



**HAL**  
open science

## Whole-genome sequencing in patients with ciliopathies uncovers a novel recurrent tandem duplication in IFT140

Véronique Geoffroy, Corinne Stoetzel, Sophie Scheidecker, Elise Schaefer, Isabelle Perrault, Séverine Bär, Ariane Kröll, Marion Delbarre, Manuela Antin, Anne-Sophie Leuvrey, et al.

### ► To cite this version:

Véronique Geoffroy, Corinne Stoetzel, Sophie Scheidecker, Elise Schaefer, Isabelle Perrault, et al.. Whole-genome sequencing in patients with ciliopathies uncovers a novel recurrent tandem duplication in IFT140. *Human Mutation*, 2018, 39 (7), pp.983-992. 10.1002/humu.23539 . hal-02371583

**HAL Id: hal-02371583**

**<https://hal.science/hal-02371583>**

Submitted on 19 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Whole Genome Sequencing in patients with ciliopathies uncovers a novel recurrent tandem duplication in IFT140

Journal:	<i>Human Mutation</i>
Manuscript ID	humu-2018-0045
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	29-Jan-2018
Complete List of Authors:	<p>geoffroy, veronique; Laboratoire de Génétique Médicale, Institut de Génétique Médicale d'Alsace, INSERM U1112, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg</p> <p>Stoetzel, Corinne; UMR_S INSERM U1112, IGMA, Laboratoire de Génétique médicale,</p> <p>Scheidecker, Sophie; UMR_S INSERM U1112, IGMA, Laboratoire de Génétique médicale,</p> <p>SCHAEFER, Elise; Hôpital de Hautepierre, GENETICS</p> <p>PERRAULT, Isabelle; INSERM U781, GENETICS</p> <p>Bär, Séverine; Department of Molecular and Cellular Genetics, UMR7156, Centre National de Recherche Scientifique (CNRS), Université de Strasbourg, Strasbourg, France</p> <p>Kröll, Ariane; UMR_S INSERM U1112, IGMA, Laboratoire de Génétique médicale,</p> <p>Delbarre, Marion; Hôpitaux universitaire de Strasbourg, Laboratoires de diagnostic génétique</p> <p>Antin, Manuela; Nouvel Hopital Civi, Laboratoire de Diagnostic Génétique</p> <p>Jaeger, Anne-Sophie; Hôpitaux universitaire de Strasbourg, Laboratoires de diagnostic génétique</p> <p>Henry, Charline; Inserm U1163-institut Imagine</p> <p>Blanché, Hélène; Fondation Jean-Dausset-Centre d'Etude du Polymorphisme Humain (CEPH), Institut de Génétique Moléculaire,</p> <p>Decker, Eva; Bioscentia,</p> <p>Kloth, Katja; Institut für Humangenetik, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany</p> <p>Klaus, Günter; University Marburg, KfH-Nierenzentrum für Kinder und Jugendliche, Marburg, Germany</p> <p>Mache, Christopher; Graz,</p> <p>Martin-Coignard, dominique; CH le Mans, Service de Génétique Médicale</p> <p>McGinn, Steven; CNRGH, Institut de Biologie François Jacob, DRF, CEA, Evry, France</p> <p>Boland, Anne; CNG</p> <p>Deleuze, Jean-François; CNG</p> <p>Friant, Sylvie; Department of Molecular and Cellular Genetics, UMR7156, Centre National de Recherche Scientifique (CNRS), Université de Strasbourg, Strasbourg, France</p> <p>SAUNIER, Sophie; Université René Descartes, ; INSERM U-574, Hopital Necker, Enfants Malades</p>

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

	ROZET, Jean-Michel; INSERM U781, GENETICS Bergmann, Carsten; Aachen University, Department of Human Genetics Dollfus, Helene; UMR_S INSERM U1112, IGMA, Laboratoire de Génétique médicale, Muller, Jean; Hôpitaux universitaire de Strasbourg, Laboratoires de diagnostic génétique; Laboratoire de Génétique Médicale, Institut de Génétique Médicale d'Alsace, INSERM U1112, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg,
Key Words:	Mainzer-Saldino syndrome, IFT140, Structural variation, tandem duplication, whole-genome sequencing, Alu-mediated recombination, copy number variation

SCHOLARONE™  
Manuscripts

For Peer Review

# Whole Genome Sequencing in patients with ciliopathies uncovers a novel recurrent tandem duplication in *IFT140*

Véronique Geoffroy<sup>1,\*</sup>, Corinne Stoetzel<sup>1,\*</sup>, Sophie Scheidecker<sup>1,2</sup>, Elise Schaefer<sup>1,3</sup>, Isabelle Perrault<sup>4</sup>, Séverine Bär<sup>5</sup>, Ariane Kröll<sup>1</sup>, Marion Delbarre<sup>2</sup>, Manuela Antin<sup>2</sup>, Anne-Sophie Leuvrey<sup>2</sup>, Charline Henry<sup>6</sup>, Hélène Blanché<sup>7</sup>, Eva Decker<sup>8</sup>, Katja Kloth<sup>9</sup>, Günter Klaus<sup>10</sup>, Christoph Mache<sup>11</sup>, Dominique Martin-Coignard<sup>12</sup>, Steven McGinn<sup>13</sup>, Anne Boland<sup>13</sup>, Jean-François Deleuze<sup>7,13</sup>, Sylvie Friant<sup>5</sup>, Sophie Saunier<sup>6</sup>, Jean-Michel Rozet<sup>4</sup>, Carsten Bergmann<sup>8,14</sup>, Hélène Dollfus<sup>1,15</sup>, Jean Muller<sup>1,2</sup>

<sup>1</sup>Laboratoire de Génétique médicale, UMR\_S INSERM U1112, IGMA, Faculté de Médecine FMTS, Université de Strasbourg, Strasbourg, France

<sup>2</sup>Laboratoires de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, Strasbourg Cedex, France

<sup>3</sup>Service de Génétique Médicale, Hôpitaux Universitaires de Strasbourg, Strasbourg, France.

<sup>4</sup>Laboratory of Genetics in Ophthalmology (LGO), INSERM UMR1163, Institute of Genetic Diseases, Imagine, Paris Descartes University, 75015 Paris, France.

<sup>5</sup>Department of Molecular and Cellular Genetics, UMR7156, Centre National de Recherche Scientifique (CNRS), Université de Strasbourg, Strasbourg, France.

<sup>6</sup>INSERM, U983, Paris Descartes University, Paris, France

<sup>7</sup>Centre d'études du polymorphisme humain-Fondation Jean Dausset, Paris, France

<sup>8</sup>Center for Human Genetics, Bioscientia, Ingelheim, Germany

<sup>9</sup>Institut für Humangenetik, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany

<sup>10</sup>University Marburg, KfH-Nierenzentrum für Kinder und Jugendliche, Marburg, Germany

<sup>11</sup>Department of Pediatrics, Medical University of Graz, Graz, Austria.

1  
2  
3 <sup>12</sup>Service de Génétique, Centre Hospitalier, CCLAD, Le Mans, France

4  
5 <sup>13</sup>CNRGH, Institut de Biologie François Jacob, DRF, CEA, Evry, France

6  
7 <sup>14</sup>Department of Medicine, University Hospital Freiburg, Freiburg, Germany

8  
9 <sup>15</sup>Centre de Référence pour les affections rares en génétique ophtalmologique, CARGO,  
10  
11 Filière SENSGENE, Hôpitaux Universitaires de Strasbourg, 67091 Strasbourg, France.

12  
13  
14  
15  
16  
17 \*equal contributor

18  
19  
20 Corresponding Author: Jean Muller

21  
22  
23 Email address: [jeanmuller@unistra.fr](mailto:jeanmuller@unistra.fr)  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Abstract

Ciliopathies represent a wide spectrum of rare diseases with overlapping phenotypes and a high genetic heterogeneity. Among those, *IFT140* is implicated in a variety of phenotypes ranging from isolated retinis pigmentosa to more syndromic cases such as the Bardet-Biedl syndrome. Using whole genome sequencing in patients with uncharacterized ciliopathies, we identified a novel recurrent tandem duplication of exon 27 to 30 (6.7 kb) in *IFT140*, c.3454-488\_4182+2588dup p.(Tyr1152\_Thr1394dup), missed by whole exome sequencing. Pathogenicity of the mutation was assessed on the patients' skin fibroblasts. Several hundreds of patients with a ciliopathy phenotype were screened and biallelic mutations were identified in 11 families representing 12 pathogenic variants of which 7 are novel. Among those unrelated families especially with a Mainzer-Saldino syndrome, 8 carried the same tandem duplication (2 at the homozygous state and 6 at the heterozygous state).

In conclusion, we demonstrated the implication of structural variations in *IFT140* related diseases expanding its mutation spectrum. We also provide evidences for a unique genomic event mediated by an *Alu-Alu* recombination occurring on a shared haplotype. We confirm that whole genome sequencing can be instrumental in the ability to detect structural variants for genomic disorders.

**Keywords:** *IFT140*, Mainzer-Saldino syndrome, structural variation, copy number variation, tandem duplication, *Alu*-mediated recombination, whole-genome sequencing.

## Background

Mainzer-Saldino syndrome (MSS, MIM 266920) is a rare (< 1/1,000,000) autosomal recessive ciliopathy characterized by severe early-onset retinal dystrophy, phalangeal cone-shaped epiphyses, chronic renal failure, and mild radiographic abnormality of the proximal femur (Perrault, et al., 2012), known to be caused by *IFT140* (Khan, et al., 2014; Perrault, et al., 2012; Schmidts, et al., 2013) and *IFT172* mutations (Halbritter, et al., 2013). Both genes have been implicated in several other ciliopathies ranging from isolated Retinis Pigmentosa (RP) to more syndromic cases such as Jeune or Bardet-Biedl syndromes (Bifari, et al., 2016; Bujakowska, et al., 2015; Schaefer, et al., 2016).

Intraflagellar transport (*IFT*) genes are involved in a bidirectional (anterograde and retrograde) transport process essential for the assembly and the maintenance of the cilia through the redistribution of ciliary proteins between the cell body and the cilium. In particular, *IFT140* a component of the IFT complex A (IFT-A) responsible of the retrograde IFT, is vital for both the development and the maintenance of outer segments of photoreceptors and has a specific role in opsin transport across the connecting cilium (Blacque, et al., 2006; Crouse, et al., 2014) of the photoreceptors. Previously, 46 different mutations in 56 families have been reported in *IFT140* (Additional File 1: Table S1), encompassing missenses, essential splice sites, stop and frameshifts mutations but no structural variations (SV) such as copy number variant (CNV). Interestingly, 10 mono allelic variants cases have also been reported with no second pathogenic allele detected to date (Additional File 1: Table S1).

In this study, we report for the first time a structural variation (tandem duplication) in *IFT140* identified by Whole Genome Sequencing (WGS) and missed by Whole Exome Sequencing (WES). The impact of this mutation was further assessed in patients' skin fibroblasts. Given

1  
2  
3 the difficulty to identify this mutation and the genomic context surrounding the breakpoints,  
4  
5 we speculated that it might have been missed by other genetic screenings and that several  
6  
7 other families might carry this mutation. We finally report 8 unrelated families carrying the  
8  
9 mutation either at the homozygous or the heterozygous state, out of 11 families identified with  
10  
11 *IFT140* biallelic mutations in total. Moreover, the characterization of the breakpoints allowed  
12  
13 to delineate a potential molecular mechanism and to design a specific duplex PCR that will  
14  
15 help screening further patients (including our cohorts).  
16  
17  
18  
19  
20

## 21 **Materials and Methods**

### 22 **Subjects**

23  
24  
25  
26  
27 Study protocols used in each cohort have been approved by the corresponding Institutional  
28  
29 Review Board or equivalent committees (as an example in Strasbourg, “Comité Protection des  
30  
31 Personnes” EST IV, N°DC-20142222), and written informed consent was given from each  
32  
33 participant or parents. Our research complies with the Declaration of Helsinki. Written  
34  
35 informed consent for open-access publication was provided by the participants or their  
36  
37 parents. DNA of additional affected and unaffected family members was requested whenever  
38  
39 it was considered informative. Skin fibroblasts were obtained for family A. Clinical data for  
40  
41 all 11 families are presented in Additional File 1: Table S2.  
42  
43  
44  
45  
46  
47

### 48 **Whole Genome Sequencing**

49  
50 WGS was performed for the two affected siblings from family A (A-II.1 and A-II.2) and their  
51  
52 parents (A-I.1 and A-I.2) by the Centre National de Génotypage (Institut de Génétique,  
53  
54 CEA). Genomic DNA was used to prepare a library for whole genome sequencing, using the  
55  
56 Illumina TruSeq DNA PCR-Free Library Preparation Kit, according to the manufacturer's  
57  
58  
59  
60



1  
2  
3 instructions. After normalization and quality control, qualified libraries have been sequenced  
4  
5 on a HiSeq2000 platform from Illumina (Illumina Inc., CA, USA), as paired-end 100 bp  
6  
7 reads. At least 3 lanes of HiSeq2000 flow cell have been produced for each sample, in order  
8  
9 to reach an average sequencing depth of 30x for each sample. Sequence quality parameters  
10  
11 have been assessed throughout the sequencing run and standard bioinformatics analysis of  
12  
13 sequencing data was based on the Illumina pipeline to generate FASTQ file for each sample.  
14  
15 The sequence reads were aligned to the reference sequence of the human genome (GRCh37)  
16  
17 using the Burrows-Wheeler Aligner (BWA V7.12) (Li and Durbin, 2010). The  
18  
19 UnifiedGenotyper and HaplotypeCaller modules of the Genome Analysis ToolKit (GATK)  
20  
21 (DePristo, et al., 2011), Platypus (<http://www.well.ox.ac.uk/platypus>) and Samtools (Li, et al.,  
22  
23 2009) were used for calling both single nucleotide variations (SNV) and small  
24  
25 insertion/deletion (indel).  
26  
27  
28  
29  
30

### 31 **Bioinformatics analysis**

32  
33 Annotation and ranking of SNV and indel were performed by VaRank (Geoffroy, et al., 2015)  
34  
35 in combination with the Alamut Batch software (Interactive Biosoftware, Rouen, France).  
36  
37 Very stringent filtering criteria were used for excluding non-pathogenic variants, in particular:  
38  
39 (1) variants represented with an allele frequency of more than 1% in public variation  
40  
41 databases either dbSNP 138 (Sherry, et al., 2001), the Exome Variant Server (NHLBI GO  
42  
43 Exome Sequencing Project, <http://evs.gs.washington.edu/EVS/>), the 1000Genomes (Genomes  
44  
45 Project, et al., 2015), the ExAC browser database(Lek, et al., 2016) or our internal exome  
46  
47 database, (2) variants in 5' UTR, 3' UTR, downstream, upstream or intronic locations without  
48  
49 pathogenic prediction of local splice effect, and (3) synonymous variants without prediction  
50  
51 of local splice effect. Variant effect on the nearest splice site was predicted using MaxEntScan  
52  
53 (Yeo and Burge, 2004), NNSplice (Reese, et al., 1997) and Splice Site Finder (Shapiro and  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Senapathy, 1987). Our analysis was focused on compound heterozygous and homozygous  
4 variants (SNV/indel/SV) consistent with a recessive mode of transmission. Structural variants  
5 were predicted using by default the CANOES program (Backenroth, et al., 2014) and  
6 annotated thanks to our in house script AnnotSV (manuscript in preparation,  
7 <http://www.lbgi.fr/AnnotSV/>) based on the classical annotations such as the Database of  
8 Genomic Variants (DGV) (MacDonald, et al., 2014). The *IFT140* nomenclature is based on  
9 the accession number NM\_014714.3 from the RefSeq database (O'Leary, et al., 2016).  
10 Genomic coordinates are defined according to GRCh37/hg19 assembly downloaded from the  
11 University of California Santa Cruz (UCSC) genome browser (Tyner, et al., 2017).  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

#### 24 **Sanger validation and segregation**

25  
26 Sanger sequencing was performed by PCR amplification with 50 ng of genomic DNA  
27 template. The primers were designed with Primer 3 (<http://frodo.wi.mit.edu/primer3>) and are  
28 detailed in Additional File 1: Table S3. Bidirectional sequencing of the purified PCR products  
29 was performed by GATC Sequencing Facilities (Konstanz, Germany).  
30  
31  
32  
33  
34  
35  
36

#### 37 **qPCR quantification**

38  
39 Absolute quantification was performed using the SyberGreen Mastermix Quantitect (Qiagen)  
40 measured on the LightCycler 480 (Roche). Amplicons were designed in exon 30 of *IFT140*  
41 and compared to two reference genes (*HBB* and *HMBS*). Each samples have been done in  
42 duplicate and standard deviation was <0.2. Standard curve have been established using 5  
43 different DNA quantity (50ng, 25ng, 10ng, 5ng and 1ng) and PCR efficiency was assessed for  
44 each amplicon ( $1.9 < \text{PCR efficiency} < 2.0$ ). Ratio between the mean concentrations from the  
45 amplicon of interest compared to reference gene was collected. The mean values for both  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2 control genes was calculated for each patient and plotted. The primers were designed with  
3  
4 Primer 3 and are detailed in Additional File 1: Table S3  
5  
6  
7

### 8 9 **RNA analysis**

10 RNA was extracted from skin fibroblasts of individual II.1 and a healthy unrelated control  
11  
12 using Rneasy RNA kit (Qiagen) then we performed reverse transcription using the iScript™  
13  
14 cDNA Synthesis Kit (BioRad, Hercules, CA).  
15  
16  
17

### 18 19 **Cell culture**

20 Fibroblasts of patients and control individuals were obtained by skin biopsy as previously  
21  
22 described (Scheidecker, et al., 2014). To induce primary cilium formation, cells were deprived  
23  
24 of serum by growth for 24 hrs in DMEM with 1% PSG but only 0.1% FCS (conditions -FCS)  
25  
26 as previously described (Stoetzel, et al., 2016).  
27  
28  
29

### 30 31 32 33 **Immunofluorescence**

34 Primary fibroblasts from patients and control individuals were grown in Nunc Lab-Tek  
35  
36 chamber slides (Thermo Scientific) and ciliogenesis was done as described above. Primary  
37  
38 cilia were labelled with an antibody directed against acetylated  $\alpha$ -tubulin highlighting the  
39  
40 axoneme<sup>41</sup>. Pictures were taken either on a fluorescence microscope (Figure 2A) or a  
41  
42 confocal microscope (Figure 2C). Primary and secondary antibodies used in this study as well  
43  
44 as their dilution are indicated in Additional File 1: Table S4.  
45  
46  
47  
48  
49

### 50 51 **Cohort screening**

52 A duplex PCR was designed to specifically detect the tandem duplication. Primers and  
53  
54 example conditions are provided in Additional File 1: Table S3 and Figure S1. Additional  
55  
56  
57  
58  
59  
60

1  
2  
3 targeted exome sequencing datasets have been reanalyzed (see Additional File 1:  
4  
5 supplementary Methods).  
6  
7

### 8 9 ***Alu* pair analysis**

10  
11 We obtained the annotations and sequences for the two *Alu* repeat in the duplication junction  
12  
13 region from the RepeatMasker track of the UCSC genome browser (Tyner, et al., 2017). We  
14  
15 subsequently aligned the *Alu* sequences and computed the identity using the online tool  
16  
17 EMBOSS Matcher (Rice, et al., 2000).  
18  
19

## 20 21 **Results**

### 22 23 24 25 **Identification of biallelic mutations in *IFT140***

26  
27  
28 Two affected individuals born from healthy non consanguineous parents (family A, Figure  
29  
30 1A), named II.1 and II.2, were referred to our lab with retinitis pigmentosa, short stature  
31  
32 (152cm and 158cm respectively at 16 and 15yo), brachydactyly, moderate renal failure and  
33  
34 overweight for II.1 (Additional File 1: Table S2). This Bardet-Biedl like phenotype prompted  
35  
36 us to perform an extensive first genetic analysis including Sanger sequencing of recurrent  
37  
38 BBS mutations (Muller, et al., 2010), targeted exome sequencing (Redin, et al., 2012) and  
39  
40 WES without success. Indeed no biallelic mutations including copy number variant could be  
41  
42 identified (see Additional File 1: Table S5). Hence, considering that affected individuals from  
43  
44 family A may harbor mutations in a region not (well) covered by the WES (i.e. deep intronic,  
45  
46 promotor...), we applied WGS to the 2 affected individuals and their parents. A combined  
47  
48 analysis including both SNV/indel and structural variant revealed biallelic mutations in  
49  
50 *IFT140*, a gene known to be responsible of several ciliopathies (ranging from isolated RP to  
51  
52 more syndromic cases such as Jeune or Mainzer-Saldino syndromes) compatible with the  
53  
54 phenotype of the affected patients.  
55  
56  
57  
58  
59  
60

1  
2  
3 First, the two affected individuals were found heterozygous for a possible splice mutation in  
4 intron 20, c.2577+25G>A, predicted to have a local splice effect as a new strong donor site.  
5 This variant was reported previously and is not present in any variation database (e.g. ExAC  
6 or 1000G). Co-segregation analysis revealed the paternal inheritance (Figure 1A, primers used  
7 are detailed in Additional File 1: Table S3). Patient's RNA analysis confirmed that a novel  
8 donor splice site within the intron 20 resulted in the incorporation of 21 additional bp  
9 (r.2577\_2578insGTGAGGGGCGCCCGCCATGGG) that are predicted to add 7 new amino  
10 acids to the protein sequence (p.Leu859\_Glu860insValArgGlyAlaArgHisGly) (Figure 1B).

11  
12  
13 Second, in *trans* to this mutation, structural variant analysis revealed a maternally inherited  
14 (Figure 1A) duplication of exons 27 to 30 (of sizes 6.7 kb). Analysis of the mapped reads  
15 revealed that the duplication is occurring in tandem in direct orientation within the gene  
16 (Additional File 1: Figure S2) with breakpoints located in intron 26 and intron 30. This  
17 tandem duplication has not been reported previously in patients and is not present in DGV.  
18 The results were confirmed by qPCR (Figure 1C) delineating the expected maternal  
19 inheritance. The duplication is predicted to be in frame and to add 243 amino acids within the  
20 tetratricopeptide repeat (TPR) domain of the IFT140 protein (c.3454-488\_4182+2588dup,  
21 p.Tyr1152\_Thr1394dup). RNA analysis revealed that the duplicated exons are transcribed  
22 (Figure 1D). Western blot analysis revealed only a single band of the approximate size of the  
23 wild type protein suggesting that the allele carrying the duplication is not detected as a protein  
24 (Figure 1E).

25  
26  
27 Immunofluorescence analysis in patients fibroblasts revealed that IFT140 is mislocalized  
28 compared to sex/age matched control cells. Indeed, when grown in ciliary conditions, patient  
29 skin fibroblasts developed primary cilia less often (Figure 2A). Thus while 75% of control  
30 cells developed a primary cilium, only about 50-55% of patient cells did so (Figure 2B). It is  
31 to notice that primary cilia did not seem to be altered since when cells were deprived of  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 serum, fixed for immunofluorescence and labeled with an anti-acetylated tubulin antibody,  
4  
5 their cilium as observed by fluorescence microscope and measured with image J (Schneider,  
6  
7 et al., 2012) was about 2-2,5  $\mu\text{m}$  both in control and patient cells (Figure 2B). Nevertheless, a  
8  
9 difference in organisation could be noticed when co-labelling was performed with an anti-  
10  
11 IFT140 antibodies (Figure 2C). Then, localisation of IFT140 at the base of the cilium could  
12  
13 be observed using confocal microscopy in 82% of ciliated control cells but only in about 14%  
14  
15 of ciliated patient fibroblasts (Figure 2D). Thus, the mutations found in the patients lead to a  
16  
17 significant decrease of cilium formation and a loss of IFT140 localisation at the base of the  
18  
19 cilium.  
20  
21

### 22 **Mutation screening in a large cohort**

23  
24 To explore additional cases of this duplication possibly missed by prior analysis, we  
25  
26 retrospectively screened our cohorts and reanalyzed available high throughput sequencing  
27  
28 panels including the *IFT140* gene. As mutations in *IFT140* are known to cause isolated to  
29  
30 syndromic retinal degeneration (Bifari, et al., 2016), this included 126 patients using the  
31  
32 Leber panel of the Imagine institute, 117 patients using the RP panel of Strasbourg Hospital ,  
33  
34 and 104 patients using the Ciliome panel of the Imagine institute (see Methods and Additional  
35  
36 File 1: methods online). We also setup a duplex PCR around the breakpoints (Additional File  
37  
38 1: Figure S1) and screened all our negative cases, including 207 from our Bardet-Biedl  
39  
40 cohort, 150 isolated RP and 40 *IFT140* heterozygous patients. All together we identified 11  
41  
42 additional families positive for biallelic mutations in *IFT140* (Figure 3 and Additional File 1:  
43  
44 Table S1) among which 10 families have a phenotype compatible with a Mainzer Saldino  
45  
46 syndrome and one patient presents with isolated retinitis pigmentosa. Among the MSS  
47  
48 patients, three patients presented a phenotype overlapping with another ciliopathy: two with  
49  
50 Jeune syndrome and one with Sensenbrenner syndrome. Eight of the 11 families carried the  
51  
52 exact same tandem duplication either at the heterozygous state (3 copies) or at the  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 homozygous state (4 copies) (Figure 3A, Additional File 1: Figure S3). Among the novel  
4  
5 mutations identified, we observed 2 missenses at the heterozygous state c.1319T>C  
6  
7 (p.Leu440Pro), c.2177C>T (p.Pro726Leu) that are both predicted deleterious (Additional File  
8  
9 1: supplementary Methods). Among the novel mutations we also identified another CNV, a  
10  
11 large deletion encompassing exon 27 to 29 (c.3454-1005\_4040+737delinsCCC, Additional  
12  
13 File 1: Figure S4) that does not share the same breakpoints compared to the tandem  
14  
15 duplication. Thus to date, 120 pathogenic alleles have been reported for *IFT140* of which the  
16  
17 tandem duplication represents 10 alleles. Interestingly this tandem duplication is the second  
18  
19 most frequent mutation for *IFT140*-affected patients.  
20  
21

### 22 **Founder effect or recurrent mechanism?**

23  
24  
25 Given the relatively high frequency (8/65 families) of this duplication in the patients  
26  
27 described with *IFT140* mutations, we performed 3 additional explorations to assess whether  
28  
29 the duplication was an independent recurrent mechanism or the same mutational event  
30  
31 derived from a common ancestry suggesting a founder effect. First, we screened a French  
32  
33 cohort of 354 WGS of healthy individuals (FREX dataset), for which CNV analysis  
34  
35 (CANOES) as well as visual inspections of the bam files (IGV batch scripting  
36  
37 (Thorvaldsdottir, et al., 2013)) did not reveal any corresponding allele, as expected due to the  
38  
39 *IFT140* diseases prevalence (e.g. MSS: 1/1,000,000). Second, we compared the haplotype at  
40  
41 the *IFT140* locus from the different patients for which NGS data was available. In total 15  
42  
43 haplotypes, including 6 carrying the duplication, from 3 different families, were analyzed (see  
44  
45 Additional File 1: Table S6). We observed a shared haplotype between the duplication  
46  
47 carriers at the *IFT140* locus, suggesting a founder effect. Third, we characterized the genomic  
48  
49 localization around the breakpoint junctions. Sanger sequencing confirmed the breakpoint's  
50  
51 coordinates (Figure 1A and Additional File 1: Figure S2) and demonstrated that the exact  
52  
53 same breakpoint is shared among all patients tested (Family C was not investigated). The  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 breakpoints overlapped two distinct repeated elements from the *Alu* family, *AluJb* and *AluJr*  
4 (Figure 4A). Pairwise alignment of the two *Alu* sequences (Additional File 1: Figure S5)  
5 revealed a conservation of 68% of identity that reached 81% in the junction region (Figure  
6 4B). The tandem duplication is resulting in an *Alu* hybrid (Figure 4C) at both sides. *Alu-Alu*  
7 recombination is a prominent mechanism underlying the formation of pathogenic SV  
8 associated with distinct diseases (Boone, et al., 2014; Stankiewicz, et al., 2009). Altogether,  
9 this led us to speculate that this tandem duplication is a rare event, which arises from a  
10 common ancestor due to an *Alu-Alu* mediated genomic recombination.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

## 24 Discussion

25  
26  
27 The advent of next generation sequencing, especially Whole Exome Sequencing (WES), in  
28 the past 10 years was a tremendous step that led to the identification of many mutations  
29 accelerating the discovery of novel genes involved in human genetic diseases (Chong, et al.,  
30 2015). Nevertheless, the diagnostic yield of WES is plateauing between 25% to 50%  
31 depending on the disease (Chong, et al., 2015; Taylor, et al., 2015) leaving many patients still  
32 with no molecular diagnosis. The use of RNA sequencing has recently extended the  
33 percentage by 10% to 66% (depending on the context: with or without a strong candidate by  
34 prior DNA sequencing) by helping identifying unseen defects or interpreting variations found  
35 in the WES (Cummings, et al., 2017; Kremer, et al., 2017). Lastly, Whole Genome  
36 Sequencing (WGS) of human patients has proven to be an instrumental tool for identifying  
37 the whole spectrum of genetic anomalies (Gilissen, et al., 2014) and will replace in a close  
38 future other genetic screening.  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

53 In line with this, we applied a range of genetic screening including targeted sequencing, WES  
54 and then finally WGS to unsolved patients affected with a ciliopathy. The combination of  
55  
56  
57  
58  
59  
60



1  
2  
3 WGS and CNV analysis was able to uncover a tandem duplication unseen by the WES and to  
4  
5 reconsider a distant splice site mutation.

6  
7 Interestingly, the tandem duplication could be identified in family A only by using the WGS  
8  
9 and not the WES data. Indeed, CANOES failed to identify the event in one of the individuals  
10  
11 (Additional File 1: Table S7), which prevented us from considering this as a mutation of  
12  
13 interest but made us rather discard it as a false positive of the method. Thereupon, it is well  
14  
15 known that CNV detection from next generation sequencing (NGS) datasets is feasible and  
16  
17 many publication reported positive results either on gene panels or WES (de Ligt, et al., 2013;  
18  
19 Redin, et al., 2012). However, this remains challenging (Tan, et al., 2014). WGS provides a  
20  
21 much more uniform distribution of sequencing-quality parameters and by nature does not  
22  
23 have the restriction of noncontiguous regions of interests (captured exons). Therefore, it is  
24  
25 more suited for SV identification (Belkadi, et al., 2015). Having that in mind, we further  
26  
27 tested the CNV detection using 2 other popular programs (see Additional File 1: Table S7)  
28  
29 which failed to detect the tandem duplication in family A patients (Additional File 1: Table  
30  
31 S7). These results confirmed many efforts are still required to ensure a high quality CNV  
32  
33 detection from NGS data, even with the large number of programs and methods available.  
34  
35 Nevertheless, thanks to the split reads from WGS data, once a CNV is detected one can easily  
36  
37 define the CNV breakpoints and further characterize the mechanism, as done in this study  
38  
39 (Additional File 1: Figure S2).  
40  
41  
42  
43

44 Analysis of the breakpoint positions and in particular the split reads allowed us to identify  
45  
46 overlapping *Alu* elements at the breakpoint junction. Alignment of the *Alu* regions  
47  
48 surrounding the junction revealed that both elements share 81% identity and exhibit a  
49  
50 microhomology region of 6 nucleotides. Because the recombination occurred between  
51  
52 homologous sequences (e.g. imperfect match substrates), it is most likely mediated by  
53  
54 mechanisms other than nonallelic homologous recombination, which are mostly recurrent  
55  
56  
57  
58  
59  
60

1  
2  
3 events (NAHR, for review see (Carvalho and Lupski, 2016)). Among the possible molecular  
4 mechanisms, that include non-homologous end joining (NHEJ) and replication-based  
5 mechanisms such as break-induced replication (BIR), microhomology-mediated BIR  
6 (MMBIR), serial replication slippage (SRS) and fork stalling and template switching  
7 (FoSTeS), we believe that the presence of a microhomology region points most likely to a  
8 MMBIR driven recombination. Given the nonrecurrent nature of those mechanisms, the  
9 common haplotype and the exact same breakpoint junction observed in most of our patients,  
10 we hypothesized that the genomic event that gave rise to this tandem duplication occurred  
11 once on a common *IFT140* haplotype.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21

22  
23 IFT140 is a 1462-amino-acid protein encoded by 31 exons and composed of 5 WD repeats  
24 and 9 tetratricopeptide repeats (TPR) both known to act as protein-protein interaction domains  
25 that might be involved in IFT140 interactions within the IFT-A complex (Zhu, et al., 2017).  
26  
27 To date, 45 different pathogenic alleles have already been described throughout the *IFT140*  
28 gene with no clustering or domain preference (Figure 4A). Most of them are private mutations  
29 but several are found multiple times like the c.634G>A (p.Gly212Arg) (7 alleles) which has  
30 been proven to affect the splicing of exon 6 (Helm, et al., 2017) or the c.2399+1G>T (5  
31 alleles) (Additional File 1: Table S1). The most frequent mutation (31 alleles) is a missense  
32 (c.1990G>A, p.Glu664Lys) that has been observed in multiple studies and especially in 11  
33 consanguineous families from the Arabian Peninsula sharing a common ancestor (Bifari, et  
34 al., 2016) which might bias the allele count. Nevertheless, the tandem duplication described in  
35 our study is the second most frequent cause of mutation in *IFT140* representing 10 alleles.  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

48  
49 In family A, the two mutations identified required extensive and careful analysis to be  
50 considered. One mutation was predicted to affect the splicing of exon 20 (c.2577+25G>A)  
51 which was confirmed at the RNA level (r.2577\_2578insGTGAGGGGCGCCCGCCATGGG),  
52 and suggested to add 7 amino acid to the protein sequence  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 (p.Leu859\_Glu860insValArgGlyAlaArgHisGly). The tandem duplication of exons 27 to 30  
4 (c.3454-488\_4182+2588dup, p.Tyr1152\_Thr1394dup) is predicted to add 243 amino acids  
5 within the TPR repeat domain. Skin fibroblast analysis revealed RNA synthesis in the patient  
6 cells but no protein could be identified by Western Blot at the predicted size. We hypothesize  
7 that the protein was not stable enough to maintain its 3D structure. Functional analysis on the  
8 patients' cells further revealed a reduced level of ciliated cells and mislocalization of the  
9 IFT140 mutant away from the cilia base.  
10  
11  
12  
13  
14  
15  
16  
17

18 The Mainzer Saldino syndrome is a syndrome characterized by skeletal phenotype with  
19 phalangeal cone-shaped epiphyses, chronic renal disease, and retinal dystrophy. Overlapping  
20 phenotypes have been described with other skeletal ciliopathies such as the Jeune or  
21 Sensenbrenner syndrome. In our cohort the diagnosis of MSS was not made in all patients  
22 before the molecular analysis. However, all of them except one with an isolated retinitis  
23 pigmentosa, fulfilled the criteria for this diagnosis after a new careful clinical and radiological  
24 examination. Indeed, 11 patients presented digits anomalies, including 8 with phalangeal  
25 cone-shaped epiphyses. Nine patients developed renal failure among which 7 have a severe  
26 and early onset renal disease (1.5 years old to 23 years old) as previously described  
27 (Schmidts, et al., 2013). The retinal dystrophy is constantly reported with a highly variable  
28 age of onset. Among our cohort, a single patient presented an isolated retinitis pigmentosa (as  
29 previously reported by others (Hull, et al., 2016)). Two patients have thoracic dystrophy  
30 which is a clinical criteria of Jeune syndrome and one patient has a craniosynostosis  
31 compatible with a Sensenbrenner syndrome. These descriptions illustrate well the overlap  
32 existing between the different ciliopathies associated with mutation in *IFT140*. On a  
33 molecular level, no genotype-phenotype correlation could be established in our cohort,  
34 especially regarding the patients carrying the duplication at the homozygous or the  
35 heterozygous state.  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Conclusions

In summary, we report here 11 novel unrelated families with mutations in *IFT140*. Among them, 8 families carry a recurrent tandem duplication of 4 exons either at the heterozygous state or the homozygous state, for which we have assessed the pathogenicity in the patients cells. This is the first time that a structural variation is reported in *IFT140* expanding the mutation spectrum for this gene. Notably, this large duplication was missed by the WES analysis but uncovered thanks to the whole genome, pointing out the power of such analysis.

## Availability of data and material

Data generated or analyzed during this study are included in the published article and the corresponding supplementary data. The raw sequencing data generated in the course of this study are not publicly available due to the protocol and the corresponding consents used that did not include such information. All variants have been submitted to ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>). Anonymised NGS data and genomic variant data files will be made available upon request from qualified investigators studying the molecular basis of genomic disorders. Datasets can be obtained via the corresponding author on reasonable request.

## Acknowledgements

We would like to thank the patients and their family for their participation. We thank Daniel Backenroth and Olivier Quenez for their help in the bioinformatics setup of CANOES, Arnaud Kress and Antony Le Béhec for informatic support, Amélie Piton, Bénédicte Gérard and Ilia Humbert for scientific discussion. We also thank Emmanuelle Génin, Pierre Lindenbaum and Richard Redon for giving us access to the genome dataset from the FREX

1  
2  
3 project and the Genomic and Bioinformatic Platform of Institut Imagine (Patrick Nitschké,  
4  
5 and Cécile Masson).  
6  
7

### 8 **Disclosure Statment**

9  
10 Eva Decker and Carsten Bergmann are employees of Bioscientia/Sonic Healthcare. Günter  
11  
12 Klaus benefited from a travel grants from The Apheresis Research Institute Cologne, the  
13  
14 DGFF (Die Deutsche Gesellschaft zur Bekämpfung von Fettstoffwechselstörungen und ihren  
15  
16 Folgeerkrankungen) and is consultant to Vifor pharma, Switzerland concerning a pediatric  
17  
18 phosphat binder trial.  
19  
20  
21

### 22 **Funding Information**

23  
24  
25 The WGS research was supported by the Laboratory of Excellence GENMED (Medical  
26  
27 Genomics) grant no. ANR-10-LABX-0013 managed by the National Research Agency  
28  
29 (ANR) part of the Investment for the Future program. Whole exome sequencing was  
30  
31 performed by the IGBMC Microarray and Sequencing platform, a member of the ‘France  
32  
33 Génomique’ consortium (ANR-10-INBS-0009) and funded by “La Fondation Maladie Rare”.  
34  
35  
36 The RP panel is supported by the French program PHRC I 2013 HUS N° 5724 and by grants  
37  
38 from the Retina France to IP, UNADEV- AVIESAN ITMO MNP to JMR. In addition, CB  
39  
40 holds a part-time faculty appointment at the University of Freiburg. His research lab receives  
41  
42 support from the Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Centre  
43  
44 (SFB) KIDGEM 1140 and the Federal Ministry of Education and Research (BMBF,  
45  
46 01GM1515C).  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

- Backenroth D, Homsy J, Murillo LR, Glessner J, Lin E, Brueckner M, Lifton R, Goldmuntz E, Chung WK, Shen Y. 2014. CANOES: detecting rare copy number variants from whole exome sequencing data. *Nucleic Acids Res* 42(12):e97.
- Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, Shang L, Boisson B, Casanova JL, Abel L. 2015. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* 112(17):5473-8.
- Bifari IN, Elkhamary SM, Bolz HJ, Khan AO. 2016. The ophthalmic phenotype of IFT140-related ciliopathy ranges from isolated to syndromic congenital retinal dystrophy. *Br J Ophthalmol* 100(6):829-33.
- Blacque OE, Li C, Inglis PN, Esmail MA, Ou G, Mah AK, Baillie DL, Scholey JM, Leroux MR. 2006. The WD repeat-containing protein IFTA-1 is required for retrograde intraflagellar transport. *Mol Biol Cell* 17(12):5053-62.
- Boone PM, Yuan B, Campbell IM, Scull JC, Withers MA, Baggett BC, Beck CR, Shaw CJ, Stankiewicz P, Moretti P and others. 2014. The Alu-rich genomic architecture of SPAST predisposes to diverse and functionally distinct disease-associated CNV alleles. *Am J Hum Genet* 95(2):143-61.
- Bujakowska KM, Zhang Q, Siemiatkowska AM, Liu Q, Place E, Falk MJ, Consugar M, Lancelot ME, Antonio A, Lonjou C and others. 2015. Mutations in IFT172 cause isolated retinal degeneration and Bardet-Biedl syndrome. *Hum Mol Genet* 24(1):230-42.
- Carvalho CM, Lupski JR. 2016. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet* 17(4):224-38.
- Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, Harrell TM, McMillin MJ, Wiszniewski W, Gambin T and others. 2015. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet* 97(2):199-215.
- Crouse JA, Lopes VS, Sanagustin JT, Keady BT, Williams DS, Pazour GJ. 2014. Distinct functions for IFT140 and IFT20 in opsin transport. *Cytoskeleton (Hoboken)* 71(5):302-10.
- Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, Bolduc V, Waddell LB, Sandaradura SA, O'Grady GL and others. 2017. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med* 9(386).
- de Ligt J, Boone PM, Pfundt R, Vissers LE, Richmond T, Geoghegan J, O'Moore K, de Leeuw N, Shaw C, Brunner HG and others. 2013. Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum Mutat* 34(10):1439-48.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M and others. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491-8.
- Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA and others. 2015. A global reference for human genetic variation. *Nature* 526(7571):68-74.
- Geoffroy V, Pizot C, Redin C, Piton A, Vasli N, Stoetzel C, Blavier A, Laporte J, Muller J. 2015. VaRank: a simple and powerful tool for ranking genetic variants. *PeerJ* 3:e796.
- Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, Kwint M, Janssen IM, Hoischen A, Schenck A and others. 2014. Genome sequencing identifies major causes of severe intellectual disability. *Nature* 511(7509):344-7.
- Halbritter J, Bizet AA, Schmidts M, Porath JD, Braun DA, Gee HY, McInerney-Leo AM, Krug P, Filhol E, Davis EE and others. 2013. Defects in the IFT-B component IFT172 cause Jeune and Mainzer-Saldino syndromes in humans. *Am J Hum Genet* 93(5):915-25.



- 1  
2  
3 Helm BM, Willer JR, Sadeghpour A, Golzio C, Crouch E, Vergano SS, Katsanis N, Davis  
4 EE. 2017. Partial uniparental isodisomy of chromosome 16 unmasks a deleterious biallelic  
5 mutation in IFT140 that causes Mainzer-Saldino syndrome. *Hum Genomics* 11(1):16.
- 6 Hull S, Owen N, Islam F, Tracey-White D, Plagnol V, Holder GE, Michaelides M, Carss K,  
7 Raymond FL, Rozet JM and others. 2016. Nonsyndromic Retinal Dystrophy due to Bi-Allelic  
8 Mutations in the Ciliary Transport Gene IFT140. *Invest Ophthalmol Vis Sci* 57(3):1053-62.
- 9 Khan AO, Bolz HJ, Bergmann C. 2014. Early-onset severe retinal dystrophy as the initial  
10 presentation of IFT140-related skeletal ciliopathy. *J AAPOS* 18(2):203-5.
- 11 Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, Haack TB, Graf E,  
12 Schwarzmayr T, Terrile C and others. 2017. Genetic diagnosis of Mendelian disorders via  
13 RNA sequencing. *Nat Commun* 8:15824.
- 14 Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH,  
15 Ware JS, Hill AJ, Cummings BB and others. 2016. Analysis of protein-coding genetic  
16 variation in 60,706 humans. *Nature* 536(7616):285-91.
- 17 Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler  
18 transform. *Bioinformatics* 26(5):589-95.
- 19 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin  
20 R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and  
21 SAMtools. *Bioinformatics* 25(16):2078-9.
- 22 MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. 2014. The Database of Genomic  
23 Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res*  
24 42(Database issue):D986-92.
- 25 Muller J, Stoetzel C, Vincent MC, Leitch CC, Laurier V, Danse JM, Helle S, Marion V,  
26 Bennouna-Greene V, Vicaire S and others. 2010. Identification of 28 novel mutations in the  
27 Bardet-Biedl syndrome genes: the burden of private mutations in an extensively  
28 heterogeneous disease. *Hum Genet* 127(5):583-93.
- 29 O'Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, Rajput B, Robbertse  
30 B, Smith-White B, Ako-Adjei D and others. 2016. Reference sequence (RefSeq) database at  
31 NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*  
32 44(D1):D733-45.
- 33 Perrault I, Saunier S, Hanein S, Filhol E, Bizet AA, Collins F, Salih MA, Gerber S, Delphin  
34 N, Bigot K and others. 2012. Mainzer-Saldino syndrome is a ciliopathy caused by IFT140  
35 mutations. *Am J Hum Genet* 90(5):864-70.
- 36 Redin C, Le Gras S, Mhamdi O, Geoffroy V, Stoetzel C, Vincent MC, Chiurazzi P, Lacombe  
37 D, Ouertani I, Petit F and others. 2012. Targeted high-throughput sequencing for diagnosis of  
38 genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alstrom  
39 syndromes. *J Med Genet* 49(8):502-12.
- 40 Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie.  
41 *J Comput Biol* 4(3):311-23.
- 42 Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open  
43 Software Suite. *Trends Genet* 16(6):276-7.
- 44 Schaefer E, Stoetzel C, Scheidecker S, Geoffroy V, Prasad MK, Redin C, Missotte I,  
45 Lacombe D, Mandel JL, Muller J and others. 2016. Identification of a novel mutation  
46 confirms the implication of IFT172 (BBS20) in Bardet-Biedl syndrome. *J Hum Genet*  
47 61(5):447-50.
- 48 Scheidecker S, Etard C, Pierce NW, Geoffroy V, Schaefer E, Muller J, Chennen K, Flori E,  
49 Pelletier V, Poch O and others. 2014. Exome sequencing of Bardet-Biedl syndrome patient  
50 identifies a null mutation in the BBSome subunit BBIP1 (BBS18). *J Med Genet* 51(2):132-6.
- 51 Schmidts M, Frank V, Eisenberger T, Al Turki S, Bizet AA, Antony D, Rix S, Decker C,  
52 Bachmann N, Bald M and others. 2013. Combined NGS approaches identify mutations in the  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 intraflagellar transport gene IFT140 in skeletal ciliopathies with early progressive kidney  
4 Disease. *Hum Mutat* 34(5):714-24.
- 5 Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image  
6 analysis. *Nat Methods* 9(7):671-5.
- 7 Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes:  
8 sequence statistics and functional implications in gene expression. *Nucleic Acids Res*  
9 15(17):7155-74.
- 10 Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001.  
11 dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29(1):308-11.
- 12 Stankiewicz P, Sen P, Bhatt SS, Storer M, Xia Z, Bejjani BA, Ou Z, Wiszniewska J, Driscoll  
13 DJ, Maisenbacher MK and others. 2009. Genomic and genic deletions of the FOX gene  
14 cluster on 16q24.1 and inactivating mutations of FOXF1 cause alveolar capillary dysplasia  
15 and other malformations. *Am J Hum Genet* 84(6):780-91.
- 16 Stoetzel C, Bar S, De Craene JO, Scheidecker S, Etard C, Chicher J, Reck JR, Perrault I,  
17 Geoffroy V, Chennen K and others. 2016. A mutation in VPS15 (PIK3R4) causes a ciliopathy  
18 and affects IFT20 release from the cis-Golgi. *Nat Commun* 7:13586.
- 19 Tan R, Wang Y, Kleinstein SE, Liu Y, Zhu X, Guo H, Jiang Q, Allen AS, Zhu M. 2014. An  
20 evaluation of copy number variation detection tools from whole-exome sequencing data. *Hum*  
21 *Mutat* 35(7):899-907.
- 22 Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, Kanapin A, Lunter G,  
23 Fiddy S, Allan C and others. 2015. Factors influencing success of clinical genome sequencing  
24 across a broad spectrum of disorders. *Nat Genet* 47(7):717-726.
- 25 Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV):  
26 high-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178-92.
- 27 Tyner C, Barber GP, Casper J, Clawson H, Diekhans M, Eisenhart C, Fischer CM, Gibson D,  
28 Gonzalez JN, Guruvadoo L and others. 2017. The UCSC Genome Browser database: 2017  
29 update. *Nucleic Acids Res* 45(D1):D626-D634.
- 30 Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to  
31 RNA splicing signals. *J Comput Biol* 11(2-3):377-94.
- 32 Zhu B, Zhu X, Wang L, Liang Y, Feng Q, Pan J. 2017. Functional exploration of the IFT-A complex in  
33 intraflagellar transport and ciliogenesis. *PLoS Genet* 13(2):e1006627.
- 34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



## Figure legends

### Figure 1. Identification of 2 mutations in *IFT140* for a family explored by WGS.

(A) Pedigree of family A, which has 2 affected individuals. Segregation analysis of the 2 mutations noted M1: 2577+25G>A and M2: 3454-488\_4182+2588dup. Example of Sanger sequencing profiles for individual II.1. The breakpoint junction between exon 30 and intron 26 are shown.

(B) M1 analysis. Sanger sequencing was performed on RNA extracted from fibroblasts of individual II.1 and a healthy unrelated control amplified between exon 18 and exon 22.

(C) M2 analysis. Quantitative real-time PCR was performed on DNA from exon 30 in all individuals from family A and one unrelated control. DNA quantity from exon 30 of *IFT140* was compared to 2 reference genes (*HBB* and *HMBS*) using the absolute quantification method.

(D) M2 RNA expression in patient's II.1 skin fibroblasts is demonstrated by PCR amplification (*PRMT9* as a control) and by Sanger sequencing.

(E) *IFT140* expression in skin fibroblasts was revealed by western blot using anti-*IFT140* antibody.

### Figure 2. Patients' fibroblasts have reduced number of ciliated cells and mislocalized *IFT140*

(A) Number of ciliated cells were counted on serum deprived (24H) control and patient's skin fibroblasts fixed for immunofluorescence and stained with anti-*IFT140* (red) and anti-acetylated tubulin (green) antibodies. Nuclei were colored in blue (Dapi).

1  
2  
3 (B) Based on 15 fields in three independent experiments (100-200 cells per experiment),  
4 mean percentages of ciliated cells are shown in a histogram together with corresponding  
5 standard deviation and p-values (n.s.: non-significant, \*:  $p < 0,0001$ ).  
6  
7

8  
9  
10 (C) IFT140 localisation was assessed on serum deprived (24H) control and patient's skin  
11 fibroblasts fixed for immunofluorescence and stained with anti-IFT140 (red) and anti-  
12 acetylated tubulin (green) antibodies. Nuclei were colored in blue (Dapi).  
13  
14  
15

16  
17 (D) Primary cilia containing IFT140 were counted based on three independent experiments  
18 (60-200 ciliated cells per experiment), mean percentages are shown in a histogram together  
19 with corresponding standard deviation and p-values (n.s.: non-significant, \*:  $p < 0,0005$ ).  
20  
21  
22  
23  
24  
25

### 26 **Figure 3. Mutations in *IFT140***

27  
28 (A) Pedigrees of 7 additional families with affected individuals carrying the tandem  
29 duplication.  
30  
31

32  
33 (B) Pedigrees of additional families with *IFT140* pathogenic variants.  
34  
35  
36  
37  
38

### 39 **Figure 4. Schematic representation of *IFT140* and of the tandem duplication**

40  
41 (A) *IFT140* extends over 105.6 kb in chromosomal region 16p13.3 and contains 31 exons  
42 encoding a 1462 amino acids protein containing two repeated domains, 5 WD repeats and 9  
43 tetratricopeptide (TPR) repeats. Previously reported mutations in *IFT140* are represented  
44 using a black "asterisk" whereas novel mutations are colored in red. In particular, the two  
45 mutations from family A are also shown in introns 20 (c.2577+25G>A) and 30 (c.3454-  
46 490\_4182+2588dup). The tandem duplication (black horizontal bar) spans from exon 27 to 30  
47 (6.7 kb). The breakpoints are located within *Alu* elements (*AluJb* and *AluJr*).  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 (B) Breakpoint junction sequence of affected individuals with tandem duplication. Breakpoint  
4  
5 junction sequence is aligned to the proximal and distal genomic references, respectively *AluJr*  
6  
7 (chr16:1565570-1565697) and *AluJb* (chr16:1571239-1571365), and color-matched.  
8  
9 Microhomology at the breakpoint is indicated in red. The aligned genomic region is 128 bp  
10  
11 long showing 81% of sequence identity.  
12

13  
14 (C) Proposed rearrangement of *Alu-Alu* mediated duplication in affected individuals.  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

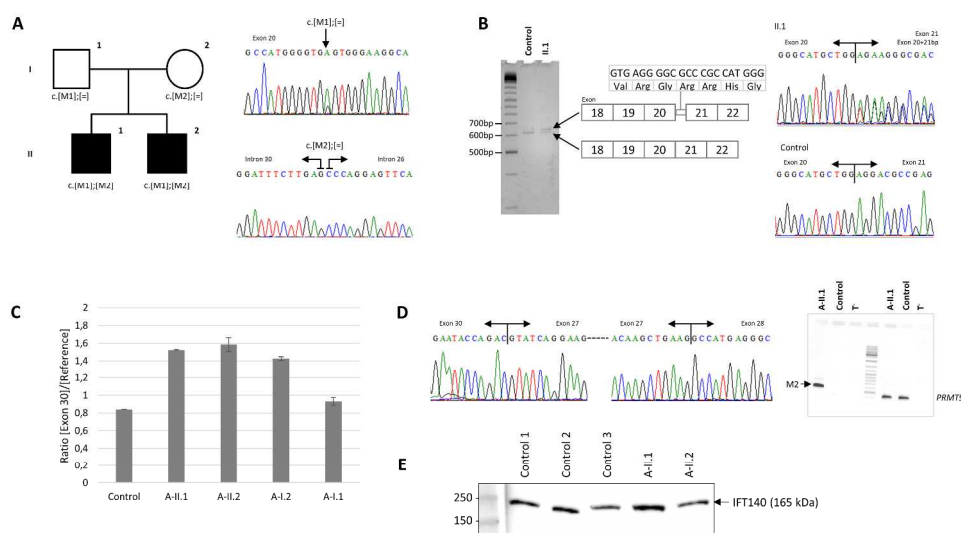


Figure 1. Identification of 2 mutations in IFT140 for a family explored by WGS.

- (A) Pedigree of family A, which has 2 affected individuals. Segregation analysis of the 2 mutations noted M1: 2577+25G>A and M2: 3454-488\_4182+2588dup. Example of Sanger sequencing profiles for individual II.1. The breakpoint junction between exon 30 and intron 26 are shown.
- (B) M1 analysis. Sanger sequencing was performed on RNA extracted from fibroblasts of individual II.1 and a healthy unrelated control amplified between exon 18 and exon 22.
- (C) M2 analysis. Quantitative real-time PCR was performed on DNA from exon 30 in all individuals from family A and one unrelated control. DNA quantity from exon 30 of IFT140 was compared to 2 reference genes (HBB and HMBS) using the absolute quantification method.
- (D) M2 RNA expression in patient's II.1 skin fibroblasts is demonstrated by PCR amplification (PRMT9 as a control) and by Sanger sequencing.
- (E) IFT140 expression in skin fibroblasts was revealed by western blot using anti-IFT140 antibody.

297x166mm (300 x 300 DPI)

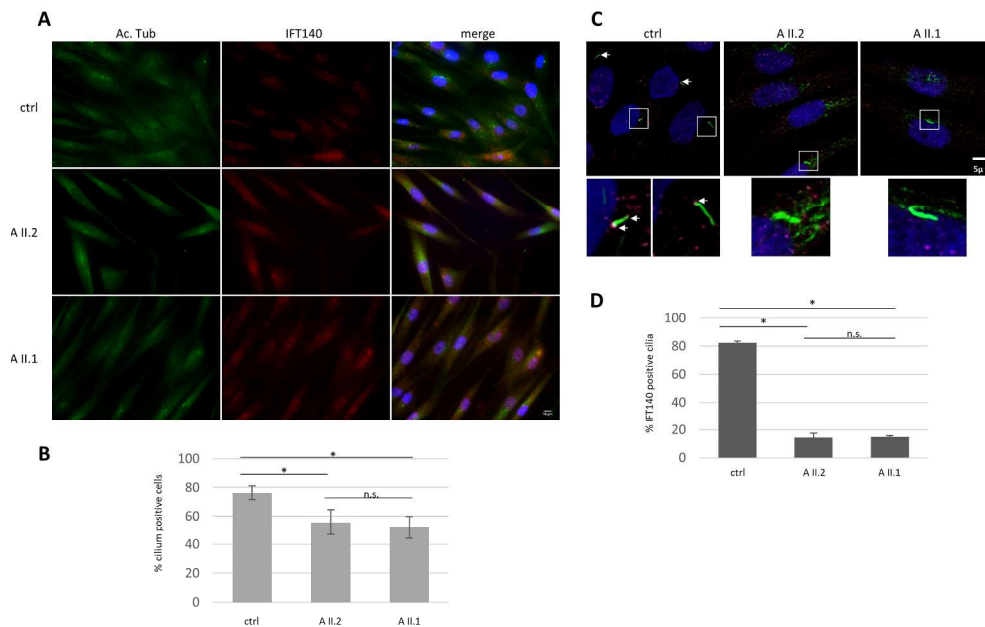


Figure 2. Patients' fibroblasts have reduced number of ciliated cells and mislocalized IFT140

(A) Number of ciliated cells were counted on serum deprived (24H) control and patient's skin fibroblasts fixed for immunofluorescence and stained with anti-IFT140 (red) and anti-acetylated tubulin (green) antibodies. Nuclei were colored in blue (Dapi).

(B) Based on 15 fields in three independent experiments (100-200 cells per experiment), mean percentages of ciliated cells are shown in a histogram together with corresponding standard deviation and p-values (n.s.: non-significant, \*:  $p < 0,0001$ ).

(C) IFT140 localisation was assessed on serum deprived (24H) control and patient's skin fibroblasts fixed for immunofluorescence and stained with anti-IFT140 (red) and anti-acetylated tubulin (green) antibodies. Nuclei were colored in blue (Dapi).

(D) Primary cilia containing IFT140 were counted based on three independent experiments (60-200 ciliated cells per experiment), mean percentages are shown in a histogram together with corresponding standard deviation and p-values (n.s.: non-significant, \*:  $p < 0,0005$ ).

220x141mm (300 x 300 DPI)

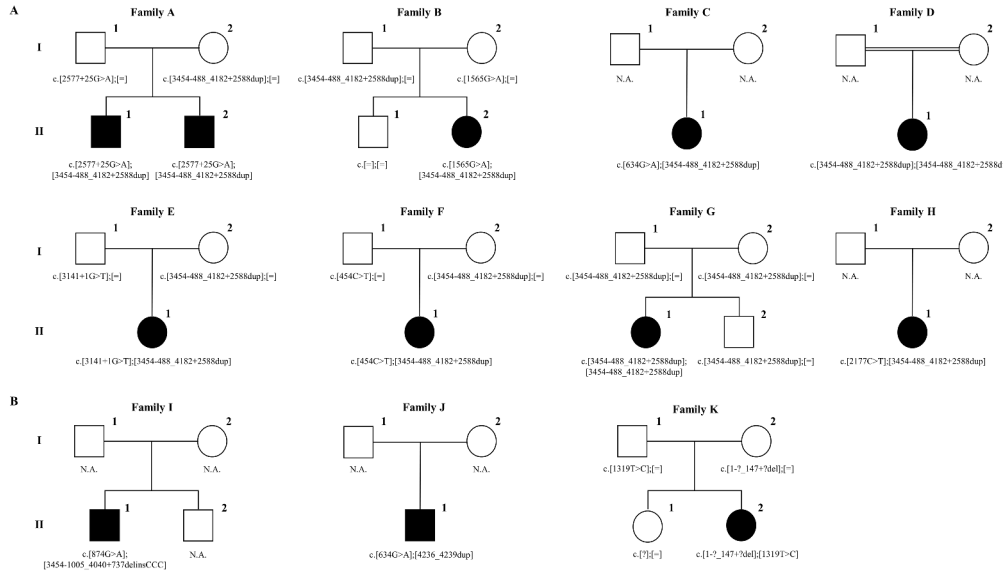


Figure 3. Mutations in IFT140  
 (A) Pedigrees of 7 additional families with affected individuals carrying the tandem duplication.  
 (B) Pedigrees of additional families with IFT140 pathogenic variants.

558x325mm (300 x 300 DPI)

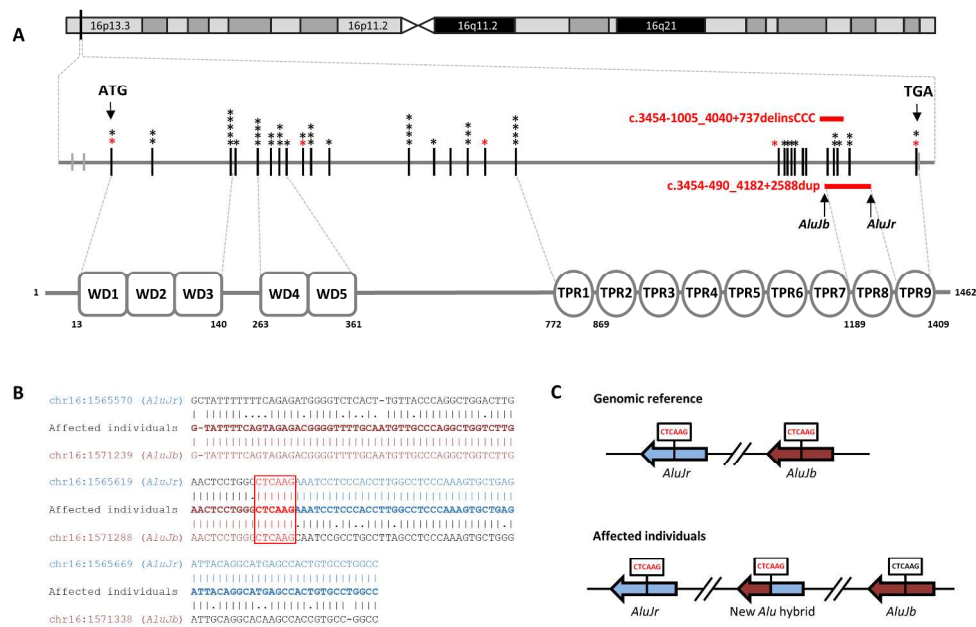


Figure 4. Schematic representation of IFT140 and of the tandem duplication  
 (A) IFT140 extends over 105.6 kb in chromosomal region 16p13.3 and contains 31 exons encoding a 1462 amino acids protein containing two repeated domains, 5 WD repeats and 9 tetratricopeptide (TPR) repeats. Previously reported mutations in IFT140 are represented using a black "asterisk" whereas novel mutations are colored in red. In particular, the two mutations from family A are also shown in introns 20 (c.2577+25G>A) and 30 (c.3454-490\_4182+2588dup). The tandem duplication (black horizontal bar) spans from exon 27 to 30 (6.7 kb). The breakpoints are located within Alu elements (AluJb and AluJr).  
 (B) Breakpoint junction sequence of affected individuals with tandem duplication. Breakpoint junction sequence is aligned to the proximal and distal genomic references, respectively AluJr (chr16:1565570-1565697) and AluJb (chr16:1571239-1571365), and color-matched. Microhomology at the breakpoint is indicated in red. The aligned genomic region is 128 bp long showing 81% of sequence identity.  
 (C) Proposed rearrangement of Alu-Alu mediated duplication in affected individuals.

232x153mm (300 x 300 DPI)

## SUPPLEMENTARY METHODS

### Whole Exome Sequencing

Whole Exome Sequencing (WES) was performed for the two affected siblings from family A (A-II.1 and A-II.2) by the IGBMC (Institute of Genetics and Molecular and Cellular Biology) Microarray and Sequencing platform. Exons of DNA samples were captured with in-solution enrichment methodology (Agilent SureSelect All Exon XT2 50 Mb Kit) and sequenced with an Illumina HiSeq 2500 instrument as paired-end 100 bp reads, in order to reach an average sequencing depth of 80x for each sample. The sequence reads were aligned to the reference sequence of the human genome (GRCh37) using the Burrows-Wheeler Aligner (BWA V7.12) (Li and Durbin, 2010). The HaplotypeCaller module of the Genome Analysis ToolKit (GATK, v.3.4.46) (DePristo, et al., 2011) was used for calling both SNV and indel.

### Targeted Exome Sequencing

#### *Leber panel*

Targeted exome sequencing of 14 full genes related to Leber congenital amaurosis was performed for 126 patients (including family B, C, D, E and I). Illumina compatible precapture barcoded genomic DNA libraries were constructed according to the manufacturer's sample preparation protocol (Ovation Ultralow, Nugen Technologies). Exons of DNA samples were captured with in-solution enrichment methodology (Agilent SureSelect custom panel) and sequenced with an Illumina HiSeq2500 (Paired-End sequencing 2x125 bases, Rapid Mode, 80 libraries per lane). Bioinformatics analysis was performed as described in Gerber *et al* (Gerber, et al., 2016) using an in house pipeline (POLYWEB).



***RP panel***

Targeted exome sequencing of 267 genes related to retinitis pigmentosa was performed for 117 patients (including family J). Exons of DNA samples were captured with in-solution enrichment methodology (Agilent QXT SureSelect custom panel) and sequenced with an Illumina NextSeq 550 instrument (Paired-End sequencing 2x150 bases, 48 libraries per lane). SNVs and indels were called with the Genome Analysis Toolkit v.3.4.46 thanks to our in house pipeline (STARK) and following the GATK best practice.

***Ciliome panel***

Targeted exome sequencing of 1221 ciliary candidate genes including genes related to ciliopathies like Jeune asphyxiating thoracic dystrophy, Mainzer-Saldino syndrome, Bardet-Biedl syndrome, Joubert syndrome and Senior Loken was performed for 364 patients (including families C, D, I and K). Sequencing was conducted using a custom SureSelect capture kit (Agilent Technologies) targeting 4.5 Mb of 20,168 exons, and performed on SOLiD5500XL (Life Technologies) or HiSeq (Illumina). Bioinformatics analysis was performed as described in Grampa *et al* (Grampa, et al., 2016) using an in house pipeline (POLYWEB). The mean depth of coverage obtained was greater than 165x, with  $\geq 89\%$  of the bases covered at least 15x. In order to evaluate duplication and large deletion events, for each individual the relative read count for each targeted region was determined as the ratio of the read count for that region divided by the total absolute read counts of all targeted regions of the design. The ratio of the relative read count of a region in a given individual over the average relative read counts in other individuals of the run resulted in the estimated copy number for that region in that individual.

**Missense pathogenicity analysis**

For the 2 novel missenses identified in this study, conservation has been assessed using the multiple sequence alignment available from the eggnoG database (KOG3717, (Huerta-Cepas, et al., 2016)) and they were analyzed thanks to the PolyPhen-2, SIFT and MutationTaster softwares (Adzhubei, et al., 2010; Kumar, et al., 2009; Schwarz, et al., 2010)

For Peer Review

Disease	Allele 1	Location	Allelic count	Allele 2	Location	Allelic count	Family
<b>Perrault <i>et al</i> 2012 (Perrault, et al., 2012) and this study</b>							
MSS	c.2399+1G>T (p.?)	Intron 19	5/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
MSS	c.932A>G (p.Tyr311Cys)	Exon 9	1/120	c.857_860del (p.Ile286Lysfs*6)	Exon 8	1/120	
MSS	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
MSS	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
MSS	c.634G>A (p.Gly212Arg)	Exon 7	7/120	c.3916dupG (p.Ala1306Glyfs*56)	Exon 29	1/120	
MSS	c.699T>G (p.Ile233Met)	Exon 7	2/120	c.699T>G (p.Ile233Met)	Exon 7	2/120	
JATD	c.2399+1G>T (p.?)	Intron 19	5/120	c.634G>A (p.Gly212Arg)	Exon 7	7/120	
MSS	c.1565G>A (p.Gly522Glu)	Exon 14	4/120	<b>c.3454-488_4182+2588dup p.(Tyr1152_Thr1394dup)</b>	Introns 26-30	10/120	Family B
JATD/MSS	c.874G>A (p.Val292Met)	Exon 8	4/120	<b>c.3454-1005_4040+737delinsCCC, (p.Tyr1152Aspfs*14)</b>	Exons 27-30	1/120	Family I
MSS	c.1727G>A (p.Arg576Gln)	Exon 15	1/120	2 <sup>nd</sup> allele not found			
MSS	c.489C>T (p.Gly163Gly), and c.488_491del (p.Glu164Thrfs*10) from creation of additional donor splice site	Exon 5	1/120	2 <sup>nd</sup> allele not found			
<b>Schmidts <i>et al</i> 2013 (Schmidts, et al., 2013)</b>							
JATD	c.1380delC (p.Asn460Lysfs*28)	Exon 12	1/120	c.874G>A (p.Val292Met)	Exon 8	4/120	
JATD	c.1565G>A (p.Gly522Glu)	Exon 14	4/120	c.874G>A (p.Val292Met)	Exon 8	4/120	
JATD	c.1565G>A (p.Gly522Glu)	Exon 14	4/120	c.874G>A (p.Val292Met)	Exon 8	4/120	
JATD	c.454C>T (p.Leu152Phe)	Exon 5	3/120	c.454C>T (p.Leu152Phe)	Exon 5	3/120	
JATD	c.2278C>T (p.Arg759*)	Exon 19	2/120	2 <sup>nd</sup> allele not found			
MSS	c.2399+1G>T (p.?)	Intron 19	5/120	c.4078T>C (p.Cys1360Arg)	Exon 30	1/120	
MSS	c.418G>A (p.Gly140Arg) and c.800A>G (p.Glu267Gly)	Exons 5/7	1/120	c.490G>T (p.Glu164*)	Exon 6	1/120	
JATD	c.2399+1G>T (p.?)	Intron 19	5/120	2 <sup>nd</sup> allele not found			
JATD	c.4058 G>C (p.Pro1353Arg)	Exon 30	1/120	2 <sup>nd</sup> allele not found			
JATD	c.2360 T>C (p.Asp787Gly)	Exon 19	1/120	2 <sup>nd</sup> allele not found			
JATD	c.1541 A>T (p.Leu514His)	Exon 14	1/120	2 <sup>nd</sup> allele not found			
JATD	c.329G>A (p.Arg110His)	Exon 4	1/120	2 <sup>nd</sup> allele not found			
JATD	c.481C>A (p.Pro161Thr)	Exon 5	2/120	2 <sup>nd</sup> allele not found			

JATD	c.481C>A (p.Pro161 Thr)	Exon 5	2/120	2 <sup>nd</sup> allele not found			
<b>Khan et al 2014 (Khan, et al., 2014)</b>							
MSS	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
MSS	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
<b>Bifari et al 2015 (Bifari, et al., 2016)</b>							
EORD	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
EORD	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
EORD	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
EORD	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
EORD	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
EORD	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
EORD	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
EORD	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
EORD	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
EORD	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	c.1990G>A (p.Glu664Lys)	Exon 17	31/120	
EORD	c.1541_1542delinsAA (p.Leu514Gln)	Exon 14	2/120	c.1541_1542delinsAA (p.Leu514Gln)	Exon 14	2/120	
<b>Xu et al 2015 (Xu, et al., 2015)</b>							
RP	c.4196T>C (p.Leu1399Pro)	Exon 31	31/120	c.1898_1901delATAA (p.Asn633Serfs*10)	Exon 17	1/120	
RP	c.3826G>A (p.Gly1276Arg)	Exon 28	1/120	c.1989C>G (p.Cys663Trp)	Exon 17	1/120	
RP	c.650_651delTG (p.Val217Glyfs*2)	Exon 7	1/120	c.212C>T (p.Pro71Leu)	Exon 4	1/120	
RP	c.2611C>T (p.Arg871Cys)	Exon 21	1/120	c.1377G>A (p.Trp459*)	Exon 12	1/120	
RP	c.1252G>C (p.Ala418Pro)	Exon 11	1/120	c.2921C>T (p.Ala974Val) presumed in trans	Exon 23	1/120	
LCA	c.1451C>T (p.Thr484Met)	Exon 13	3/120	c.985T>C (p.Cys329Arg)	Exon 9	1/120	
LCA	c.2368G>A (p.Glu790Lys)	Exon 19	1/120	c.1655_1656delAG (p.Glu522Glyfs*6)	Exon 14	1/120	
<b>Beheshtian et al 2015 (Beheshtian, et al., 2015)</b>							
RP	c.3827G>A (p.Gly1276Glu)	Exon 28	2/120	c.3827G>A (p.Gly1276Glu)	Exon 28	2/120	
<b>Weisschuh et al 2016 (Weisschuh, et al., 2016)</b>							
RP	c.472C>T (p.Arg158Trp)	Exon 5	1/120	c.1565G>A (p.Gly522Glu)	Exon 14	4/120	
<b>Hull et al 2016 (Hull, et al., 2016)</b>							

RP	c.1451C>T (p.Thr484Met)	Exon 13	3/120	c.1451C>T (p.Thr484Met)	Exon 13	3/120	
RP	c.2399+1G>T (p.?)	Intron 19	5/120	c.2815T>C (p.Ser939Pro)	Exon 22	1/120	
RP	c.998G>A (p.Cys333Tyr)	Exon 9	4/120	c.998G>A (p.Cys333Tyr)	Exon 9	4/120	
RP	c.998G>A (p.Cys333Tyr)	Exon 9	4/120	c.998G>A (p.Cys333Tyr)	Exon 9	4/120	
RP	c.1021G>A (p.Ala341Thr)	Exon 10	1/120	c.1422_23insAA (p.Arg475Asnfs*14)	Exon 12	1/120	
<b>Pena-Padilla et al 2016 (Pena-Padilla, et al., 2016)</b>							
OTCS	c.3141+1G>T (p.?)	Exon 24	2/120	c.-11_6del (p.?)	Exon 3	1/120	
<b>Bayat et al 2017 (Bayat, et al., 2017)</b>							
SB	c.634G>A (p.Gly212Arg)	Exon 7	7/120	c.2278C>T (p.Arg759*)	Exon 19	2/120	
<b>Helm et al 2017 (Helm, et al., 2017)</b>							
MSS	c.634G>A (p.Gly212Arg)	Exon 7	7/120	c.634G>A (p.Gly212Arg)	Exon 7	7/120	
<b>This study</b>							
MSS	<b><i>c.2577+25G&gt;A (p.Leu859_Glu860insValArgGlyAlaArgHisGly)</i></b>	Intron 20	1/120	<b><i>c.3454-488_4182+2588dup (p.Tyr1152_Thr1394dup)</i></b>	Introns 26-30	10/120	Family A
MSS	c.634G>A (p.Gly212Arg)	Exon 7	7/120	<b><i>c.3454-488_4182+2588dup (p.Tyr1152_Thr1394dup)</i></b>	Introns 26-30	10/120	Family C
MSS	<b><i>c.3454-488_4182+2588dup (p.Tyr1152_Thr1394dup)</i></b>	Introns 26-30	10/120	<b><i>c.3454-488_4182+2588dup (p.Tyr1152_Thr1394dup)</i></b>	Introns 26-30	10/120	Family D
JATD/MSS	c.3141+1G>T (p.?)	Exon 24	2/120	<b><i>c.3454-488_4182+2588dup (p.Tyr1152_Thr1394dup)</i></b>	Introns 26-30	10/120	Family E
MSS	c.454C>T (p.Leu152Phe)	Exon 5	3/120	<b><i>c.3454-488_4182+2588dup (p.Tyr1152_Thr1394dup)</i></b>	Introns 26-30	10/120	Family F
MSS	<b><i>c.3454-488_4182+2588dup (p.Tyr1152_Thr1394dup)</i></b>	Introns 26-30	10/120	<b><i>c.3454-488_4182+2588dup (p.Tyr1152_Thr1394dup)</i></b>	Introns 26-30	10/120	Family G
MSS/SB	<b><i>c.2177C&gt;T (p.Pro726Leu)</i></b>	Exon 18	1/120	<b><i>c.3454-488_4182+2588dup (p.Tyr1152_Thr1394dup)</i></b>	Introns 26-30	10/120	Family H
RP	c.634G>A (p.Gly212Arg)	Exon 7	7/120	<b><i>c.4236_4239dup (p.Tyr1414Leufs*48)</i></b>	Exon 31	1/120	Family J
MSS (mild)	<b><i>c.1-?_147+?del, start loss</i></b>	Exon 3	1/120	<b><i>c.1319T&gt;C (p.Leu440Pro)</i></b>	Exon 11	1/120	Family K

**Table S1.** Previously reported and novel mutations in *IFT140*.

Mutations are reported only once per family. Those reported in this study are in bold and those described for the first time are also in italic. Mutation allelic counts are reported for only one affected individual per family. It is to notice that some studies (i.e. Bifari *et al*) have analyzed consanguineous families from the same region increasing the global allelic count. EORD: early-onset retinal dystrophy; JATD: Jeune asphyxiating thoracic dystrophy; MSS: Mainzer-Saldino Syndrome; OCTS: Opitz trigonocephaly C syndrome; RP: retinitis pigmentosa; SB: Sensenbrenner syndrome.

1  
2  
3  
4

Individual	A-II.1	A-II.2	B-II.1	C-II.2	D-II.1	E	F	G	H	I-II.1	J-II.1	K-II.2
Birth	10/01/1998 France, Italy	31/08/1999 France, Italy	17/03/1978 Serbia, Yugoslavia	26/08/1976 France	27/12/1968 France	20/09/2010 Caucasian	11/02/2007 Austria	21/05/1993 Austria		16/04/1986	05/06/1993	14/02/2008
Sex	M	M	F	F	F	F	F	F	F	M	M	F
Suspected diagnostic	MSS	MSS	MSS	MSS	MSS	Jeune/MSS	MSS	MSS	MSS/Sensenbrenner	Jeune/MSS	Isolated RP	MSS (mild)
Onset	14 yo	15 yo	3 yo	6 yo	45 yo	Neonatal	4 yo	1 <sup>st</sup> yo	Neonatal			7 yo
Short stature (SD)	+ (-3 SD)	+ (-2 SD)	+ (<-5 SD)	+ (-3 SD)	+ (<-5 SD)	+ disproportionate (-3 SD)	(-1 SD)	(+1.5 SD)				(-1 SD)
Skeletal features	Broad long bones	Hip anomalies (narrow pelvis with narrow iliac wings, lack of coverage of superior femoral epiphyses)	Hip anomalies Metaphyseal defect (forearms) Micromelia Scoliosis	Metaphyseal defect (long bones)	Short and broad bones Bilateral ulnar valgus Marked curved radius Hip anomalies (large pelvis, short femoral heads) Varus of the feet	Hypoplastic thorax Short, slightly broader ribs Hip anomalies (hip dysplasia, broad epiphyses, steep acetabular roofs) Genu varum Hyperlordosis Brachymelia No premature cranosynostosis	Moderate hips anomalies	No	Craniosynostosis Hip anomalies not known	Thoracic dystrophy		
Digit anomalies	Brachymetacarpia (1, 3, 4 digits) Brachymetatarsia	Short hands Brachydactyly and brachymetacarpia (1, 4 and 5 digits) brachymetatarsia	Short hands Cone-shaped phalangeal epiphyses (hands)	Cone-shaped phalangeal epiphyses (hands)	Brachytelephalangy Brachymesophalangy with clinodactyly of the 5 <sup>th</sup> digits (feet and hands) Enlarged metaphyses (feet) No cone-shaped epiphyses due to the age of the patient (52 yo)	Short hands Brachyphalangy Cone-shaped phalangeal epiphyses (hands) Pes planovalgus Incomplete syndactyly of digits 3-4 (toes) Brachydactyly of 3rd toe Delayed bone age	Short hands Brachymesophalangy Cone-shaped phalangeal epiphyses (feet, hands)	Short hands Brachymesophalangy Cone-shaped phalangeal epiphyses (feet, hands)	Short hands Brachymesophalangy Cone-shaped phalangeal epiphyses	Cone-shaped phalangeal epiphyses		Cone-shaped phalangeal epiphyses (second phalanges of the 2, 3, 4 digits)
Retinal dystrophy	Hemeralopia (10 yo) VA: 9/10RLE	Hemeralopia (7 yo) Divergent strabismus VA: 7/10RLE RP Altered ERG	Onset (3 mo) Nystagmus Night blindness VA: 2/10RLE Maculopathy Altered ERG	Night blindness VA: 4/10RLE	Divergent strabismus (2 yo) Hemeralopia (10 yo) RP (16 yo) VA: 2.4/10 RE 1/20 LE Advanced RP VF: very reduced on right side	Congenital pendular nystagmus Strabismus Hyperopia Astigmatism Severe visual reduction	Myopia RP (4 yo)	Myopia RP (1 yo)	Nystagmus Strabismus Refraction +2 Retinal dystrophy not known Papillae edges not well differentiated	ERG: no signal left	RP slow evolution	No eye pursuit (1 mo) Bilateral divergent strabismus (5 mo) Nystagmus (6 mo) Early onset advanced RP (3 yo) ERG: severe and primary damage of cones and rods
Renal involvement	Moderate renal failure Microalbuminuria Renal US normal	Moderate renal failure Microalbuminuria Renal US normal Arterial hypertension	End stage renal failure (5 yo) Dialysis Transplant Nephronophthisis with polyuria polydipsia (6 yo) Arterial hypertension (24/12) Renal US (Several cortical cysts (10 yo), echogenicity, kidney's size: R:5mm L:59mm)	End stage renal failure (7 yo) Dialysis Transplant (7 yo) Nephronophthisis with polyuria polydipsia (6 yo) Arterial hypertension (24/12) Renal US (no cysts, echogenicity, kidney's size: R:5mm L:59mm) Obsolescent/hyalinized glomerulus tubulo-interstitial nephritis important arterial damage	End stage renal failure (10 yo) Dialysis (10 yo) Transplantation (16 yo)	End stage renal failure (1.5 yo) Peritoneal dialysis (2 yo) Transplant (2.5 yo) Hydronephrosis (4 mo) Renal US (small hyperechoic kidney, right kidney with 3 chambered cysts) Arterial hypertension Renal anemia	End stage renal failure (4.7 yo) Renal US (Small kidneys (<3th centile, increased echogenicity, reduced cortico-medullary differentiation, no cysts) Proteinuria, microhematuria	End stage renal failure (23 yo) Renal US (cysts, low-normal size, increased echogenicity, loss of cortico-medullary differentiation (11 yo) Glomerular and tubular proteinuria	End stage renal failure (Dialysis (3.8 yo) Transplant CKD stage V Renal US (small cysts, small kidney) Glomerular and tubular proteinuria No hematuria			Renal function normal Renal ultrasound normal
Hepatic involvement	No	No	Portal space fibrosis Hemosiderosis		No	Slightly hyperechoic liver Slightly enlarged	Yes	No	Normal liver function tests Hepatomegaly			
Pancreatic fibrosis	No	No			No	No				Yes (cystic)		
Neurological features	No	No	No			Asphyxia at birth Slight Developmental delay (head control 4 mo, sitting 8 mo, walking 24 mo, first words 18 mo, slow speech development) No ataxia, no seizures No autistic features	No	Hypotonia at birth Developmental delay	Developmental delay No cerebellar ataxia, but oculomotor apraxia Intraocular hypertension No seizures			
Cerebral anomalies	No	No				Generalized brain atrophy (CT scan 4 mo)	No		No		No	No
Dysmorphic features	Yes (high hairline, frontal bossing, ptosis, long philtrum, prominent chin)	Yes (high forehead, frontal bossing, long philtrum, prominent chin)			No	Yes (high hairline, frontal bossing, hypertelorism, broad nasal ridge, deep-set eyes, anteverted nares, long philtrum, everted lower lip, prominent chin, deep-set ears)	No	No	Yes			
Oral anomalies	No	No	No		No	No	No	No	No			
Other	Asthma					Truncal adiposity, breast tissue Splenomegaly Obstructive bronchitis Recurrent infections	Recurrent respiratory infections	Late onset obesity	Premature birth 28 weeks of gestational age		Overweight when younger (9 yo)	Joint luxations with spontaneous subluxations

**Table S2** Clinical Features of MSS-affected individuals harboring compound-heterozygous mutations in *IFT140*. CKD: Cystic Kidney Disease, CT: Computed Tomography, ERG: Electroretinogram, F: Female, L: Left, LE: Left Eye, M: Male, mo: month old, MSS: Mainzer-Saldino Syndrome, R: Right, RE: Right Eye, RLE: Right and left Eye, RP: Retinitis Pigmentosa, SD: standard deviation, US: Ultrasound, VA: Visual acuity, VF: visual field, yo: years old.

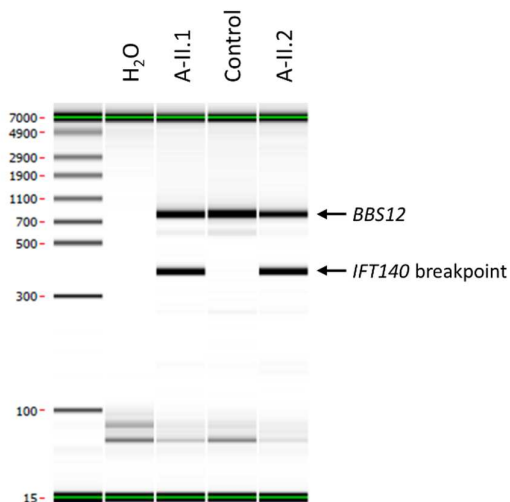
Application	Gene	RefSeq	Name	Forward (5'-3')	Reverse (5'-3')
DNA c.2577+25G>A	<i>IFT140</i>	NM_014714.3	IFT140-ex20	GATGACTACTTGCCCCAGGA	CACACCTGACAGGCACACAC
DNA breakpoint TD			IFT140-intron30F_intron26R	CCGGAGTAGCTGGGATTACA	CTGTCTAGCTGGGGTGGGTA
DNA c.1565G>A			IFT140-ex14	TGAGCGAGTGAATGAGTGGA	CGGTGTTAAAATGGGCACCT
cDNA c.2577+25G>A			IFT140-RT-PCR-ex18-19F_ex22R	ATCAAAAAGTGAGGCCGTCTG	GCGTGTCCGACTTCTCGTA
cDNA TD			IFT140-RT-PCR-ex30F_ex28R	AAGCAGTGTGAGCTGCTCCT	GGGCACAAGCGTCATAAAAG
cDNA control	<i>PRMT9</i>	NM_001304458	PRMT9-RT-PCR-ex3F_ex6R	GTGGAACGCTGGCACTTTAT	CAATTAGCACTTTCACTTTGG
qPCR	<i>IFT140</i>	NM_014714.3	IFT140-QPCR-ex30F_ex30R	CATAGAAAGTTTCCTCCTGCCAT	TCTGGTTCCTCCAGGAGCA
qPCR	<i>HMBS</i>	NM_000190	HMBS-QPCR-	CACAGCACTCCCCTGACAAC	GAGGGAGGCGATAGTAGGACC
qPCR	<i>HBB</i>	NM_000518	BGLOB-QPCR-ex1F_ex1R	ACACAACGTGTTCACTAGC	CAACTTCATCCACGTTACCC
Duplex PCR	<i>IFT140</i>	NM_014714.3	IFT140-intron30F_intron26R	CCCAAAGCGTCATTTCTGGTGTTT	CCCAGGGTCAGGTGATCTTCTTA
	<i>BBS12</i>	NM_152618	BBS12-ex2F_ex2R	AGTTCTCATTGAGGGTGACCT	GCCAGAGATGAAGCCAGCCA

**Table S3** List and characteristics of primers used in this study. TD: Tandem duplication of exon 27-30

<b>Name primary antibodies</b>	<b>Applications</b>	<b>Producer</b>	<b>Conditions</b>
Mouse anti-Acetylated $\alpha$ -Tubulin	WB, IF	Abcam #ab24610	IF: 1/200
Rabbit anti-IFT140	WB, IF	Proteintech #17460-1-AP	IF: 1/100 WB: 1/1000
Mouse anti- $\beta$ -Tubulin	WB	Euromedex # TUB-2A2	1/2500
<b>Name secondary antibodies</b>			
Donkey anti-Mouse IgG (H+L), FITC conjugate	IF	Thermo Scientific #A16012	1/500
Goat anti-Rabbit IgG (H+L) Alexa Fluor 568 conjugate	IF	Thermo Scientific #A-11011	1/500
Chicken anti-Rabbit IgG-HRP	WB	Santa Cruz #sc-2955	1/5000
Goat anti mouse IgG-HRP	WB	Santa Cruz #sc-2060	1/5000

**Table S4.** List of antibodies used in this study. IF: immunofluorescence, WB Western Blot.



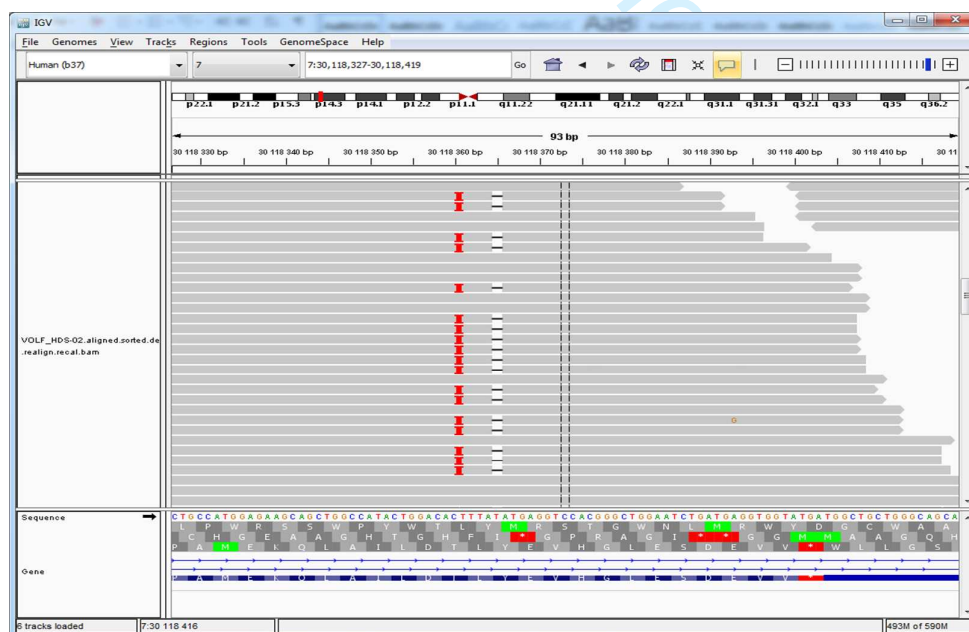


**Figure S1.** Duplex PCR profiles on a LabChip GX (Perkin Elmer) for a rapid detection of the *IFT140* tandem duplication (see **Table S3** for primers details). Normal samples harbor a single band (*BBS12*) whereas homozygous or heterozygous carriers for the tandem duplication show an additional band specific for the duplication breakpoint.

Patient	A-II.1		A-II.2	
	SNV	indel	SNV	indel
<b>Type of sequence variant</b>				
Total number of variants	49489	2168	49418	2133
After exclusion of variants with an allele frequency >1% (extracted from dbSNP, the Exome Variant Server and the ExAC database or our internal exome database)	6528	1038	6517	1017
After exclusion of variants found at the homozygous state and more than once at the heterozygous state in 90 control exomes	1159	90	1029	72
After exclusion of 5'UTR, 3'UTR, downstream, upstream and intron locations without local splice effect prediction (from the "localSpliceEffect" field of Alamut-Batch)	517	20	458	12
After exclusion of synonymous variants without local splice effect prediction (from the "localSpliceEffect" field of Alamut-Batch)	417	20	373	12
After selection of variants consistent with recessive transmission (compound heterozygous, homozygous variants).	2 compound heterozygous in <i>PLEKHA8</i>			

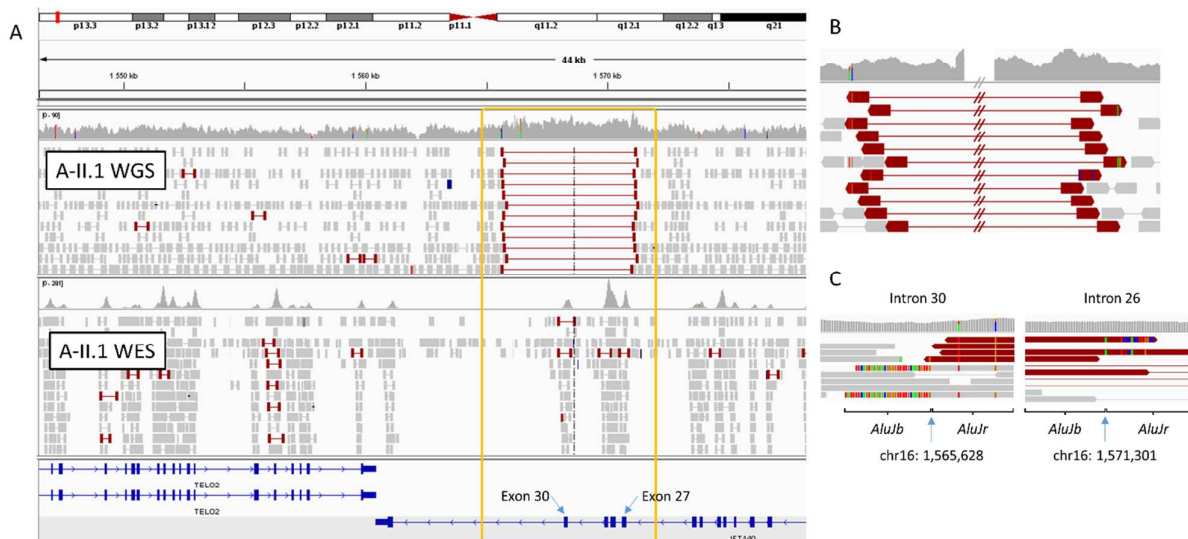
**Table S5** Summary of the whole exome sequencing results from family A

WES data processing and variant calling revealed respectively 51,657 and 51,551 SNV and indels in individual A-II.1 and A-II.2. Variant filtering using stringent criteria reduced the number of genetics variants to respectively 437 and 385 variants per proband. To identify variants consistent with autosomal recessive inheritance, we kept only compound shared heterozygous or homozygous variants, reducing the number of variants to 2 heterozygous variants in the *PLEKHA8* gene: NM\_001197026.1:c.1517\_1518insAT (p.Leu507Phefs\*27) and NM\_001197026.1:c.1522del (p.Tyr508Metfs\*25). Both variants being always supported on the same reads (as shown in the IGV view below), they are in *cis* and are probably an artifact due to a bad reads alignment. So, no candidate gene was found by the WES analyses.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

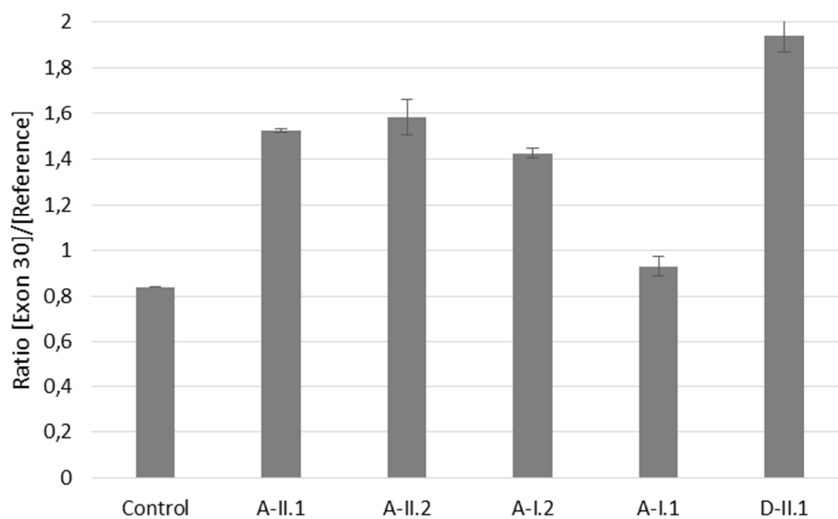


**Figure S2.** Next generation sequencing data from patient A-II.1 displayed around *IFT140*, in particular from exon 27 to exon 30 highlighting the tandem duplication.

(A) Comparison between whole genome sequencing (WGS) and whole exome sequencing (WES) data from patient A-II.1. The reads are displayed as pairs and sorted according to their insert size using IGV (Thorvaldsdottir, et al., 2013).

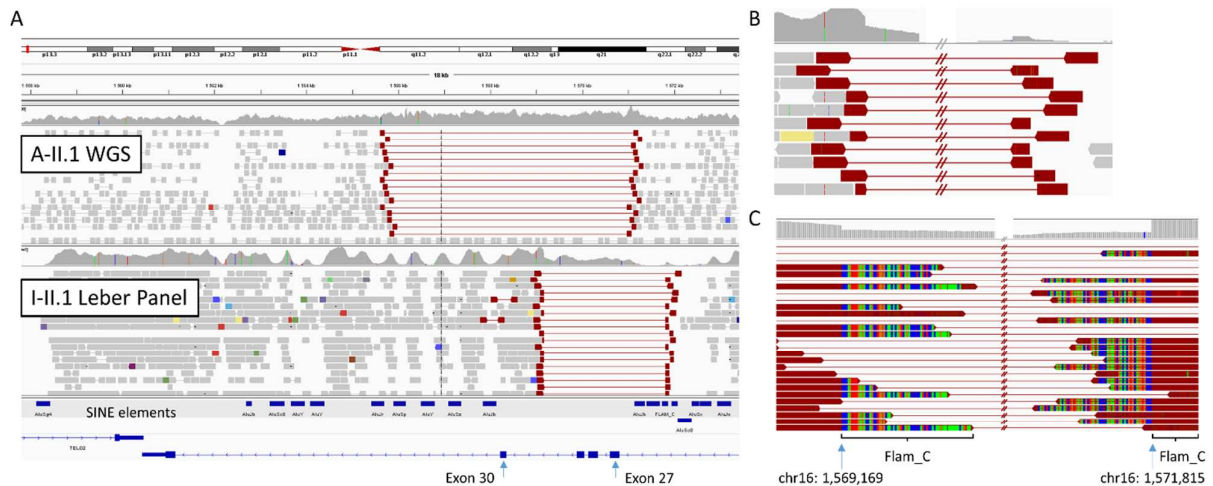
(B) Highlight of the WGS read pairs at the breakpoint junctions showing the typical drawing (read pairs pointing in opposite direction) for tandem duplications in direct orientation.

(C) Highlight of the WGS read pairs at the breakpoints including soft clipped bases. The left breakpoint in intron 30 overlaps the *AluJr* while the right breakpoint in intron 26 overlaps the *AluJb*. Analysis of the reads and the split reads that are misaligned allowed us to define the exact position of the duplication.



**Figure S3.** Homozygous duplication in *IFT140* validated by quantitative real-time PCR.

Quantitative real-time PCR was performed on DNA from exon 30 in all individuals from family A and D as well as in one unrelated control. DNA quantity from exon 30 of *IFT140* was compared to 2 reference genes (*HBB* and *HMBS*) using the absolute quantification method.



**Figure S4.** Next generation sequencing data for family A and I displayed around *IFT140*, in particular from exon 27 to exon 30.

(A) Comparison between whole genome sequencing (WGS) from patient A-II.1 (tandem duplication carrier) and targeted exome sequencing using the full *IFT140* gene as a target region from patient I-II.1 (deletion of exon 27 to 29). The reads are displayed as pairs and sorted according to their insert size using IGV (Thorvaldsdottir, et al., 2013). The two events do not share the same breakpoints.

(B) Highlight of the read pairs from patient I-II.1 at the breakpoint junctions showing the typical drawing (read pairs pointing at each other) for a genomic deletion.

(C) Highlight of the read pairs from patient I-II.1 at the breakpoints including soft clipped bases. The right breakpoint overlaps a SINE element from the *Alu* family (Flam\_C) but not the left breakpoint. Analysis of the split reads that are misaligned allows us to define the exact position of the deletion (c.3454-1005\_4040+737delinsCCC).









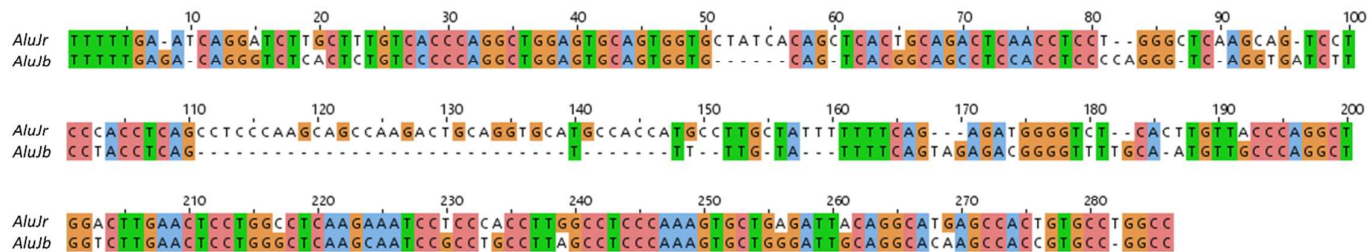


1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

	rs138669173	1654525	C	G	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	rs7197288	1654753	C	A	C/A	C/C	C/C	C/C	C/A	C/C	C/A	C/A	A/A	C/C	A/A	C/A	C/A	A/A	C/A
	rs2667679	1654946	G	A	G/A	G/G	G/G	G/G	G/A	G/G	G/A	G/G	A/A	G/G	G/A	G/A	G/A	G/A	G/G
	rs56359342	1654948	G	A	G/A	G/G	G/G	G/G	G/A	G/G	G/A	G/G	A/A	G/G	G/A	G/A	G/A	G/A	G/G
	rs11745556	1655073	T	C	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	rs35220693	1655439	C	T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/T	C/C	C/T	C/C	C/C
	rs6600157	1656321	G	A	G/A	C/G	C/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	rs142169797	1656564	C	T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	rs2859310	1656765	A	G	A/G	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/G	G/G	A/A	G/G	A/G	A/G	A/A
	rs542474562	1656927	GCACA	G	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA
	rs7192480	1657012	C	T	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/T	T/T	C/C	C/T	C/T	C/T	C/T	C/C
	rs34935994	1657034	CAT	C	CAT/C	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT	CAT/CAT
	rs1894649	1657284	C	G	C/G	C/C	C/C	C/C	C/C	C/C	C/G	C/G	G/G	C/C	C/G	C/G	C/G	C/G	C/C
	rs743963	1657350	A	T	A/T	A/A	A/A	A/A	A/T	A/A	A/T	A/A	T/T	A/A	A/T	A/T	A/T	A/T	A/A
	rs3784840	1657754	C	T	C/C	C/C	C/C	C/C	C/T	C/C	C/T	C/C	T/T	C/C	C/T	C/T	C/T	C/T	C/C
	rs743964	1657853	C	T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/T	C/C	C/T	C/C	C/C
	rs763152	1658056	T	C	T/C	T/T	T/T	T/T	T/C	T/T	T/C	T/C	C/C	T/T	C/C	T/C	T/C	T/C	T/T
	rs763153	1658086	G	C	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/C	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	rs35888283	1658101	TAG	T	TAG/T	TAG/TAG	TAG/TAG	TAG/TAG	TAG/TAG	TAG/TAG	TAG/TAG	TAG/TAG	T/T	TAG/TAG	TAG/TAG	TAG/TAG	TAG/TAG	TAG/TAG	TAG/TAG
	rs9937922	1658154	C	G	C/G	C/C	C/C	C/C	C/C	C/C	C/G	C/G	G/G	C/C	G/G	C/G	G/G	C/G	C/C
	rs115729443	1658454	C	T	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C
	rs11438070	1658632	C	CA	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	rs28519681	1659128	T	C	T/C	T/T	T/T	T/T	T/C	T/T	T/C	T/C	C/C	T/T	C/C	T/C	T/C	T/C	T/T
	rs28633318	1659723	T	G	T/G	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	rs1057983	1660274	A	G	A/G	A/A	A/A	A/A	A/G	A/A	A/G	A/G	G/G	A/A	G/G	A/G	G/G	A/G	A/A
	rs150781872	1661491	C	CA	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	G/C	C/C	C/C	C/C	C/C	C/C
	rs2859314	1661701	A	G	A/G	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/G	A/A	G/G	A/A	A/G	A/G	A/A
	rs553272330	1661814	CA	C	CA/CA	CA/CA	CA/CA	CA/CA	CA/CA	CA/CA	CA/CA	CA/CA	CA/C	CA/CA	CA/CA	CA/CA	CA/CA	CA/CA	CA/CA

**Table S6.** Haplotype analysis for ~100kb around *IFT140* on chromosome 16.

SNP analysis at the *IFT140* locus for carrier individuals of the tandem duplication and in controls sequenced either by WGS or Target Exome Sequencing (TES, by the Leber panel which include the full *IFT140* gene). In total, 6 carrier individuals from 3 families were analysed (family A, B and C) and indicated. Affected patients are depicted with an “\*”. Six control samples are also shown. For each individual, SNP detected in *IFT140* were extracted and listed. Homozygous SNP are shown in green (reference allele) or in blue (alternative allele) and heterozygous SNP in orange. The duplication region is highlighted in grey.



**Figure S5** *AluJb* and *AluJr* alignment

*AluJb* and *AluJr* sequences were extracted from IGV (Thorvaldsdottir, et al., 2013) at the breakpoints of the tandem duplications. Pairwise alignment of the nucleotide sequences and percent identity has been computed using Jalview (Waterhouse, et al., 2009).

Technique	Individuals	CANOES	Pindel	CNVnator
Ciliome panel	D-II.1	DUP, intron26-intron30	67 DEL, 13 INV, 35 SI	NA
Ciliome panel	K-II.2	2 DEL, 3DUP (DUP, intron27-exon31)	36 DEL, 1 INV, 6 SI, 1 TD	NA
Leber panel	B-I.1	DUP, intron26-intron30	DUP, intron26-intron30 (675 DEL, 70 INV, 233 SI, 5 TD)	NA
Leber panel	B-I.2	0	(688 DEL, 62 INV, 211 SI, 2 TD)	NA
Leber panel	B-II.1	DUP, intron26-intron30	DUP, intron26-intron30 (595 DEL, 55 INV, 193 SI, 2 TD)	NA
Leber panel	C-II.1	DUP, intron26-intron30	DUP, intron26-intron30 (842 DEL, 120 INV, 417 SI, 5 TD)	NA
Leber panel	I-II.1	1 DEL	(659 DEL, 115 INV, 339 SI, 2 TD)	NA
WES	A-II.1	0	10 DEL, 3 INV, 10 SI, 1 TD	NA
WES	A-II.2	DUP, intron27-intron30	13 DEL, 3 INV, 9 SI	NA
WGS	A-II.1	DUP, exon26-intron31	14 DEL, 3 INV, 4 SI	0
WGS	A-II.2	DUP, exon26-intron31	10 DEL, 2 INV, 3 SI	0
WGS	A-I.2	DUP, exon26-intron31	17 DEL, 1 INV, 2 SI	0
WGS	A-I.1	0	53 DEL, 1 INV, 1 SI	0

**Table S7.** Summary of the CNV detected in *IFT140* by using 3 different tools in 13 individuals.

CNV detection has been performed on 13 individuals sequenced with 4 different approaches: the Ciliome panel, the Leber panel, the WES and the WGS. Three tools have been run with their default options: CANOES (v1.0), Pindel (v0.2.5b9) and CNVnator (v0.3.3) (Abyzov, et al., 2011; Backenroth, et al., 2014; Ye, et al., 2009). The number of CNV detected in *IFT140* are reported here, as well as the presence or not of the tandem duplication of exons 26 to 30. It is to notice that Pindel simultaneously calls small indel and structural variations, increasing thus the number of detected variants. NA, not applicable; DUP, duplication; DEL, deletion; INV, inversion; SI, short insertion; TD, tandem duplication.

**SUPPLEMENTARY REFERENCES**

- Abyzov A, Urban AE, Snyder M, Gerstein M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* 21(6):974-84.
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nature Methods* 7(4):248-9.
- Backenroth D, Homsy J, Murillo LR, Glessner J, Lin E, Brueckner M, Lifton R, Goldmuntz E, Chung WK, Shen Y. 2014. CANOES: detecting rare copy number variants from whole exome sequencing data. *Nucleic Acids Res* 42(12):e97.
- Bayat A, Kerr B, Douzgou S, Study DDD. 2017. The evolving craniofacial phenotype of a patient with Sensenbrenner syndrome caused by IFT140 compound heterozygous mutations. *Clin Dysmorphol* 26(4):247-251.
- Beheshtian M, Saeed Rad S, Babanejad M, Mohseni M, Hashemi H, Eshghabadi A, Hajizadeh F, Akbari MR, Kahrizi K, Riazi Esfahani M and others. 2015. Impact of whole exome sequencing among Iranian patients with autosomal recessive retinitis pigmentosa. *Arch Iran Med* 18(11):776-85.
- Bifari IN, Elkhamary SM, Bolz HJ, Khan AO. 2016. The ophthalmic phenotype of IFT140-related ciliopathy ranges from isolated to syndromic congenital retinal dystrophy. *Br J Ophthalmol* 100(6):829-33.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M and others. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491-8.
- Gerber S, Alzayady KJ, Burglen L, Bremond-Gignac D, Marchesin V, Roche O, Rio M, Funalot B, Calmon R, Durr A and others. 2016. Recessive and Dominant De Novo ITPR1 Mutations Cause Gillespie Syndrome. *Am J Hum Genet* 98(5):971-980.
- Grampa V, Delous M, Zaidan M, Ody G, Thomas S, Elkhartoufi N, Filhol E, Niel O, Silbermann F, Lebreton C and others. 2016. Novel NEK8 Mutations Cause Severe Syndromic Renal Cystic Dysplasia through YAP Dysregulation. *PLoS Genet* 12(3):e1005894.
- Helm BM, Willer JR, Sadeghpour A, Golzio C, Crouch E, Vergano SS, Katsanis N, Davis EE. 2017. Partial uniparental isodisomy of chromosome 16 unmasks a deleterious biallelic mutation in IFT140 that causes Mainzer-Saldino syndrome. *Hum Genomics* 11(1):16.
- Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M and others. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44(D1):D286-93.
- Hull S, Owen N, Islam F, Tracey-White D, Plagnol V, Holder GE, Michaelides M, Carss K, Raymond FL, Rozet JM and others. 2016. Nonsyndromic Retinal Dystrophy due to Bi-Allelic Mutations in the Ciliary Transport Gene IFT140. *Invest Ophthalmol Vis Sci* 57(3):1053-62.
- Khan AO, Bolz HJ, Bergmann C. 2014. Early-onset severe retinal dystrophy as the initial presentation of IFT140-related skeletal ciliopathy. *J AAPOS* 18(2):203-5.
- Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* 4(7):1073-81.
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589-95.
- Pena-Padilla C, Marshall CR, Walker S, Scherer SW, Tavares-Macias G, Razo-Jimenez G, Bobadilla-Morales L, Acosta-Fernandez E, Corona-Rivera A, Mendoza-Londono R and others. 2016. Compound heterozygous mutations in the IFT140 gene cause Opitz trigonocephaly C syndrome in a patient with typical features of a ciliopathy. *Clin Genet*.

- 1  
2  
3 Perrault I, Saunier S, Hanein S, Filhol E, Bizet AA, Collins F, Salih MA, Gerber S, Delphin  
4 N, Bigot K and others. 2012. Mainzer-Saldino syndrome is a ciliopathy caused by IFT140  
5 mutations. *Am J Hum Genet* 90(5):864-70.
- 6 Schmidts M, Frank V, Eisenberger T, Al Turki S, Bizet AA, Antony D, Rix S, Decker C,  
7 Bachmann N, Bald M and others. 2013. Combined NGS approaches identify mutations in the  
8 intraflagellar transport gene IFT140 in skeletal ciliopathies with early progressive kidney  
9 Disease. *Hum Mutat* 34(5):714-24.
- 10 Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. 2010. MutationTaster evaluates  
11 disease-causing potential of sequence alterations. *Nat Methods* 7(8):575-6.
- 12 Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV):  
13 high-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178-92.
- 14 Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2--a  
15 multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189-91.
- 16 Weisschuh N, Mayer AK, Strom TM, Kohl S, Glockle N, Schubach M, Andreasson S, Bernd  
17 A, Birch DG, Hamel CP and others. 2016. Mutation Detection in Patients with Retinal  
18 Dystrophies Using Targeted Next Generation Sequencing. *PLoS One* 11(1):e0145951.
- 19 Xu M, Yang L, Wang F, Li H, Wang X, Wang W, Ge Z, Wang K, Zhao L, Li H and others.  
20 2015. Mutations in human IFT140 cause non-syndromic retinal degeneration. *Hum Genet*  
21 134(10):1069-78.
- 22  
23 Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to  
24 detect break points of large deletions and medium sized insertions from paired-end short  
25 reads. *Bioinformatics* 25(21):2865-71.
- 26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60