas other more common variants were represented in samples R1 and F1. The protein encoded by this gene does not have any known protein domain, and it is partially conserved among other species. We cannot propose any hypothesis on the function of this gene, and its potential involvement in PFAPA without the complete screen of ethnically matched controls. A deep functional dissection of the function of the protein for this gene is needed before any conclusions can be drawn.

We then decided to investigate the possibility that mild mutations, or rare polymorphisms in other known autoinflammatory genes could cause the PFAPA phenotype. This hypothesis is currently matter of debate, and the diagnosis of milder forms of autoinflammatory disease and PFAPA is often overlapping, leading to confusion in patients and their families. In particular, rare polymorphisms (common in 2-4% of the healthy population) have been associated to PFAPA or milder forms of autoinflammatory disease, mostly because no other strong changes in candidate genes have been identified [195, 197, 207]. Our analysis revealed that although we identified

some rare missense changes in some of the inflammatory genes, the frequencies of these variants were not different from those that were present in the general population, demonstrating that they are probably not enough alone to cause the disease. Moreover, some of these changes did not co-segregate with the phenotype, or were also present in healthy members of the family. One example is the novel change found in *NLRP12,* p.R211C, present in a well conserved domain and with a predicted strong impact on protein function. Recently a similar missense mutation (p.R352C) in the same NACTH domain of this gene was associated with genetically unexplained periodic fevers [171], with a PFAPA-like phenotype, and was shown to cause a functional defect in Caspase-1 signaling. The fact that the unaffected members of a family carried this mutant allele did not exclude the involvement of this gene in the etiology of the autoinflammatory syndrome, but it could indicate a reduced penetrance effect probably due to some genetic modifiers. It is interesting to notice that the affected patient (R1) also carried another very rare change in the gene MEFV (p.F425Y) with function unknown. The intriguing hypothesis of digenic or multigenic mutations causing the impairment of the inflammasome seemed to be confirmed by the identification of two rare polymorphisms often associated with the disease in family A. In fact we identified a change in the *MEFV* gene (p.E148Q) and in *NLRP3* (p.Q703K), both common polymorphisms with a possible functional effect on the protein. In particular, for p.Q703K, functional impairment of *NLRP3* has been clearly demonstrated [208]. Unfortunately this digenic pattern of inheritance was not identified in the other affected subjects, reducing the strength of our finding. Nevertheless we cannot exclude, based on this observation, that other rare variants present in other inflammasome-associated genes could be involved in the etiology of PFAPA. To check this possibility we carefully screened some of the genes that compose the inflammasome set (in particular the NLR family members) based on current literature [209-211].

We identified a few novel variants in *NLRP2*, *NLRP5* and *NOD1*, as well as some very rare variants in other inflammasome-associated genes, such as *NLRP10, NLRP4*, *NLRP11* and *NLRC4*. Interestingly, the majority of these genes were involved in IL-1β inflammasome activation, or in the NF-κB pathway. For example, new evidence demonstrated the high similarity and redundancy between the structure of NLRP1, NLRP6, NLRP10, NLRP3 and NLRP12 [212]. Although the function within the inflammasome for many of these members has not yet been clarified, structural similarities among NLRs suggest that all these components could somehow mediate inflammation in response to different stimuli and trigger IL-1β production [211].

Moreover, we identified two samples carrying a specific rare haplotype in *NLRP1* that has recently been associated to vitiligo and autoimmune diseases, and connected to an increase of IL-1β release in patients carrying the haplotype [213]. In addition, dominant mutations in the same gene have been shown to cause systemic inflammation in mouse models [214]. Another interesting observation from the analysis on the effect of protein structure was that the majority of identified variants seemed to be present in the NACHT or very close to the NACHT domain. This domain has an ATPase activity and is crucial for the self-oligomerization (usually heptamers or examers) of NLRs to form the structure necessary for inflammasome assembly [215]. NACTH domain mutations have been associated with structural changes that lead to a continuous activation of the inflammasome complex. A link between mutations affecting NACHT and different autoinflammatory diseases has been observed in *NLRP3* and *NOD1* for example [155, 216], reinforcing the hypothesis of a gain-of-function effect of these mutations on the complex formation. It is interesting to discover that every single affected individual in this cohort presents more than one rare variant in one of the inflammasome-composing genes. To test if this observation was statistically significant, we compared the probability of finding a rare change in one of these genes in our cohort compared to the general population. To reduce the error due to the different chemistry and output of different machines, we performed this analysis by extracting data from a collection of 400 exomes of healthy subjects from a large screening project performed in Lausanne. Unfortunately there were no statistically significant values for tested genes, although relatively low p values were seen for *NLRP10, NLRP2* and *MEFV* genes. It is likely that the size of the cohort was too small to obtain enough statistical power. Another technical limitation was linked to the fact that we could not analyze the genotype of the NLR genes in controls individuals, reducing the power of this analysis. A specific pipeline should be designed to extract this data from the VCF files of the control individuals.

To conclude, although we did not successfully identify any common genes involved in PFAPA, the fact that several rare variants were present in inflammasome-associated genes, and in particular in the NACHT domain, could suggest the implication of these in the etiology of this disease, although the analyzed dataset was too small to drive any clear conclusion. Moreover, since for many autoinflammatory diseases the causative variant has not been found, or the same mutation has been associated to different severities of the disease [155], the presence of more than one variant in the autoinflammatory genes could indicate a more complex genetic basis of these diseases.

For this reason, the exome sequencing of all these patients with unsolved autoinflammatory disease, and a more complete check of inflammasome-associated genes could help to elucidate the genotype-phenotype correlation for all of these syndromes, and maybe clarify the causes of PFAPA.

"And now here my secret, a very simple secret; is only with the heart that one can see rightly, what is essential is invisible to the eye." Antoine de Saint-Exupery, French writer, 1900

Table 6. List of novel variants found in each family.

FAMILY A

Gene	NT change	AA change	A1	A2	SIFT	Polyphen
LPCAT1	c.1241C>T	p.P414L	+/-	+/-	Damaging	Benign
PWWP2A	c.1972A>T	p.I658F	+/-	+/-	Damaging	Probably damaging
MAPK9	c.212A>C	p.Y71C	+/-	+/-	Damaging	Probably damaging
KPNA7	c.1487T>C	p.L496S	+/-	+/-	Damaging	Benign
PGM5	c.1553C>T	p.T518I	+/-	+/-	Damaging	Probably damaging
PNLIPRP1	c.854A>G	p.Y285C	+/-	+/-	Damaging	Probably damaging
SPON1	c.2044A>C	p.K682Q	+/-	+/-	Damaging	Probably damaging
E2F8	c.2461G>A	p.E821K	+/-	+/-	Damaging	Probably damaging
VWF	c.236G>T	p.G79V	+/-	+/-	Damaging	Probably damaging
TMOD2	c.365C>T	p.A122V	+/-	+/-	Damaging	Benign
MEF2A	c.1024_1026delCGA	p.342delQ	-/-	-/-	-	-
GFOD2	c.1100C>T	p.T367M	+/-	+/-	Damaging	Probably damaging
MAPK12	c.611C>T	p.T204M	+/-	+/-	Damaging	Probably damaging

FAMILY B

Gene	NT change	AA change	B1	B7	SIFT	Polyphen
NEMF	c.2353T>C	p.S785P	+/-	+/-	Damaging	Probably damaging
BRF1	c.14150A>G	p.N35S	+/-	+/-	Tolerated	Probably damaging
ERVV-1	c.129A>G	p.N432D	+/-	+/-	Activating	-

FAMILY O

Gene	NT change	AA change	O1	O5	SIFT	Polyphen
AGRN	c.1729G>A	p.E577K	+/-	+/-	Damaging	Probably Damaging
POU3F1	c.744T>G	p.D248E	+/-	+/-	Tolerated	Benign
LRRC41	c.974G>A	p.R325Q	+/-	+/-	Tolerated	Benign
ADAMTSL4	c.1432G>A	p.G478R	+/-	+/-	Damaging	Probably Damaging
ASPM	c.1951A>G	p.K651E	+/-	+/-	Tolerated	Benign
IL24	c.*130delCAAA; c.*56delCAAA		+/-	+/-		
MTMR14	c.1433+946C>T; c.1498C>T	p.R500C	+/-	+/-	Damaging	Benign
LOC100287163/ZNF717	c.392G>A	p.G131E	+/-	+/-	Tolerated	Benign
CLRN1	c.34A>G	p.M12V	+/-	+/-	Tolerated	Benign
MAGEF1	c.70G>A	p.D24N	+/-	+/-	Tolerated	Benign
SLC26A1; IDUA	c.1817C>T; c.576+2006C>T; c.299+1173G>A	p.A606V	+/-	+/-	Tolerated	Benign
ANK2	c.5459C>T; c.4399+3828C>T; c.4426+3828C>T	p.P1820L	+/-	+/-		Benign
DDX60L	c.3266A>G	p.D1089G	+/-	+/-	Tolerated	Benign
MTMR12	c.1748C>T	p.T583I	+/-	+/-	Damaging	Probably Damaging
ORC3	c.22T>A; c.451T>A	p.L151M; p.L8M	+/-	+/-	Tolerated	Benign
AIMP2	c.452C>T	p.T151M	+/-	+/-	Damaging	Possibly Damaging
PCLO	c.14965G>A	p.V4989I	+/-	+/-		Benign

FLNC	c.5200-400C>G; c.5208C>G	p.D1736E	+/-	+/-	Tolerated	Benign
RNF170; HOOK3; MIR4469	c.-858G>T; c.-8620+210C>A; c.-26411+210C>A; c.-8410C>A		+/-	+/-		
SNX16	c.869G>A; c.956G>A	p.C319Y; p.C290Y	+/-	+/-	Activating	Benign
SPAG1	c.1870A>G	p.K624E	+/-	+/-	Tolerated	Benign
CSMD3	c.6878C>T; c.6566C>T; c.6758C>T	p.T2189I; p.T2253I; p.T2293I	+/-	+/-	Damaging	Possibly Damaging
PLEC	c.3044C>T; c.2948C>T; c.3455C>T; c.3056C>T; c.3125C>T; c.3002C>T; c.2978C>T	p.P1019L; p.P1152L; p.P1001L; p.P993L; p.P1015L; p.P1042L; p.P983L	+/-	+/-		Possibly Damaging
UBAC1	c.704C>T	p.P235L	+/-	+/-	Damaging	Possibly Damaging
NOTCH1	c.6130G>A	p.A2044T	+/-	+/-	Tolerated	Possibly Damaging
RPP30	c.23A>G	p.D8G	+/-	+/-	Tolerated	Benign
ABCC2	c.3762G>A	p.W1254*	+/-	+/-		
SEC23IP	c.907G>A	p.V303I	+/-	+/-	Tolerated	Benign
MUC2	c.*5438G>T	p.C1813F	+/-	+/-		
RTN3	c.1814C>T; c.1757C>T; c.1478C>T; c.143-29675C>T; c.199+15409C>T	p.T605I; p.T586I; p.T493I	+/-	+/-	Tolerated	Benign
PVRL1	c.1318G>T	p.V440F	+/-	+/-	Damaging	Benign

M6PR	c.684C>G; c.454-503C>G	p.F228L	+/-	+/-	Tolerated	Benign
KLRC3	c.692T>C	p.I231T	+/-	+/-	Damaging	Benign
STAB2	c.3461G>A	p.R1154Q	+/-	+/-	Tolerated	Probably Damaging
ACAD10	c.772G>C; c.690+670G>C	p.V258L	+/-	+/-	Tolerated	Benign
FRY	c.724G>C	p.A242P	+/-	+/-	Tolerated	Benign
ALG5	c.267C>A	p.S89R	+/-	+/-	Tolerated	Benign
ZNF410	c.1257+1190C>T; c.1460C>T; c.810+1190C>T; c.1398+1190C>T; c.1179+1190C>T	p.P487L	+/-	+/-	Tolerated	
EIF3J	c.599C>T	p.T200I	+/-	+/-	Tolerated	Possibly Damaging
KIF7	c.1643C>T	p.P548L	+/-	+/-	Tolerated	Benign
IFT140	c.1244A>C	p.Q415P	+/-	+/-	Tolerated	Probably Damaging
SSH2	c.197G>A	p.R66Q	+/-	+/-	Tolerated	Benign
ADCYAP1	c.483G>C	p.R161S	+/-	+/-	Tolerated	Benign
ODF3L2	c.551C>T	p.T184M	+/-	+/-	Damaging	Probably Damaging
MUC16	c.22355G>C	p.S7452T	+/-	+/-		Benign
RASAL3	c.2822G>T	p.R941L	+/-	+/-	Tolerated	Benign
C19orf44	c.1567G>A	p.V523M	+/-	+/-	Tolerated	Probably Damaging
ZNF772	c.118_120delGAG	p.40delE	+/-	+/-		
ZSCAN18	c.1630G>A; c.1462G>A; c.1054G>A	p.A488T; p.A544T; p.A352T	+/-	+/-	Tolerated	Benign

ACSS1	c.2041G>A; c.2047G>A; c.1663-860G>A; c.1684G>A	p.D562N; p.D683N; p.D681N	+/-	+/-	Tolerated	Benign
MN1	c.1058C>T	p.P353L	+/-	+/-	Tolerated	Benign
LARGE	c.52C>T	p.L18F	+/-	+/-	Tolerated	Benign
mir-658; ANKRD54	c.-101C>T		+/-	+/-		
MAGEC2/MAGEC3	c.656C>T	p.T219M	+/-	+/-	Damaging	Benign

Patient F

Gene	NT change	AA change	F1	SIFT	Polyphen
MYOM3	c.3571delG	p.V1191fs	-/+		
AGL	c.3383C>T; c.3332C>T; c.3335C>T	p.A1128V; p.A1112V; p.A1111V	-/+	Damaging	Probably Damaging
CD247	c.394C>T; c.397C>T	p.R132W; p.R133W	-/+	Damaging	Probably Damaging
METTL13	c.-2050+210C>A; c.38C>A	p.S13Y	-/+	Damaging	Probably Damaging
DENND1B	c.566A>G	p.N189S	-/+	Damaging	Probably Damaging
EML6	c.5081C>G	p.A1694G	-/+	Damaging	Benign
MBD5	c.3887G>A	p.R1296Q	-/+	Activating	Benign
TTN	c.32564_32565insAGA; c.13283-37642_13283-37641insAGA; c.13658-37642_13658-37641insAGA; c.13859-37642_13859-37641insAGA;	p.9928_9929insE; p.11172_11173insE; p.10855_10856insE	-/+		

	c.29783_29784insAGA; c.33515_33516insAGA				
FAM171B	c.129_130insCAGCAG	p.43_44insQQ	-/-		
QARS	c.2290C>T	p.R764*	-/+		
TMEM45A	c.583T>G	p.F195V	-/+	Damaging	Benign
MYH15	c.758A>G	p.N253S	-/+	Damaging	Probably Damaging
IGSF10	c.4948A>T	p.I1650L	-/+	Damaging	Probably Damaging
RAP2B	c.122G>T	p.R41L	-/+	Damaging	Benign
IFT80	c.1822G>C; c.2233G>C	p.D745H; p.D608H	-/+	Damaging	Probably Damaging
CHRD	c.2390C>T	p.T797M	-/+	Damaging	Possibly Damaging
ZNF141	c.1423T>G	p.*475G	-/+		
KLB	c.2327_2329delTCT	p.777delF	-/+		
PDCL2	c.327_329delAGA	p.109delE	-/+		
DNAH5	c.6217A>G	p.M2073V	-/+	Damaging	Benign
IL6ST	c.1469T>C; c.1652T>C; c.*4243T>C	p.I551T; p.I490T	-/+	Damaging	Probably Damaging
MCC	c.63_64insGGCAGCGGC	p.21_22insGSG	-/+		

DNAJC18	c.508A>C	p.I170L	-/+	Damaging	Probably Damaging
PSD2	c.11A>G	p.D4G	-/+	Damaging	Benign
AFAP1L1	c.491G>T	p.S164I	-/+	Damaging	Benign
WWC1	c.2866C>T	p.P956S	-/+	Activating	Benign
ARL10	c.116G>A	p.R39H	-/+	Damaging	Benign
HDGFL1	c.542_543insGGC	p.181_182insA	-/+		
HIST1H2BA; HIST1H2AA	c.87_88insT; c.-469insA	p.K30fs	-/+		
DLK2	c.457G>A	p.G153S	-/+	Damaging	Probably Damaging
SNX14	c.1213C>T; c.1345C>T	p.L405F; p.L449F	-/+	Damaging	Probably Damaging
ENPP3	c.642+2T>C		-/+		
FERD3L	c.236_237insGGA	p.79_80insE	-/+		
TMEM229A	c.599_601delAGC	p.200delQ	-/+		
ST18	c.1708A>C	p.T570P	-/+	Damaging	Probably Damaging
UBE2W	c.67T>C	p.S23P	-/+		
LOC100506422	c.420+1G>A		-/+		
ALG2; SEC61B	c.-667T>C; c.19A>G	p.S7G	-/+	Damaging	Possibly Damaging

OR13C3	c.533C>A	p.S178Y	-/+	Damaging	Probably Damaging
GTF3C4	c.1807G>T	p.D603Y	-/+	Damaging	Possibly Damaging
PAPSS2	c.1848_1850delGGA; c.1833_1835delGGA	p.617delE; p.612delE	-/+		
ARL3	c.23G>A	p.R8H	-/+	Damaging	Possibly Damaging
OR51B5; OR51B6	c.-3386-5174T>C; c.577A>G	p.N193D	-/+	Damaging	Probably Damaging
PIK3C2A	c.2942C>T	p.A981V	-/+	Damaging	Probably Damaging
ROM1	c.719A>T	p.D240V	-/+	Damaging	Probably Damaging
MACROD1	c.656C>T	p.P219L	-/+	Damaging	Probably Damaging
SLC5A8	c.933G>C	p.W311C	-/+	Damaging	Probably Damaging
SGCG	c.623G>T	p.G208V	-/+	Damaging	Probably Damaging
DAOA-AS1; DAOA	c.392C>T; c.308C>T; c.179C>T	p.T131I; p.T103I; p.T60I	-/+	Damaging	Benign
SLC8A3	c.682G>A	p.V228M	-/+	Damaging	Probably Damaging
RIN3	c.2772C>G	p.F924L	-/+	Damaging	Possibly Damaging
mir-368			-/+		
NIPA2	c.1004_1005insTGA; c.947_948insTGA	p.334_335insD; p.315_316insD	-/+		
SPG21	c.683A>G; c.764A>G	p.N228S; p.N255S	-/+	Damaging	Probably Damaging

DNAH3	c.262_263insAAC	p.87_88insQ	-/+		
ERN2	c.2816_2818delTCC	p.939delL	-/+		
C16orf78	c.436C>T	p.R146W	-/+	Damaging	Probably Damaging
ENO3	c.963_964insT; c.1092_1093insT	p.W322fs; p.W365fs	-/+		
MIR4733; NF1	c.-919T>C		-/+		
FAM20A	c.727C>T; c.313C>T	p.R243*; p.R105*	-/+		
ODF3L2	c.574G>C	p.A192P	-/+	Damaging	Possibly Damaging
SAFB2	c.2290C>T	p.R764C	-/+	Damaging	Benign
STXBP2	c.1163C>A; c.1172C>A	p.P388Q; p.P391Q	-/+	Damaging	Probably Damaging
ZNF44	c.431C>T; c.287C>T	p.T144I; p.T96I	-/+	Activating	Benign
BEST2	c.101G>A	p.R34Q	-/+	Tolerated	Possibly Damaging
DDX39A	c.1072G>A	p.V358I	-/+	Damaging	Possibly Damaging
ZNF253	c.829G>A	p.V277I	-/+	Activating	Benign
FAM187B	c.758C>T	p.T253M	-/+	Damaging	Probably Damaging
PAF1	c.1190T>C; c.1404T>C	p.D468D; p.M397T	-/+	Damaging	
FCGBP	c.3553C>T	p.P1185S	-/+		Probably Damaging

CADM4	c.*50insC		-/+		
PPP1R37	c.1108T>A	p.S370T	-/+	Damaging	
PRR12	c.2396C>A	p.P799H	-/+		Probably Damaging
KRTAP19-3	c.161delG	p.G54fs	-/+		
MAP3K15	c.361G>T	p.D121Y	-/+	Damaging	
UBL4A; SLC10A3	c.-863A>G; c.*60A>G		-/+		

Patient 2940

Gene	NT change	AA change	2940	SIFT	Polyphen
TCTEX1D4	c.360_386delTGCGCGTGCACAGCGTGCCCTGGAGGC	p.121_129delARAQRALEA	-/+		
CDCP2	c.1224_1225insGC	p.M409fs	-/-		
ARHGAP29	c.2876C>T	p.A959V	-/+	Tolerated	Benign
TCHH	c.4378C>G	p.H1460D	-/+		Benign
OR10K1	c.835A>C	p.T279P	-/+	Damaging	Probably Damaging
IGSF9	c.829C>T	p.R277W	-/+	Damaging	Probably Damaging
MIR1295B; mir-1295; FMO3	c.133-2007A>G		-/+		
CFHR5	c.437A>G	p.E146G	-/+	Damaging	Possibly Damaging

DNAH14	c.1258C>G	p.R420G	-/+	Damaging	Benign
TMEM18	c.107C>T	p.A36V	-/+	Activating	Benign
NCOA1	c.484G>A	p.V162M	-/+	Damaging	Possibly Damaging
IL36B	c.13+1G>A		-/+		
IMP4	c.832T>C	p.Y278H	-/+	Damaging	Benign
LRP1B	c.5456C>T	p.S1819F	-/+	Damaging	Benign
CWC22	c.952C>T	p.R318C	-/+	Damaging	Probably Damaging
FSIP2	c.257C>T	p.A86V	-/+	Damaging	
FSIP2	c.1895A>T	p.Y632F	-/+	Damaging	
USP40	c.3152G>A	p.R1051Q	-/+		Benign
KIAA2018	c.4418_4419insGCA	p.1473_1474insQ	-/+		
GK5	c.8G>A	p.G3E	-/+	Damaging	Benign
CLCN2	c.376A>G; c.508A>G	p.I126V; p.I170V	-/+	Damaging	Possibly Damaging
TMEM175	c.124G>A	p.A42T	-/+	Damaging	Probably Damaging
DOK7	c.1120_1121insGCCT; c.190_191insGCCT; c.*341insGCCT	p.A378fs; p.A68fs	-/+		
DSPP	c.3654_3656delCAG	p.1219delS	-/+		
RRH	c.730delA	p.I244fs	-/+		
PPID	c.437_438delAT	p.H146fs	-/+		
C5orf42	c.3130C>T	p.R1044C	-/+	Damaging	Benign
ARSB	c.29C>T	p.P10L	-/+	Damaging	Benign

MCC	c.63_64insGGC	p.21_22insG	-/-		
SNX2	c.*127insACAC		-/+		
FSTL4	c.814G>T	p.V272F	-/+	Damaging	Benign
BTN2A2	c.94+1432C>T; c.427C>T	p.R143C	-/+	Damaging	Possibly Damaging
MAPK13	c.564C>G	p.Y188*	-/+		
POLR1C	c.766T>C	p.C256R	-/+	Damaging	Probably Damaging
PRIM2	c.1138G>A	p.D380N	-/+		Probably Damaging
TAAR5	c.689C>A	p.A230D	-/+	Damaging	Probably Damaging
MIR4644; FAM120B	c.2017+212C>T		-/+		
PDGFA	c.*166A>G; c.*2046A>G		-/+		
FTSJ2	c.542C>A	p.T181N	-/+	Damaging	Benign
LOC389458; RBAK-LOC389458	c.273_275delTGC	p.92delA	-/+		
SMO	c.47_48insGCT	p.16_17insL	-/+		
ZNF425	c.1376_1377insGC	p.H460fs	-/+		
MBOAT4; LEPROTL1	c.730G>A; c.280-4751C>T	p.V244M	-/+	Activating	Benign
TRPS1	c.1157A>T	p.K386I	-/+	Damaging	Benign
FER1L6	c.1607T>C	p.L536P	-/+	Damaging	Probably Damaging
NDUFB9	c.347C>T	p.P116L	-/+	Damaging	Probably Damaging
EPPK1	c.2569C>T	p.R857C	-/+	Damaging	Possibly

					Damaging
PDCD1LG2	c.542G>A	p.R181H	-/+	Damaging	Probably Damaging
TUSC1	c.142_143insGCG	p.47_48insG	-/+		
FOXE1	c.505_510delGCCG CC	p.169_170delAA	-/-		
COL27A1	c.502G>A	p.V168I	-/+	Damaging	Benign
ANK3	c.12077C>A; c.1808-795C>A; c.4409-795C>A; c.4388-795C>A	p.S4026Y	-/+	Damaging	Probably Damaging
UNC5B	c.14_25delGCGGAG CTCGGG	p.6_9delGARG	-/+		
TLL2	c.2432G>C	p.G811A	-/+	Damaging	Probably Damaging
CALHM2	c.187G>A	p.V63M	-/+	Damaging	Probably Damaging
CTBP2	c.2158G>C; c.538G>C	p.G180R; p.G720R	-/+	Damaging	Possibly Damaging
TSPAN4	c.230_232delAGG; c.38_40delAGG	p.78delE; p.14delE	-/+		
MUC2	c.4745C>G	p.T1582R	-/+		Benign
MUC2	c.*4913C>T	p.T1638I	-/+		
MUC2	c.*5394C>G	p.N1798K	-/+		
NAV2	c.4109C>T; c.6926C>T; c.6917C>T; c.7094C>T; c.6725C>T	p.S2242F; p.S2306F; p.S2309F; p.S1370F; p.S2365F	-/+	Damaging	Probably Damaging
INPPL1	c.3224C>G	p.P1075R	-/+	Damaging	Benign
USP35	c.125T>C	p.L42P	-/+	Damaging	Possibly Damaging

NCAPD2	c.3218C>G	p.S1073C	-/+	Damaging	Possibly Damaging
KRT79	c.332C>T	p.P111L	-/+	Damaging	Benign
DHX37	c.504_505insGAG	p.168_169insE	-/+		
ANKLE2	c.961G>A	p.E321K	-/+	Tolerated	Benign
TBC1D4	c.2683A>G	p.K895E	-/+	Damaging	Benign
ANKRD10	c.1003_1005delGAG	p.335delE	-/+		
MYH6	c.3508G>A	p.E1170K	-/+	Damaging	Possibly Damaging
DCAF4	c.1106C>T; c.1226C>T; c.1388C>T; c.1406C>T; c.1343C>T	p.S463L; p.S448L; p.S409L; p.S369L; p.S469L	-/+	Damaging	Possibly Damaging
mir-154			-/+		
SPINT1	c.962-1delG; c.914-1delG		-/+		
TLN2	c.1069A>G	p.T357A	-/+	Damaging	Probably Damaging
CHD2; MIR3175	c.62+3104G>C		-/+		
SYNM	c.2353_2355delAAG	p.785delK	-/+		
UNKL	c.1663C>A; c.169C>A	p.P57T; p.P555T	-/+	Activating	
COX6A2	c.203G>A	p.R68H	-/+	Damaging	Probably Damaging
SOCS7	c.535_536insAGC	p.179_180insQ	-/+		
COPZ2; mir-148	c.112-254delA		-/+		

COPZ2	c.65G>A	p.G22E	-/+		
LRRC37A3 (includes others)	c.4705-3_4705-2insT		-/+		
BAHCC1	c.1272_1274delCTT	p.H424_L425delinsQ	-/+		
BAHCC1	c.*6616T>C	p.F2206L	-/+		
TBCD;	c.1318+17390A>G;				Probably
ZNF750	c.131T>C	p.M44T	-/+	Damaging	Damaging
ZNF521	c.2603C>T	p.S868F	-/+	Damaging	Possibly Damaging
MBD2	c.345_346insGGC	p.115_116insG	-/+		
CTDP1	c.1510C>T; c.1867C>T	p.R623C; p.R504C	-/+	Damaging	Probably Damaging
ODF3L2	c.574G>C	p.A192P	-/+	Damaging	Possibly Damaging
S1PR5	c.892A>G	p.N298D	-/+	Damaging	Probably Damaging
ZNF878	c.1111C>A	p.Q371K	-/+	Damaging	Benign
ZNF490	c.752A>C	p.E251A	-/+	Damaging	Benign
JUND	c.752G>A	p.S251N	-/+	Tolerated	Probably Damaging
ZNF676	c.1511G>A	p.R504H	-/+	Damaging	Benign
ZNF675	c.1555C>T	p.R519*	-/+		
RASGRP4	c.826C>T; c.664-1487C>T; c.1117C>T; c.955-1487C>T; c.1015C>T; c.955-45C>T; c.1075C>T	p.R359C; p.R276C; p.R373C; p.R339C	-/+	Damaging	Benign
KLK5	c.485G>T	p.R162I	-/+	Damaging	Benign
NLRP5	c.821G>A	p.R274Q	-/+	Activating	Benign

ZNF446	c.1091G>A	p.G364E	-/+	Activating	Benign
CPXM1	c.1030G>A	p.G344R	-/+	Damaging	Probably Damaging
AP5S1	c.53C>T	p.T18M	-/+	Damaging	Benign
PAK7	c.100G>A	p.G34S	-/+	Damaging	Probably Damaging
CHEK2; HSCB	c.-7384delCCC; c.8_10delGGG; c.-22691delCCC	p.4delG	-/+		
NEFH	c.1960_1961insCCCCTGAGAAGGCCAAGT	p.654_655insPEKAKS	-/-		

Patient 2941					
Gene	**NT change**	**AA change**	**2941**	**SIFT**	**Polyphen**
ENO1	c.641C>T; c.362C>T	p.A121V; p.A214V	-/+	Damaging	Possibly Damaging
CYP4Z1	c.1051G>A	p.G351R	-/+	Activating	Possibly Damaging
IGSF3	c.3122_3123insGGA; c.3062_3063insGGA	p.1020_1021insE; p.1040_1041insE	-/-		
APOB	c.2722G>A	p.E908K	-/+	Damaging	Probably Damaging
MYO1B	c.*51T>C		-/+		
MYRIP	c.*148insT		-/+		
ENAM	c.2641G>A	p.G881R	-/+	Damaging	Possibly Damaging
PDZD2	c.4952G>T	p.C1651F	-/+	Activating	Benign
NSD1	c.1457C>A; c.650C>A	p.S217Y; p.S486Y	-/+	Damaging	Possibly Damaging
PRELID1; MXD3	c.*375A>T; c.*174T>A		-/+		
MUC22	c.817A>G	p.T273A	-/-		
	c.818C>A	p.T273K	-/-		
	c.837_838insGC	p.I280fs	-/-		

	c.838A>T	p.I280F	-/-		
	c.3623T>C	p.V1208A	-/-		
	c.3628G>A	p.A1210T	-/-		
	c.3629C>T	p.A1210V	-/-		
	c.3631G>A	p.E1211K	-/-		
MAFA	c.618_623delCCACCA	p.207_208delHH	-/-		
C9orf43	c.887_888insGCA	p.296_297insQ	-/-		
MUC6	c.5018T>G	p.L1673R	-/+		Benign
	c.4847T>C	p.V1616A	-/+		Possibly Damaging
MGA	c.652T>G	p.L218V	-/+	Damaging	Benign
TP53BP1	c.1347_1352delTATCCC; c.1362_1367delTATCCC	p.450_451delIP; p.455_456delIP	-/+		
ABCA8	c.4036G>A	p.V1346M	-/+	Damaging	Probably Damaging
CHMP6	c.495+2T>C		-/+		
CACNA1A	c.*185delCAG; c.6973_6975delCAG; c.6991_6993delCAG	p.2331delQ; p.2325delQ	-/-		

Patient 2942

Gene	NT change	AA change	2941	SIFT	Polyphen
ABCA4	c.3755A>T	p.E1252V	-/+	Damaging	Benign
IGFN1	c.9311C>T	p.P3104L	-/+	Damaging	
ORMDL1	c.428A>G	p.H143R	-/+	Damaging	Probably Damaging
MAP1B	c.4202C>T	p.P1401L	-/+	Activating	Possibly Damaging
RAD50	c.3454C>T	p.R1152*	-/+		
DNAH8	c.13288G>A	p.V4430M	-/+	Damaging	

mir-3689			-/+		
SYNPO2L	c.*1606C>G		-/+		
BMPR1A	c.782A>G	p.K261R	-/+	Damaging	Probably Damaging
NPAT	c.2572_2573insA	p.T858fs	-/+		
SELPLG	c.460C>G; c.412C>G	p.P138A; p.P154A	-/+	Activating	Benign
MYH6	c.4828C>T	p.R1610C	-/+	Damaging	Probably Damaging
MYH6	c.3584G>A	p.R1195H	-/+	Damaging	Probably Damaging
BRF1	c.1163_1165delGTG; c.1082_1084delGTG; c.818_820delGTG; c.449_451delGTG; c.551_553delGTG	p.273delG; p.361delG; p.150delG; p.184delG; p.388delG	-/+		
ZKSCAN2	c.2396G>T	p.G799V	-/+	Damaging	Probably Damaging
ZKSCAN2	c.1499G>C	p.W500S	-/+	Damaging	Probably Damaging
CDH20	c.*86A>G		-/+		
SLC5A5	c.643G>A	p.G215R	-/+	Damaging	Probably Damaging
ACOT8	c.622T>C	p.W208R	-/+	Damaging	Probably Damaging
P2RX6	c.1015G>A; c.937G>A	p.V339I; p.V313I	-/+	Activating	Benign

REFERENCES

1. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(12): p. 5463-7.
2. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
3. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
4. Shendure, J., et al., *Accurate multiplex polony sequencing of an evolved bacterial genome.* Science, 2005. **309**(5741): p. 1728-32.
5. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors.* Nature, 2005. **437**(7057): p. 376-80.
6. Dressman, D., et al., *Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(15): p. 8817-22.
7. Benaglio, P. and C. Rivolta, *Ultra high throughput sequencing in human DNA variation detection: a comparative study on the NDUFA3-PRPF31 region.* PloS one, 2010. **5**(9).
8. Housby, J.N. and E.M. Southern, *Fidelity of DNA ligation: a novel experimental approach based on the polymerisation of libraries of oligonucleotides.* Nucleic acids research, 1998. **26**(18): p. 4259-66.
9. Zhou, X., et al., *The next-generation sequencing technology and application.* Protein Cell, 2010. **1**(6): p. 520-36.
10. Fedurco, M., et al., *BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies.* Nucleic acids research, 2006. **34**(3): p. e22.
11. Abecasis, G.R., et al., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.
12. Schweiger, M.R., et al., *The power of NGS technologies to delineate the genome organization in cancer: from mutations to structural variations and epigenetic alterations.* Cancer Metastasis Rev, 2011. **30**(2): p. 199-210.
13. Ding, L., et al., *Analysis of next-generation genomic data in cancer: accomplishments and challenges.* Human molecular genetics, 2010. **19**(R2): p. R188-96.
14. Ng, S.B., et al., *Massively parallel sequencing and rare disease.* Human molecular genetics, 2010. **19**(R2): p. R119-24.
15. Turner, E.H., et al., *Massively parallel exon capture and library-free resequencing across 16 genomes.* Nature methods, 2009. **6**(5): p. 315-6.
16. Bamshad, M.J., et al., *Exome sequencing as a tool for Mendelian disease gene discovery.* Nature reviews. Genetics, 2011. **12**(11): p. 745-55.
17. Flicek, P. and E. Birney, *Sense from sequence reads: methods for alignment and assembly.* Nature methods, 2009. **6**(11 Suppl): p. S6-S12.

18. Li, H., J. Ruan, and R. Durbin, *Mapping short DNA sequencing reads and calling variants using mapping quality scores.* Genome research, 2008. **18**(11): p. 1851-8.
19. Li, R., et al., *SNP detection for massively parallel whole-genome resequencing.* Genome research, 2009. **19**(6): p. 1124-32.
20. Nielsen, R., et al., *Genotype and SNP calling from next-generation sequencing data.* Nature reviews. Genetics, 2011. **12**(6): p. 443-51.
21. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
22. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
23. Li, R., et al., *SOAP2: an improved ultrafast tool for short read alignment.* Bioinformatics, 2009. **25**(15): p. 1966-7.
24. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome research, 2010. **20**(9): p. 1297-303.
25. Willoughby, C.E., et al., *Anatomy and physiology of the human eye: effects of mucopolysaccharidoses disease on structure and function - a review.* Clinical and Experimental Ophthalmology, 2010. **38**: p. 2-11.
26. Mccaa, C.S., *The Eye and Visual Nervous-System - Anatomy, Physiology and Toxicology.* Environmental Health Perspectives, 1982. **44**(Apr): p. 1-8.
27. Sung, C.H. and J.Z. Chuang, *The cell biology of vision.* The Journal of cell biology, 2010. **190**(6): p. 953-63.
28. Bok, D., *The retinal pigment epithelium: a versatile partner in vision.* Journal of cell science. Supplement, 1993. **17**: p. 189-95.
29. Kandel, E.R., J.H. Schwartz, and T.M. Jessell, *Principles of neural science*. 4th ed2000, New York: McGraw-Hill, Health Professions Division. xli, 1414 p.
30. Rodieck, R.W., *Central nervous system: afferent mechanisms.* Annual review of physiology, 1971. **33**: p. 203-40.
31. Yau, K.W. and R.C. Hardie, *Phototransduction motifs and variations.* Cell, 2009. **139**(2): p. 246-64.
32. Arendt, D., *Evolution of eyes and photoreceptor cell types.* Int J Dev Biol, 2003. **47**(7-8): p. 563-71.
33. Muniz, A., et al., *A novel cone visual cycle in the cone-dominated retina.* Experimental eye research, 2007. **85**(2): p. 175-84.
34. Lamb, T.D. and E.N. Pugh, Jr., *Dark adaptation and the retinoid cycle of vision.* Progress in retinal and eye research, 2004. **23**(3): p. 307-80.
35. Molday, R.S., L.L. Molday, and C.J. Loewen, *Role of subunit assembly in autosomal dominant retinitis pigmentosa linked to mutations in peripherin 2.* Novartis Foundation symposium, 2004. **255**: p. 95-112; discussion 113-6, 177-8.
36. Roof, D.J. and J.E. Heuser, *Surfaces of rod photoreceptor disk membranes: integral membrane components.* The Journal of cell biology, 1982. **95**(2 Pt 1): p. 487-500.
37. Steinberg, R.H., S.K. Fisher, and D.H. Anderson, *Disc morphogenesis in vertebrate photoreceptors.* The Journal of comparative neurology, 1980. **190**(3): p. 501-8.
38. Molday, R.S. and L.L. Molday, *Differences in the protein composition of bovine retinal rod outer segment disk and plasma membranes isolated by a*

ricin-gold-dextran density perturbation method. The Journal of cell biology, 1987. **105**(6 Pt 1): p. 2589-601.
39. Tsacopoulos, M., et al., *Trafficking of molecules and metabolic signals in the retina.* Progress in retinal and eye research, 1998. **17**(3): p. 429-42.
40. Mazelova, J., et al., *Ciliary targeting motif VxPx directs assembly of a trafficking module through Arf4.* The EMBO journal, 2009. **28**(3): p. 183-92.
41. Tai, A.W., et al., *Rhodopsin's carboxy-terminal cytoplasmic tail acts as a membrane receptor for cytoplasmic dynein by binding to the dynein light chain Tctex-1.* Cell, 1999. **97**(7): p. 877-87.
42. Insinna, C., et al., *Analysis of a zebrafish dync1h1 mutant reveals multiple functions for cytoplasmic dynein 1 during retinal photoreceptor development.* Neural Dev, 2010. **5**: p. 12.
43. Davis, E.E., M. Brueckner, and N. Katsanis, *The emerging complexity of the vertebrate cilium: new functional roles for an ancient organelle.* Developmental cell, 2006. **11**(1): p. 9-19.
44. Satir, P. and S.T. Christensen, *Overview of structure and function of mammalian cilia.* Annual review of physiology, 2007. **69**: p. 377-400.
45. Reiter, J.F., O.E. Blacque, and M.R. Leroux, *The base of the cilium: roles for transition fibres and the transition zone in ciliary formation, maintenance and compartmentalization.* EMBO Rep, 2012. **13**(7): p. 608-18.
46. Yang, J., et al., *Rootletin, a novel coiled-coil protein, is a structural component of the ciliary rootlet.* The Journal of cell biology, 2002. **159**(3): p. 431-40.
47. Besharse, J.C., D.M. Forestner, and D.M. Defoe, *Membrane assembly in retinal photoreceptors. III. Distinct membrane domains of the connecting cilium of developing rods.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 1985. **5**(4): p. 1035-48.
48. Liu, Q., et al., *RP1 is required for the correct stacking of outer segment discs.* Investigative ophthalmology & visual science, 2003. **44**(10): p. 4171-83.
49. Gao, J., et al., *Progressive photoreceptor degeneration, outer segment dysplasia, and rhodopsin mislocalization in mice with targeted disruption of the retinitis pigmentosa-1 (Rp1) gene.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(8): p. 5698-703.
50. Liu, Q., J. Zuo, and E.A. Pierce, *The retinitis pigmentosa 1 protein is a photoreceptor microtubule-associated protein.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 2004. **24**(29): p. 6427-36.
51. Jansen, H.G., et al., *Development and degeneration of retina in rds mutant mice: ultraimmunohistochemical localization of opsin.* Experimental eye research, 1987. **44**(3): p. 347-61.
52. Sung, C.H. and A.W. Tai, *Rhodopsin trafficking and its role in retinal dystrophies.* International review of cytology, 2000. **195**: p. 215-67.
53. Kozminski, K.G., et al., *A motility in the eukaryotic flagellum unrelated to flagellar beating.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(12): p. 5519-23.
54. Rosenbaum, J.L. and G.B. Witman, *Intraflagellar transport.* Nat Rev Mol Cell Biol, 2002. **3**(11): p. 813-25.

55. Scholey, J.M., *Kinesin-II, a membrane traffic motor in axons, axonemes, and spindles.* The Journal of cell biology, 1996. **133**(1): p. 1-4.
56. Marszalek, J.R., et al., *Genetic evidence for selective transport of opsin and arrestin by kinesin-II in mammalian photoreceptors.* Cell, 2000. **102**(2): p. 175-87.
57. Mikami, A., et al., *Molecular structure of cytoplasmic dynein 2 and its distribution in neuronal and ciliated cells.* Journal of cell science, 2002. **115**(Pt 24): p. 4801-8.
58. Tsujikawa, M. and J. Malicki, *Intraflagellar transport genes are essential for differentiation and survival of vertebrate sensory neurons.* Neuron, 2004. **42**(5): p. 703-16.
59. Pazour, G.J., et al., *The intraflagellar transport protein, IFT88, is essential for vertebrate photoreceptor assembly and maintenance.* The Journal of cell biology, 2002. **157**(1): p. 103-13.
60. Keady, B.T., Y.Z. Le, and G.J. Pazour, *IFT20 is required for opsin trafficking and photoreceptor outer segment development.* Molecular biology of the cell, 2011. **22**(7): p. 921-30.
61. Zaghloul, N.A. and N. Katsanis, *Mechanistic insights into Bardet-Biedl syndrome, a model ciliopathy.* J Clin Invest, 2009. **119**(3): p. 428-37.
62. Nachury, M.V., et al., *A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis.* Cell, 2007. **129**(6): p. 1201-13.
63. Wei, Q., et al., *The BBSome controls IFT assembly and turnaround in cilia.* Nature cell biology, 2012. **14**(9): p. 950-7.
64. Hurd, T.W. and F. Hildebrandt, *Mechanisms of nephronophthisis and related ciliopathies.* Nephron. Experimental nephrology, 2011. **118**(1): p. e9-14.
65. Otto, E.A., et al., *Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy.* Nature genetics, 2010. **42**(10): p. 840-50.
66. Jiang, S.T., et al., *Essential role of nephrocystin in photoreceptor intraflagellar transport in mouse.* Human molecular genetics, 2009. **18**(9): p. 1566-77.
67. Coene, K.L., et al., *OFD1 is mutated in X-linked Joubert syndrome and interacts with LCA5-encoded lebercilin.* American journal of human genetics, 2009. **85**(4): p. 465-81.
68. Singla, V., et al., *Ofd1, a human disease gene, regulates the length and distal structure of centrioles.* Developmental cell, 2010. **18**(3): p. 410-24.
69. Valente, E.M., et al., *Mutations in CEP290, which encodes a centrosomal protein, cause pleiotropic forms of Joubert syndrome.* Nature genetics, 2006. **38**(6): p. 623-5.
70. Sayer, J.A., et al., *The centrosomal protein nephrocystin-6 is mutated in Joubert syndrome and activates transcription factor ATF4.* Nature genetics, 2006. **38**(6): p. 674-81.
71. Baala, L., et al., *Pleiotropic effects of CEP290 (NPHP6) mutations extend to Meckel syndrome.* Am J Hum Genet, 2007. **81**(1): p. 170-9.
72. Leitch, C.C., et al., *Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome.* Nature genetics, 2008. **40**(4): p. 443-8.

73. den Hollander, A.I., et al., *Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis.* American journal of human genetics, 2006. **79**(3): p. 556-61.
74. Chang, B., et al., *In-frame deletion in a novel centrosomal/ciliary protein CEP290/NPHP6 perturbs its interaction with RPGR and results in early-onset retinal degeneration in the rd16 mouse.* Human molecular genetics, 2006. **15**(11): p. 1847-57.
75. Kim, J., S.R. Krishnaswami, and J.G. Gleeson, *CEP290 interacts with the centriolar satellite component PCM-1 and is required for Rab8 localization to the primary cilium.* Human molecular genetics, 2008. **17**(23): p. 3796-805.
76. Murga-Zamalloa, C.A., et al., *Accumulation of the Raf-1 kinase inhibitory protein (Rkip) is associated with Cep290-mediated photoreceptor degeneration in ciliopathies.* The Journal of biological chemistry, 2011. **286**(32): p. 28276-86.
77. Tsang, W.Y., et al., *CP110 suppresses primary cilia formation through its interaction with CEP290, a protein deficient in human ciliary disease.* Developmental cell, 2008. **15**(2): p. 187-97.
78. Hartong, D.T., E.L. Berson, and T.P. Dryja, *Retinitis pigmentosa.* Lancet, 2006. **368**(9549): p. 1795-809.
79. Li, Z.Y., D.E. Possin, and A.H. Milam, *Histopathology of bone spicule pigmentation in retinitis pigmentosa.* Ophthalmology, 1995. **102**(5): p. 805-16.
80. Berson, E.L., P. Gouras, and M. Hoff, *Temporal aspects of the electroretinogram.* Archives of ophthalmology, 1969. **81**(2): p. 207-14.
81. den Hollander, A.I., et al., *Leber congenital amaurosis: genes, proteins and disease mechanisms.* Progress in retinal and eye research, 2008. **27**(4): p. 391-419.
82. Koenekoop, R.K., *An overview of Leber congenital amaurosis: a model to understand human retinal development.* Surv Ophthalmol, 2004. **49**(4): p. 379-98.
83. Bowne, S.J., et al., *A dominant mutation in RPE65 identified by whole-exome sequencing causes retinitis pigmentosa with choroidal involvement.* Eur J Hum Genet, 2011. **19**(10): p. 1074-81.
84. Marlhens, F., et al., *Mutations in RPE65 cause Leber's congenital amaurosis.* Nature genetics, 1997. **17**(2): p. 139-41.
85. Kremer, H., et al., *Usher syndrome: molecular links of pathogenesis, proteins and pathways.* Human molecular genetics, 2006. **15 Spec No 2**: p. R262-70.
86. Bonnet, C. and A. El-Amraoui, *Usher syndrome (sensorineural deafness and retinitis pigmentosa): pathogenesis, molecular diagnosis and therapeutic approaches.* Curr Opin Neurol, 2012. **25**(1): p. 42-9.
87. Liu, X., et al., *Myosin VIIa, the product of the Usher 1B syndrome gene, is concentrated in the connecting cilia of photoreceptor cells.* Cell motility and the cytoskeleton, 1997. **37**(3): p. 240-52.
88. Maerker, T., et al., *A novel Usher protein network at the periciliary reloading point between molecular transport machineries in vertebrate photoreceptor cells.* Human molecular genetics, 2008. **17**(1): p. 71-86.

89. van Wijk, E., et al., *Usher syndrome and Leber congenital amaurosis are molecularly linked via a novel isoform of the centrosomal ninein-like protein.* Human molecular genetics, 2009. **18**(1): p. 51-64.
90. Kersten, F.F., et al., *The mitotic spindle protein SPAG5/Astrin connects to the Usher protein network postmitotically.* Cilia, 2012. **1**(1): p. 2.
91. Sahly, I., et al., *Localization of Usher 1 proteins to the photoreceptor calyceal processes, which are absent from mice.* The Journal of cell biology, 2012. **199**(2): p. 381-99.
92. Mansergh, F.C., et al., *Retinitis pigmentosa and progressive sensorineural hearing loss caused by a C12258A mutation in the mitochondrial MTTS2 gene.* American journal of human genetics, 1999. **64**(4): p. 971-85.
93. Kajiwara, K., E.L. Berson, and T.P. Dryja, *Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci.* Science, 1994. **264**(5165): p. 1604-8.
94. Rivolta, C., et al., *Retinitis pigmentosa and allied diseases: numerous diseases, genes, and inheritance patterns.* Human molecular genetics, 2002. **11**(10): p. 1219-27.
95. den Hollander, A.I., et al., *Lighting a candle in the dark: advances in genetics and gene therapy of recessive retinal dystrophies.* The Journal of clinical investigation, 2010. **120**(9): p. 3042-53.
96. Mears, A.J., et al., *Nrl is required for rod photoreceptor development.* Nature genetics, 2001. **29**(4): p. 447-52.
97. Freund, C.L., et al., *Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor.* Cell, 1997. **91**(4): p. 543-53.
98. Cao, H., et al., *Temporal and Tissue Specific Regulation of RP-Associated Splicing Factor Genes PRPF3, PRPF31 and PRPC8-Implications in the Pathogenesis of RP.* PloS one, 2011. **6**(1): p. e15860.
99. Linder, B., et al., *Systemic splicing factor deficiency causes tissue-specific defects: a zebrafish model for retinitis pigmentosa.* Human molecular genetics, 2011. **20**(2): p. 368-77.
100. Venturini, G., et al., *CNOT3 is a modifier of PRPF31 mutations in retinitis pigmentosa with incomplete penetrance.* PLoS genetics, 2012. **8**(11): p. e1003040.
101. Seyedahmadi, B.J., et al., *Comprehensive screening of the USH2A gene in Usher syndrome type II and non-syndromic recessive retinitis pigmentosa.* Experimental eye research, 2004. **79**(2): p. 167-73.
102. Littink, K.W., et al., *Mutations in the EYS gene account for approximately 5% of autosomal recessive retinitis pigmentosa and cause a fairly homogeneous phenotype.* Ophthalmology, 2010. **117**(10): p. 2026-33, 2033 e1-7.
103. Klevering, B.J., et al., *Microarray-based mutation analysis of the ABCA4 (ABCR) gene in autosomal recessive cone-rod dystrophy and retinitis pigmentosa.* European journal of human genetics : EJHG, 2004. **12**(12): p. 1024-32.
104. Dryja, T.P., et al., *Frequency of mutations in the gene encoding the alpha subunit of rod cGMP-phosphodiesterase in autosomal recessive retinitis pigmentosa.* Investigative ophthalmology & visual science, 1999. **40**(8): p. 1859-65.

105. McLaughlin, M.E., et al., *Mutation spectrum of the gene encoding the beta subunit of rod phosphodiesterase among patients with autosomal recessive retinitis pigmentosa.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(8): p. 3249-53.
106. Hartong, D.T., et al., *Insights from retinitis pigmentosa into the roles of isocitrate dehydrogenases in the Krebs cycle.* Nature genetics, 2008. **40**(10): p. 1230-4.
107. Gu, S., et al., *Autosomal recessive retinitis pigmentosa locus RP28 maps between D2S1337 and D2S286 on chromosome 2p11-p15 in an Indian family.* Journal of medical genetics, 1999. **36**(9): p. 705-7.
108. Kumar, A., et al., *Confirmation of linkage and refinement of the RP28 locus for autosomal recessive retinitis pigmentosa on chromosome 2p14-p15 in an Indian family.* Molecular vision, 2004. **10**: p. 399-402.
109. Ramamurthy, V. and M. Cayouette, *Development and disease of the photoreceptor cilium.* Clinical genetics, 2009. **76**(2): p. 137-45.
110. Rachel, R.A., T. Li, and A. Swaroop, *Photoreceptor sensory cilia and ciliopathies: focus on CEP290, RPGR and their interacting proteins.* Cilia, 2012. **1**(1): p. 22.
111. Coene, K.L., et al., *The ciliopathy-associated protein homologs RPGRIP1 and RPGRIP1L are linked to cilium integrity through interaction with Nek4 serine/threonine kinase.* Human molecular genetics, 2011. **20**(18): p. 3592-605.
112. van Reeuwijk, J., H.H. Arts, and R. Roepman, *Scrutinizing ciliopathies by unraveling ciliary interaction networks.* Human molecular genetics, 2011. **20**(R2): p. R149-57.
113. Di Gioia, S.A., et al., *FAM161A, associated with retinitis pigmentosa, is a component of the cilia-basal body complex and interacts with proteins involved in ciliopathies.* Human molecular genetics, 2012. **21**(23): p. 5174-84.
114. Zach, F., et al., *The retinitis pigmentosa 28 protein FAM161A is a novel ciliary protein involved in intermolecular protein interaction and microtubule association.* Human molecular genetics, 2012. **21**(21): p. 4573-86.
115. Roepman, R. and U. Wolfrum, *Protein networks and complexes in photoreceptor cilia.* Sub-cellular biochemistry, 2007. **43**: p. 209-35.
116. Boldt, K., et al., *Tandem affinity purification of ciliopathy-associated protein complexes.* Methods in cell biology, 2009. **91**: p. 143-60.
117. Gloeckner, C.J., et al., *Tandem affinity purification of protein complexes from mammalian cells by the Strep/FLAG (SF)-TAP tag.* Methods in molecular biology, 2009. **564**: p. 359-72.
118. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry.* Analytical chemistry, 2003. **75**(17): p. 4646-58.
119. Letteboer, S.J. and R. Roepman, *Versatile screening for binary protein-protein interactions by yeast two-hybrid mating.* Methods in molecular biology, 2008. **484**: p. 145-59.
120. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic acids research, 2009. **37**(1): p. 1-13.

121. Gherman, A., E.E. Davis, and N. Katsanis, *The ciliary proteome database: an integrated community resource for the genetic and functional dissection of cilia.* Nature genetics, 2006. **38**(9): p. 961-2.
122. Bandah-Rozenfeld, D., et al., *Homozygosity mapping reveals null mutations in FAM161A as a cause of autosomal-recessive retinitis pigmentosa.* Am J Hum Genet, 2010. **87**(3): p. 382-91.
123. Liu, Q., et al., *The proteome of the mouse photoreceptor sensory cilium complex.* Molecular & cellular proteomics : MCP, 2007. **6**(8): p. 1299-317.
124. Venoux, M., et al., *Poc1A and Poc1B act together in human cells to ensure centriole integrity.* Journal of cell science, 2013. **126**(Pt 1): p. 163-75.
125. Sarig, O., et al., *Short stature, onychodysplasia, facial dysmorphism, and hypotrichosis syndrome is caused by a POC1A mutation.* Am J Hum Genet, 2012. **91**(2): p. 337-42.
126. Yang, Z., E.A. Roberts, and L.S. Goldstein, *Functional analysis of mouse kinesin motor Kif3C.* Molecular and cellular biology, 2001. **21**(16): p. 5306-11.
127. Muresan, V., et al., *KIF3C and KIF3A form a novel neuronal heteromeric kinesin that associates with membrane vesicles.* Molecular biology of the cell, 1998. **9**(3): p. 637-52.
128. Zhao, C., et al., *Kinesin-2 family in vertebrate ciliogenesis.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(7): p. 2388-93.
129. Bandah-Rozenfeld, D., et al., *Homozygosity mapping reveals null mutations in FAM161A as a cause of autosomal-recessive retinitis pigmentosa.* American journal of human genetics, 2010. **87**(3): p. 382-91.
130. Ramskold, D., et al., *An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data.* PLoS Comput Biol, 2009. **5**(12): p. e1000598.
131. Chen, S., et al., *Crx, a novel Otx-like paired-homeodomain protein, binds to and transactivates photoreceptor cell-specific genes.* Neuron, 1997. **19**(5): p. 1017-30.
132. Sohocki, M.M., et al., *A range of clinical phenotypes associated with mutations in CRX, a photoreceptor transcription-factor gene.* Am J Hum Genet, 1998. **63**(5): p. 1307-15.
133. Freund, C.L., et al., *De novo mutations in the CRX homeobox gene associated with Leber congenital amaurosis.* Nature genetics, 1998. **18**(4): p. 311-2.
134. Corbo, J.C., et al., *CRX ChIP-seq reveals the cis-regulatory architecture of mouse photoreceptors.* Genome research, 2010. **20**(11): p. 1512-25.
135. Ozgul, R.K., et al., *Exome sequencing and cis-regulatory mapping identify mutations in MAK, a gene encoding a regulator of ciliary length, as a cause of retinitis pigmentosa.* Am J Hum Genet, 2011. **89**(2): p. 253-64.
136. Boldt, K., et al., *Disruption of intraflagellar protein transport in photoreceptor cilia causes Leber congenital amaurosis in humans and mice.* J Clin Invest, 2011. **121**(6): p. 2169-80.
137. Follit, J.A., et al., *The intraflagellar transport protein IFT20 is associated with the Golgi complex and is required for cilia assembly.* Molecular biology of the cell, 2006. **17**(9): p. 3781-92.

138. Hosch, J., B. Lorenz, and K. Stieger, *RPGR: role in the photoreceptor cilium, human retinal disease, and gene therapy.* Ophthalmic genetics, 2011. **32**(1): p. 1-11.
139. Ashe, A., et al., *Mutations in mouse Ift144 model the craniofacial, limb and rib defects in skeletal ciliopathies.* Human molecular genetics, 2012. **21**(8): p. 1808-23.
140. Delaval, B., et al., *The cilia protein IFT88 is required for spindle orientation in mitosis.* Nature cell biology, 2011. **13**(4): p. 461-8.
141. Fischler, M.P. and W.H. Reinhart, *[Fever: friend or enemy?].* Schweiz Med Wochenschr, 1997. **127**(20): p. 864-70.
142. Dinarello, C.A., *Biologic basis for interleukin-1 in disease.* Blood, 1996. **87**(6): p. 2095-147.
143. Kumar, H., T. Kawai, and S. Akira, *Pathogen recognition by the innate immune system.* Int Rev Immunol, 2011. **30**(1): p. 16-34.
144. Sahoo, M., et al., *Role of the inflammasome, IL-1beta, and IL-18 in bacterial infections.* ScientificWorldJournal, 2011. **11**: p. 2037-50.
145. Martinon, F., K. Burns, and J. Tschopp, *The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta.* Molecular cell, 2002. **10**(2): p. 417-26.
146. Mariathasan, S., et al., *Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf.* Nature, 2004. **430**(6996): p. 213-8.
147. Miao, E.A., et al., *Pseudomonas aeruginosa activates caspase 1 through Ipaf.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(7): p. 2562-7.
148. Terra, J.K., et al., *Cutting edge: resistance to Bacillus anthracis infection mediated by a lethal toxin sensitive allele of Nalp1b/Nlrp1b.* J Immunol, 2010. **184**(1): p. 17-20.
149. Kanneganti, T.D., et al., *Pannexin-1-mediated recognition of bacterial molecules activates the cryopyrin inflammasome independent of Toll-like receptor signaling.* Immunity, 2007. **26**(4): p. 433-43.
150. Kanneganti, T.D., et al., *Critical role for Cryopyrin/Nalp3 in activation of caspase-1 in response to viral infection and double-stranded RNA.* The Journal of biological chemistry, 2006. **281**(48): p. 36560-8.
151. Touitou, I. and I. Kone-Paut, *Autoinflammatory diseases.* Best practice & research. Clinical rheumatology, 2008. **22**(5): p. 811-29.
152. Touitou, I., *Inheritance of autoinflammatory diseases: shifting paradigms and nomenclature.* J Med Genet, 2013.
153. Martinon, F., et al., *Gout-associated uric acid crystals activate the NALP3 inflammasome.* Nature, 2006. **440**(7081): p. 237-41.
154. Masters, S.L., et al., *Horror autoinflammaticus: the molecular pathophysiology of autoinflammatory disease (*).* Annu Rev Immunol, 2009. **27**: p. 621-68.
155. Aksentijevich, I., et al., *The clinical continuum of cryopyrinopathies: novel CIAS1 mutations in North American patients and a new cryopyrin model.* Arthritis Rheum, 2007. **56**(4): p. 1273-85.
156. Dowds, T.A., et al., *Cryopyrin-induced interleukin 1beta secretion in monocytic cells: enhanced activity of disease-associated mutants and requirement for ASC.* The Journal of biological chemistry, 2004. **279**(21): p. 21924-8.

157. Gershoni-Baruch, R., et al., *Familial Mediterranean fever: prevalence, penetrance and genetic drift.* Eur J Hum Genet, 2001. **9**(8): p. 634-7.
158. Chae, J.J., et al., *The B30.2 domain of pyrin, the familial Mediterranean fever protein, interacts directly with caspase-1 to modulate IL-1beta production.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(26): p. 9982-7.
159. Seshadri, S., et al., *Pyrin levels in human monocytes and monocyte-derived macrophages regulate IL-1beta processing and release.* J Immunol, 2007. **179**(2): p. 1274-81.
160. Hesker, P.R., et al., *Genetic loss of murine pyrin, the Familial Mediterranean Fever protein, increases interleukin-1beta levels.* PloS one, 2012. **7**(11): p. e51105.
161. Chae, J.J., et al., *Targeted disruption of pyrin, the FMF protein, causes heightened sensitivity to endotoxin and a defect in macrophage apoptosis.* Molecular cell, 2003. **11**(3): p. 591-604.
162. Yu, J.W., et al., *Cryopyrin and pyrin activate caspase-1, but not NF-kappaB, via ASC oligomerization.* Cell Death Differ, 2006. **13**(2): p. 236-49.
163. Chae, J.J., et al., *Gain-of-function Pyrin mutations induce NLRP3 protein-independent interleukin-1beta activation and severe autoinflammation in mice.* Immunity, 2011. **34**(5): p. 755-68.
164. Lachmann, H.J., et al., *Clinical and subclinical inflammation in patients with familial Mediterranean fever and in heterozygous carriers of MEFV mutations.* Rheumatology, 2006. **45**(6): p. 746-50.
165. Kalyoncu, M., et al., *Are carriers for MEFV mutations "healthy"?* Clin Exp Rheumatol, 2006. **24**(5 Suppl 42): p. S120-2.
166. Houten, S.M., et al., *Mutations in MVK, encoding mevalonate kinase, cause hyperimmunoglobulinaemia D and periodic fever syndrome.* Nature genetics, 1999. **22**(2): p. 175-7.
167. Hoffmann, G., et al., *Mevalonic aciduria--an inborn error of cholesterol and nonsterol isoprene biosynthesis.* The New England journal of medicine, 1986. **314**(25): p. 1610-4.
168. Kuijk, L.M., et al., *HMG-CoA reductase inhibition induces IL-1beta release through Rac1/PI3K/PKB-dependent caspase-1 activation.* Blood, 2008. **112**(9): p. 3563-73.
169. Jeru, I., et al., *Mutations in NALP12 cause hereditary periodic fever syndromes.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(5): p. 1614-9.
170. Williams, K.L., et al., *The CATERPILLER protein monarch-1 is an antagonist of toll-like receptor-, tumor necrosis factor alpha-, and Mycobacterium tuberculosis-induced pro-inflammatory signals.* The Journal of biological chemistry, 2005. **280**(48): p. 39914-24.
171. Jeru, I., et al., *Identification and functional consequences of a recurrent NLRP12 missense mutation in periodic fever syndromes.* Arthritis Rheum, 2011. **63**(5): p. 1459-64.
172. McDermott, M.F., et al., *Germline mutations in the extracellular domains of the 55 kDa TNF receptor, TNFR1, define a family of dominantly inherited autoinflammatory syndromes.* Cell, 1999. **97**(1): p. 133-44.

173. Lobito, A.A., et al., *Abnormal disulfide-linked oligomerization results in ER retention and altered signaling by TNFR1 mutants in TNFR1-associated periodic fever syndrome (TRAPS).* Blood, 2006. **108**(4): p. 1320-7.
174. Todd, I., et al., *Mutant forms of tumour necrosis factor receptor I that occur in TNF-receptor-associated periodic syndrome retain signalling functions but show abnormal behaviour.* Immunology, 2004. **113**(1): p. 65-79.
175. Huggins, M.L., et al., *Shedding of mutant tumor necrosis factor receptor superfamily 1A associated with tumor necrosis factor receptor-associated periodic syndrome: differences between cell types.* Arthritis Rheum, 2004. **50**(8): p. 2651-9.
176. Marshall, G.S., et al., *Syndrome of periodic fever, pharyngitis, and aphthous stomatitis.* The Journal of pediatrics, 1987. **110**(1): p. 43-6.
177. Padeh, S., et al., *Periodic fever, aphthous stomatitis, pharyngitis, and adenopathy syndrome: clinical characteristics and outcome.* The Journal of pediatrics, 1999. **135**(1): p. 98-101.
178. Padeh, S., N. Stoffman, and Y. Berkun, *Periodic fever accompanied by aphthous stomatitis, pharyngitis and cervical adenitis syndrome (PFAPA syndrome) in adults.* Isr Med Assoc J, 2008. **10**(5): p. 358-60.
179. Cantarini, L., et al., *A case of resistant adult-onset periodic fever, aphthous stomatitis, pharyngitis and cervical adenitis (PFAPA) syndrome responsive to anakinra.* Clin Exp Rheumatol, 2012. **30**(4): p. 593.
180. Colotto, M., et al., *PFAPA syndrome in a young adult with a history of tonsillectomy.* Intern Med, 2011. **50**(3): p. 223-5.
181. Cazzato, M., et al., *A case of adult periodic fever, aphthous stomatitis, pharyngitis, and cervical adenitis (PFAPA) syndrome associated with endocapillary proliferative glomerulonephritis.* Clin Rheumatol, 2013. **32 Suppl 1**: p. 33-6.
182. Stojanov, S., et al., *Periodic fever, aphthous stomatitis, pharyngitis, and adenitis (PFAPA) is a disorder of innate immunity and Th1 activation responsive to IL-1 blockade.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(17): p. 7148-53.
183. Vigo, G. and F. Zulian, *Periodic fevers with aphthous stomatitis, pharyngitis, and adenitis (PFAPA).* Autoimmun Rev, 2012. **12**(1): p. 52-5.
184. Kolly, L., et al., *Periodic fever, aphthous stomatitis, pharyngitis, cervical adenitis syndrome is linked to dysregulated monocyte IL-1beta production.* The Journal of allergy and clinical immunology, 2012.
185. Feder, H.M., Jr., *Cimetidine treatment for periodic fever associated with aphthous stomatitis, pharyngitis and cervical adenitis.* The Pediatric infectious disease journal, 1992. **11**(4): p. 318-21.
186. Garavello, W., M. Romagnoli, and R.M. Gaini, *Effectiveness of adenotonsillectomy in PFAPA syndrome: a randomized study.* J Pediatr, 2009. **155**(2): p. 250-3.
187. Licameli, G., et al., *Long-term surgical outcomes of adenotonsillectomy for PFAPA syndrome.* Arch Otolaryngol Head Neck Surg, 2012. **138**(10): p. 902-6.
188. Licameli, G., et al., *Effect of adenotonsillectomy in PFAPA syndrome.* Arch Otolaryngol Head Neck Surg, 2008. **134**(2): p. 136-40.

189. Hofer, M., N. Mahlaoui, and A.M. Prieur, *A child with a systemic febrile illness - differential diagnosis and management.* Best practice & research. Clinical rheumatology, 2006. **20**(4): p. 627-40.
190. Long, S.S., *Syndrome of Periodic Fever, Aphthous stomatitis, Pharyngitis, and Adenitis (PFAPA)--what it isn't. What is it?* J Pediatr, 1999. **135**(1): p. 1-5.
191. Sampaio, I.C., M.J. Rodrigo, and J.G. Monteiro Marques, *Two siblings with periodic fever, aphthous stomatitis, pharyngitis, adenitis (PFAPA) syndrome.* The Pediatric infectious disease journal, 2009. **28**(3): p. 254-5.
192. Valenzuela, P.M., et al., *Syndrome of periodic fever, aphthous stomatitis, pharyngitis, and adenitis (PFAPA) in siblings.* Clin Rheumatol, 2009. **28**(10): p. 1235-7.
193. Cochard, M., et al., *PFAPA syndrome is not a sporadic disease.* Rheumatology, 2010. **49**(10): p. 1984-7.
194. Wurster, V.M., et al., *Long-term follow-up of children with periodic fever, aphthous stomatitis, pharyngitis, and cervical adenitis syndrome.* J Pediatr, 2011. **159**(6): p. 958-64.
195. Dagan, E., et al., *MEFV, TNF1rA, CARD15 and NLRP3 mutation analysis in PFAPA.* Rheumatol Int, 2010. **30**(5): p. 633-6.
196. Kovacs, L., et al., *Elevated immunoglobulin D levels in children with PFAPA syndrome.* Neuro Endocrinol Lett, 2010. **31**(6): p. 743-6.
197. Gattorno, M., et al., *Differentiating PFAPA syndrome from monogenic periodic fevers.* Pediatrics, 2009. **124**(4): p. e721-8.
198. Toplak, N., et al., *An international registry on autoinflammatory diseases: the Eurofever experience.* Annals of the rheumatic diseases, 2012. **71**(7): p. 1177-82.
199. Firmann, M., et al., *The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome.* BMC Cardiovasc Disord, 2008. **8**: p. 6.
200. Brickell, P.M., *Immortalization of human B-lymphocytes by epstein-barr virus.* Methods in molecular biology, 1992. **8**: p. 213-8.
201. Pattnaik, S., et al., *Customisation of the exome data analysis pipeline using a combinatorial approach.* PloS one, 2012. **7**(1): p. e30080.
202. Hagen, K., et al., *The validity of questionnaire-based diagnoses: the third Nord-Trondelag Health Study 2006-2008.* J Headache Pain, 2010. **11**(1): p. 67-73.
203. Myers, M.D., L.L. Dragone, and A. Weiss, *Src-like adaptor protein down-regulates T cell receptor (TCR)-CD3 expression by targeting TCRzeta for degradation.* The Journal of cell biology, 2005. **170**(2): p. 285-94.
204. Park, S.K., H. Qiao, and M.A. Beaven, *Src-like adaptor protein (SLAP) is upregulated in antigen-stimulated mast cells and acts as a negative regulator.* Molecular immunology, 2009. **46**(10): p. 2133-9.
205. Park, S.K. and M.A. Beaven, *Mechanism of upregulation of the inhibitory regulator, src-like adaptor protein (SLAP), by glucocorticoids in mast cells.* Molecular immunology, 2009. **46**(3): p. 492-7.
206. Carty, M., et al., *The human adaptor SARM negatively regulates adaptor protein TRIF-dependent Toll-like receptor signaling.* Nature immunology, 2006. **7**(10): p. 1074-81.

207. Pelagatti, M.A., et al., *Long-term clinical profile of children with the low-penetrance R92Q mutation of the TNFRSF1A gene.* Arthritis Rheum, 2011. **63**(4): p. 1141-50.
208. Verma, D., et al., *The Q705K polymorphism in NLRP3 is a gain-of-function alteration leading to excessive interleukin-1beta and IL-18 production.* PloS one, 2012. **7**(4): p. e34977.
209. Zambetti, L.P., et al., *The rhapsody of NLRPs: master players of inflammation...and a lot more.* Immunol Res, 2012. **53**(1-3): p. 78-90.
210. Correa, R.G., S. Milutinovic, and J.C. Reed, *Roles of NOD1 (NLRC1) and NOD2 (NLRC2) in innate immunity and inflammatory diseases.* Bioscience reports, 2012. **32**(6): p. 597-608.
211. Schroder, K. and J. Tschopp, *The inflammasomes.* Cell, 2010. **140**(6): p. 821-32.
212. Anderson, J.P., et al., *Structural, expression, and evolutionary analysis of mouse CIAS1.* Gene, 2004. **338**(1): p. 25-34.
213. Levandowski, C.B., et al., *NLRP1 haplotypes associated with vitiligo and autoimmunity increase interleukin-1beta processing via the NLRP1 inflammasome.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(8): p. 2952-6.
214. Masters, S.L., et al., *NLRP1 inflammasome activation induces pyroptosis of hematopoietic progenitor cells.* Immunity, 2012. **37**(6): p. 1009-23.
215. Proell, M., et al., *The Nod-like receptor (NLR) family: a tale of similarities and differences.* PloS one, 2008. **3**(4): p. e2119.
216. Albrecht, M., et al., *Structural localization of disease-associated sequence variations in the NACHT and LRR domains of PYPAF1 and NOD2.* FEBS Lett, 2003. **554**(3): p. 520-8.

ABBREVIATIONS

List of abbreviations used in the text in alphabetical order:

ADE2	Adenine
AIS	Acquired immune system
arRP	Autosomal recessive retinitis pigmentosa
ASC	Apoptosis associated speckle protein containing CARD domain
BB	Basal body
BBS	Bardet-Biedl syndrome
bp	Base pairs
BWT	Burrows-Wheelers transform
BWA	Burrows-Wheelers transform based algorithm
CAPS	Cryopyrin associated periodic syndromes
CARD	Caspase activator and recruitment domain
CASP-1	Caspase-1
CBR	CRX binding region
CBS	Corticosteroid binding region
CC	Connecting cilium
CCD	Charge-coupled device
cGMP	Cyclic guanosin monophosphate
cM	Centimorgan
CP	Calycal process
CRBP	Cellular retinol-binding protein
CRX	Cone-rod homeobox protein
Da	Dalton
DHC1b/2	Dynein heavy chain 1b
DMEM	Dulbecco's modified Eagles medium
DMSO	Dimethyl-sulfoxide
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
ELM	External limiting membrane
EM	Electron microscopy
ER	Endoplasmic reticulum
ERG	Electroretinogram
FCS	Fetal calf serum
FMF	Familiar Mediterranean fever
g	Gravitational force
GAII	Genome analyzer II
GAL4-BD	GAL4 binding domain
GATK	Genome analysis tool kit
GDP	Guanosin diphosphate
GO	Gene ontology
GRK1	G-protein coupled receptor kinase 1

GTP	Guanosin triphosphate
h	Hour
HEK293T	Human embrionic kidney 293 Large T antigen
het	Heterozygous
hg19	Human genome reference 19
HIDS	Hyperimmunoglobulinemia syndrome
HIS3	Histidine
homo	Homozygous
HPLC	High performances liquid chromatography
IDH3B	Isocitrate dehydrogenase 3 beta
IFT	Intraflagellar transport
IIS	Innate immune system
INL	Inner nuclear layer
Indel	Insertion/deletion
IPL	Inner plexyform layer
IRBP	Inter-photoreceptor retinoid binding protein
IS	Photoreceptor inner segment
KAP1	kinesin associated protein 1
KO	Knockout
LacZ	Beta-galactosidase
LC-MS/MS	Liquid chromatography and tandem mass spectrometry
LCA	Leber congenital amaurosis
lncRNA	Long non-coding RNA
LOD	Logarithm of odds
LPS	Lipopolysaccharide
LRAT	Lecithin: retinol acyl-transferase
LRR	Leucine rich domain
MAF	Minor allele frequency
MAP	Microtubule associated protein
MAPK	Mitogen-activated protein kinase
MAQ	Mapping and assembly with quality
Mb	Megabases
MEL1	Alpha galactosidase
miRNA	Micro RNA
mM	Milli molar
MVK	Mevalonate kinase
N	Nucleus
NACHT	Conserved n NAIP, CIITA, HET-E, and TP-1 domain
NGS	Next generation sequencing
NLRs	NOD-like receptors
nm	Nanometers
NMD	Non-sense mediated mRNA decay
NOMID	Neonatal-onset mutisystem inflammatory disease
NPHP	Nephronophtisis protein
NRL	Neural retina leucine zipper
OFD1	Oro-facial digital 1

ONL	Outer nuclear layer
OPL	Outer plexyform layer
OS	Photoreceptor outer segment
PAGE	Polyacrylamide gel electrophoresis
PAMPS	Pathogen-associated molecular patterns
PBLs	Peripheral blood mononucleate cells
PCR	Polymerase chain reaction
PDE	Phospodiesterase
PFAPA	Periodic fever with aphtous, stomatitis, pharyngitis and cervical adenitis
PPM	Parts per million
PRP	Pre-mRNA processing
PYD	Pyrin domain
RGB	Red, green, blue
RNA	Ribonucleic acid
RP	Retinitis pigmentosa
RP28	Locus retinitis pigmentosa 28
RPE	Retinal pigmented epitelium
RPMI	Roswell Park Memorial Institute medium
SDS	Sodium dodecyl sulfate
SF	Streptavidin/FLAG
siRNA	Small interfering RNA
SNPs	Single nucleotide polymorphisms
SOAP2	Short oligonucleotide analysis package 2
TAP	Tandem affinity purification
TBS	Tris-phosphate buffer
TLRs	Toll like receptors
TNF	Tumor necrosis factor
TPR	Tetratricopeptide
TRAPS	TNF receptor associated syndrome
UCSC	University of California Santa Cruz
USH1	Usher syndrome 1 proteins
VCF	Variant call format file
Y2H	Yeast -two hybrid