The American Journal of Human Genetics, Volume *109*

Supplemental information

Mutations in *SCNM1* cause orofaciodigital syndrome due to minor intron splicing defects affecting primary cilia

Asier Iturrate, Ana Rivera-Barahona, Carmen-Lisset Flores, Ghada A. Otaify, Rasha Elhossini, Marina L. Perez-Sanz, Julián Nevado, Jair Tenorio-Castano, Juan Carlos Triviño, Francesc R. Garcia-Gonzalo, Francesca Piceci-Sparascio, Alessandro De Luca, Leopoldo Martínez, Tugba Kalaycı, Pablo Lapunzina, Umut Altunoglu, Mona Aglan, Ebtesam Abdalla, and Victor L. Ruiz-Perez

TABLE OF CONTENTS

I. Supplemental Figures

Figure S1

A

B

AGGGTGGGACCTATGGCAGGGACGGGCCCAACCCTCTTCCCCTGGACCACTTATCCTCTTCTCTACCCCTGCTCAGCTTTG
CTTGTGCCATCTGCCCCCATCGACCGGTACTGGACACCCTGGCCATGCTGACTGCCCACCGTGCAGGCAAGAAACATCTGT
CCAGTAAGTTAGGGGGAAGACGGGATGGGGAATAAACCCTCGAAATCTCTGCACACCACTCTTGGTGCTATGCTTTTAATT
CTGTTTCCCTTTCTCCT***CAG***GCTTGCAGCTTTTCTATGGCAAGAAGCAGCCGGGAAAGGAAAGAAAGCAGAATCCAAAACA
TCAGAATGAATTGAGAAGGGAAGAAACCAGGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGG
CGGGCGGATCACGAGGTCAGGAGATCGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAT
TAGCCGGGCGTGGTAGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGGGAATGGCGTGAACCCGGGAGGCG
GAGCTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAA
AAAAAAAA**GAGAAGGGAAGAAACCA**AAGCTGAG***GTAATC***AGAGAGTGGAGAGTGTGTTCTAGGCAAGACCCTGCTGCACAC
ACCAGGCTGGAAGGGCGGAGGGCTGAAGGCTCTGACTCTGTCTCCCATCTCCTTACCAGGCTCCTCTGCTAACTCAGACAC
GACTTATCACCCAGAGTGCTCTGCACAGAGCTCCCCACTATAACAGTTGCTGCCGCCG

C

```
P4       1 GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACGAGGTC 70
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AluYc1   1 GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACGAGGTC 70

P4      71 AGGAGATCGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGG 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AluYc1  71 AGGAGATCGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGG 140

P4     141 GCGTGGTAGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGGGAATGGCGTGAACCCGGGA 210
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
AluYc1 141 GCGTGGTAGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGA 210

P4     211 GGCGGAGCTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCT 280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AluYc1 211 GGCGGAGCTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCT 280

P4     281 C 281
             |
AluYc1 281 C 281
```

Figure S1. Characterization of the *SCNM1* mutation identified in the proband of family 3 (P4). A) Schematic diagram showing *SCNM1* exon 4 in green and the AluYc1 interrupting this exon detected in P4 in blue. Intronic sequences are represented with dotted lines. Direct repeats corresponding to the target site duplication (TSD) flanking the Alu are indicated as red arrowheads. **B)** Genomic DNA sequence of *SCNM1* exon 4 (highlighted in green), AluYc1 (blue) and *SCNM1* exon 4 upstream and downstream adjacent sequences demonstrating the site of the Alu insertion in P4. TSD nucleotides are underlined, and the 5´ and 3´ splice sites of *SCNM1* exon 4 are in bold italics and shadowed in grey. **C)** DNA alignment showing only one mismatch between the sequence of the AluYc1 inserted in *SCNM1* exon 4 of P4 (P4) and the AluYc1 consensus sequence (AluYc1) obtained from the GIRI Repbase database (https://www.girinst.org/Repbase/). Nucleotide alignment was generated using Censor software tool[1].

Figure S2


C2H2 - ZF SCNM1 domain
SCNM1 acidic domain
NP_076946.1 [Homo sapiens]
Q8K136.1 [Mus musculus]
XP_003642719.1 [Gallus gallus]
NP_001089700.1 [Xenopus laevis]
XP_011603538.1 [Takifugu rubripes]
XP_035671606.1 [Branchiostoma floridae]
XP_030848545.1 [Strongylocentrotus purpuratus]
XP_025206979.1 [Melanaphis sacchari]
ELT87453.1 [Capitella teleta]
XP_005101981.1 [Aplysia californica]
XP_002107820.1 [Trichoplax adhaerens]
ORY06941.1 [Basidiobolus meristosporus]
XP_004368322.1 [Acanthamoeba castellanii]
XP_028765868.1 [Prosopis alba]
XP_004248741.1 [Solanum lycopersicum]
XP_008664530.1 [Zea mays]

Figure S2. Multiple protein sequence alignment of SCNM1 orthologs. Alignment of SCNM1 orthologous proteins from species of different eukaryotic kingdoms conducted with the MUSCLE algorithm[2]. Protein accession numbers used in the alignment are indicated before the name of each species. Amino acids that are conserved in ≥75% of sequences are colored. Regions of the alignment corresponding to C2H2 zinc finger (C2H2-ZF) and acidic domains are highlighted with an upper band in pink or blue, respectively.

Figure S3

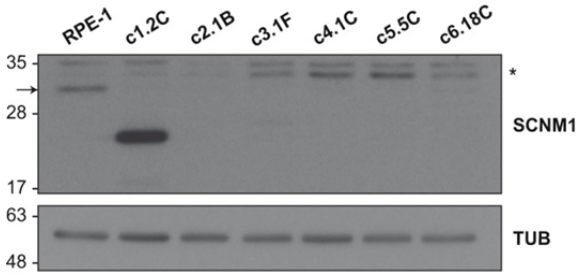

Figure S3. Western Blot analysis of SCNM1 in *SCNM1*-KO RPE-1 clones. Anti-SCNM1 immunoblotting of lysates from different *SCNM1*-KO RPE-1 clones (c1.2C, c2.1B, c3.1F, c4.1C, c5.5C and c6.18C) used in this study which were generated by CRISPR-Cas9 editing, and parental control RPE-1 cells (RPE-1). Wild-type SCNM1 is identified with an arrow, and the asterisk designates nonspecific bands. The lower molecular weight band in c1.2C indicates that an aberrant truncated SCNM1 protein is synthesized in this clone. The same could apply to c3.1F where a faint different band, also with lower molecular weight than the wild-type protein, is detected by the anti-SCNM1 antibody. Tubulin (TUB) was used as loading control (n=3).

Figure S4

A

U12 introns

Total U12 introns analyzed	MORE retained in affected individual (padj<0.05)	MORE retained in affected individual (padj<0.05) (%)	LESS retained in affected individual (padj<0.05)	LESS retained in affected individual (padj<0.05) (%)
489	25	5,11%	0	0%

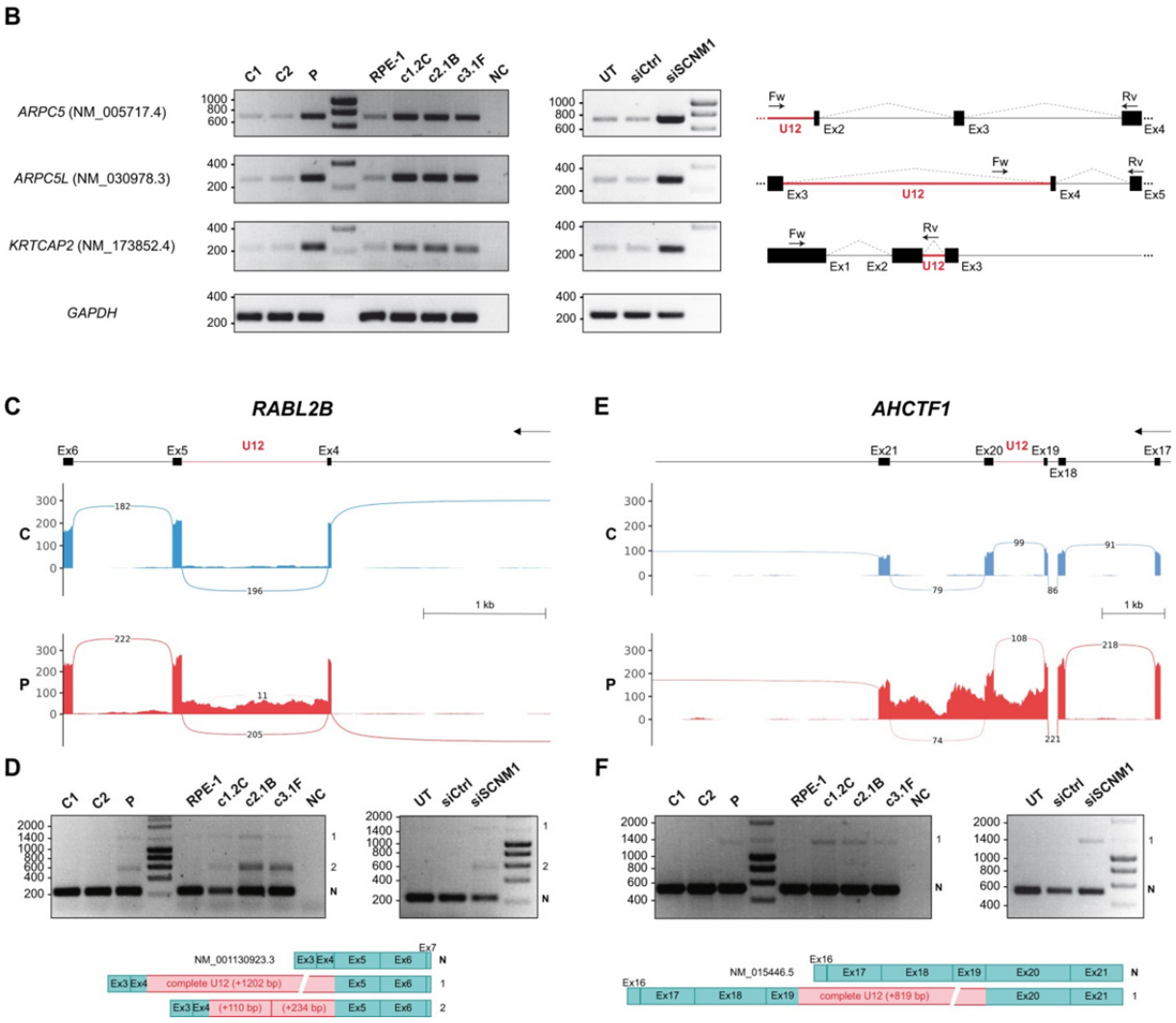

Figure S4. U12 intron retention and alternative splicing defects are increased in SCNM1-deficient cells. **A)** Summary of results from bioinformatics analysis of RNA-seq data using intEREst R package[3] showing number and percentage of differentially retained U12 introns between fibroblasts derived from an affected individual (P2) and control fibroblasts. **B)** Representative examples of exon-intron(U12) semi-quantitative RT-PCR assays. Agarose gel images of exon-intron(U12) RT-PCR for *ARPC5*, *ARPC5L* and *KRTCAP2* showing higher yield of RT-PCR products in SCNM1-deficient cells compared to

their respective controls (n=2). Primary fibroblasts [controls (C1, C2), affected individual P2 (P)], *SCNM1*-KO RPE-1 clones [*SCNM1*-KO clones: c1.2C, c2.1B, c3.1F; parental control cells: RPE-1], and siRNA treated RPE-1 cells [un-transfected (UT), transfected with non-targeted siRNA (siCtrl), transfected with siRNA against *SCNM1* (siSCNM1)]. NC: no cDNA control. Exon-exon primers for the housekeeping gene *GAPDH* were used as control for the quantity of cDNA template. Schematic representations of the genes analyzed are shown to the right of corresponding gel images. The arrows indicate the position of the forward (Fw) and reverse (Rv) RT-PCR primers used. Boxes are exons and the lines introns. U12 introns are highlighted with a red line. Dotted lines represent canonical splicing events. **C-F)** Characterization of *RABL2B* (C-D) and *AHCTF1* (E-F) U12 intron splicing. In panels C) and E), a schematic representation of the genomic region containing the U12 intron of *RABL2B* and *AHCTF1* is represented on top. Boxes indicate exons and lines introns. U12 introns are represented with a red line and the arrows specify the direction of transcription. Sashimi plots of the genomic region of interest from control fibroblasts (C; in blue) and fibroblasts derived from P2 (P; in red) are shown below. In sashimi plots the mean counts from two independent RNA-Seq experiments is represented. Splicing events supported by a minimum of 10 reads are denoted with lines and the number of reads supporting each junction is indicated. Panels D) and F) show representative images of agarose gels corresponding to exon-exon RT-PCR analysis of *RABL2B* and *AHCTF1* in primary fibroblasts, RPE-1 *SCNM1*-KO clones and siRNA-treated RPE-1 cells. Sample labelling is as in panel B (n=2). Normal transcript isoforms are designated with an N and the numbers on the right of each gel denote abnormally spliced products detected in this assay. Schematic representation of the exon composition of RT-PCR products is underneath showing, in each case, partial or complete inclusion of the U12 intron (red).

Figure S5

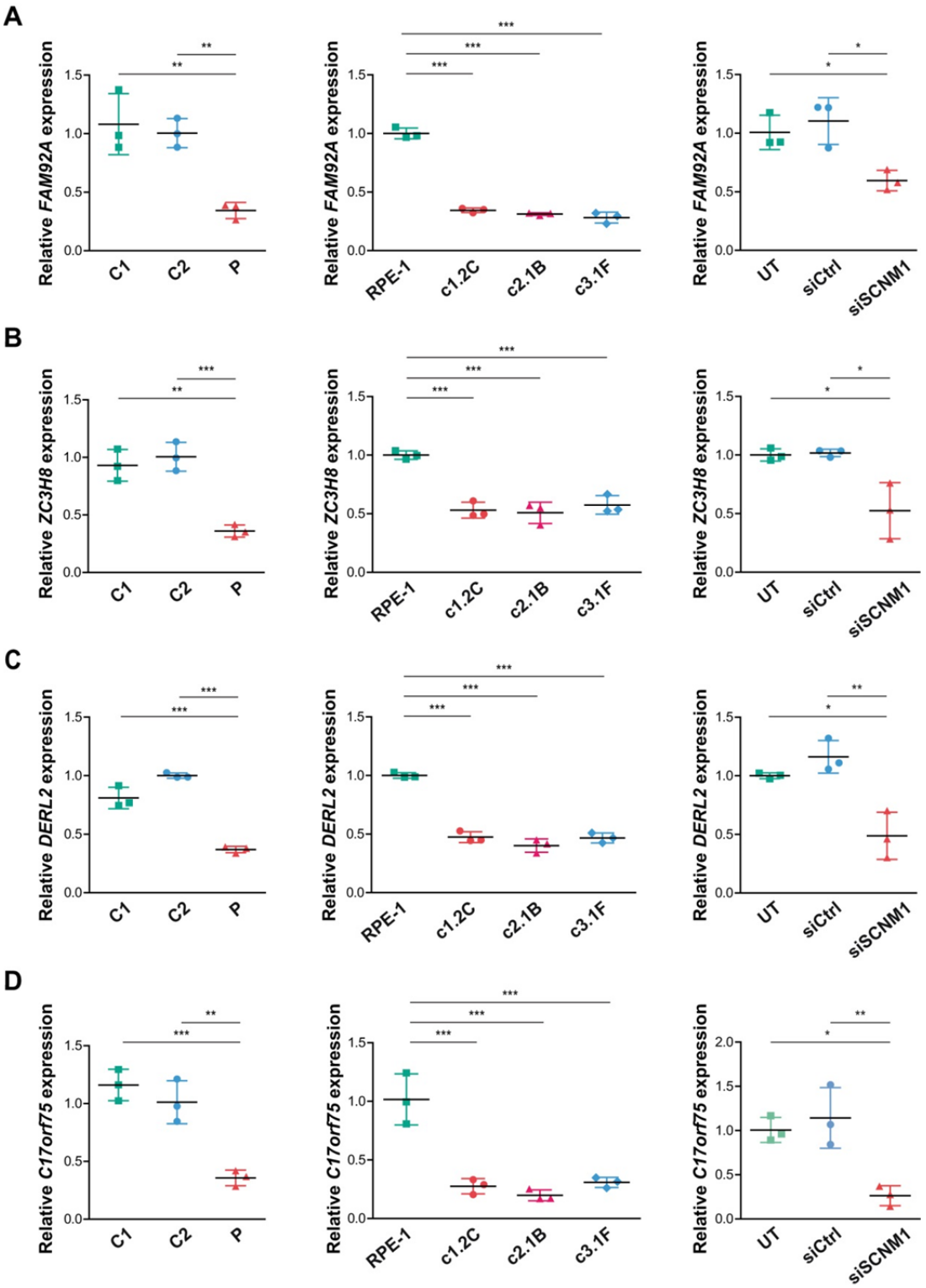

Figure S5. Absence of SCNM1 causes a severe reduction in the transcript levels of several U12 intron-containing genes. A-D) Relative quantification of mRNA levels of *FAM92A*, *DERL2*, *ZC3H8* and *C17orf75* expression by RT-qPCR in primary fibroblasts

[controls (C1, C2), affected individual P2 (P)], *SCNM1*-KO RPE-1 clones [*SCNM1*-KO clones: c1.2C, c2.1B, c3.1F; parental control cells: RPE-1], and siRNA treated RPE-1 cells [un-transfected (UT), transfected with non-targeted siRNA (siCtrl), transfected with siRNA against SCNM1 (siSCNM1)]. Taqman assays were used to quantify *FAM92A*, *DERL2* and *ZC3H8* expression using the geometric mean of *GAPDH* and *GUSB* mRNA levels to normalize data. *C17orf75* mRNA levels were quantified using SYBR Green and RT-qPCR specific primers and were normalized against *GUSB* expression. The ΔCt mean value of C2, RPE-1 or UT was used as calibrator sample in the corresponding experiments. Scatter plots show the mean ± SD (n=3). **p*<0.05, ***p*<0.01, ****p*<0.001. One-way ANOVA with Tukey´s multiple comparison test.

Figure S6

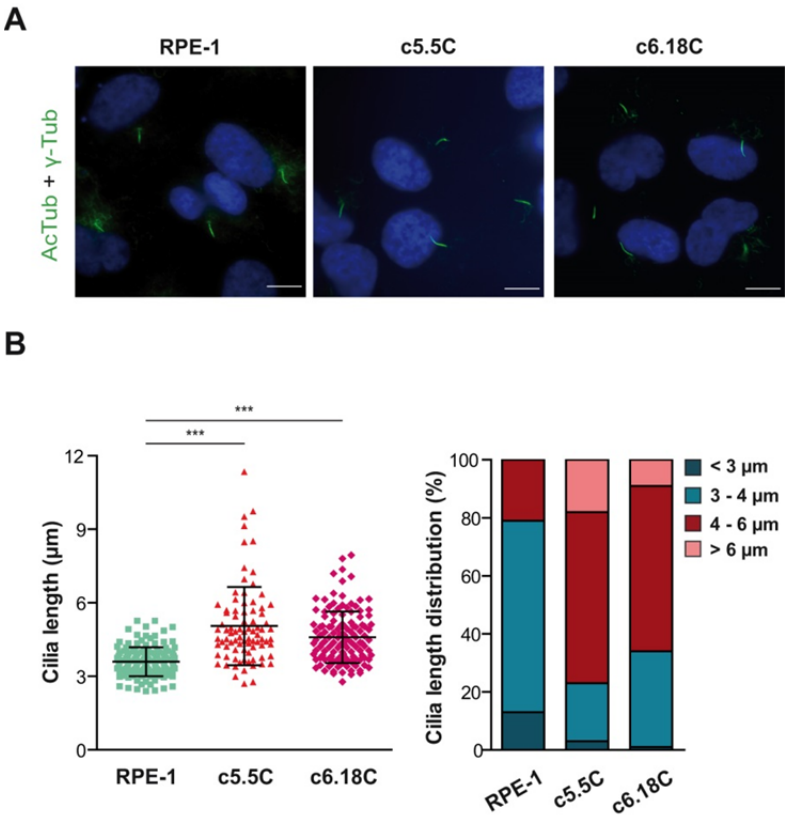

Figure S6. *SCNM1*-KO RPE-1 clones exhibit elongated primary cilia. A) Representative immunofluorescence images showing cilia with increased length in *SCNM1*-KO RPE-1 clones c5.5C and c6.18C compared to RPE-1 parental control cells. Green [acetylated tubulin (AcTub) + gamma tubulin (γ-Tub)]: cilia; blue (DAPI): nuclei. Scale bars: 10 µm. **B)** Graphs representing cilia length and distribution of cilia length in *SCNM1*-KO RPE-1 clones (c5.5C and c6.18C) and RPE-1 control cells. At least 85 cilia from 3 different experiments were measured per cell line. Data are presented as mean ± SD. ***p<0.001. Kruskal-Wallis test with Dunn's multiple comparison test.

Figure S7

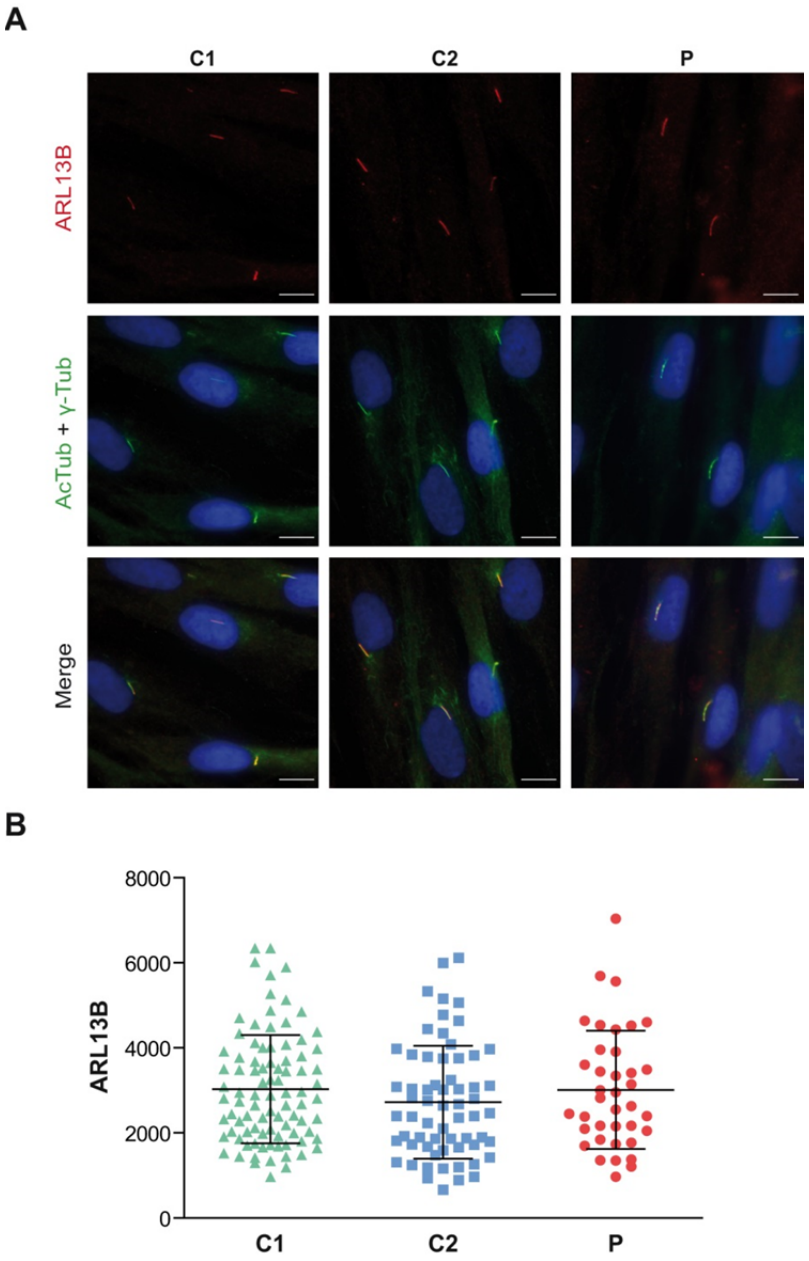

Figure S7. ARL13B ciliary localization is normal in SCNM1-deficient primary fibroblasts. A) Representative images showing positive ARL13B immunostaining (red) in primary cilia of fibroblasts from controls (C1, C2) and affected individual P2 (P) (n=3). Green [acetylated tubulin (AcTub) + gamma tubulin (γ-Tub)]: cilia, DAPI (blue): nuclei. Scale bars: 10 µm. **B)** Scatter dot plot of ciliary ARL13B integrated density (y-axis) showing no statistically significant differences between fibroblasts from controls (C1, C2) and affected individual P2 (P). At least 38 ARL13B positive cilia were analyzed per cell line from two independent experiments. Data are presented as mean ± SD. Kruskal-Wallis test with Dunn´s multiple comparison test.

Figure S8

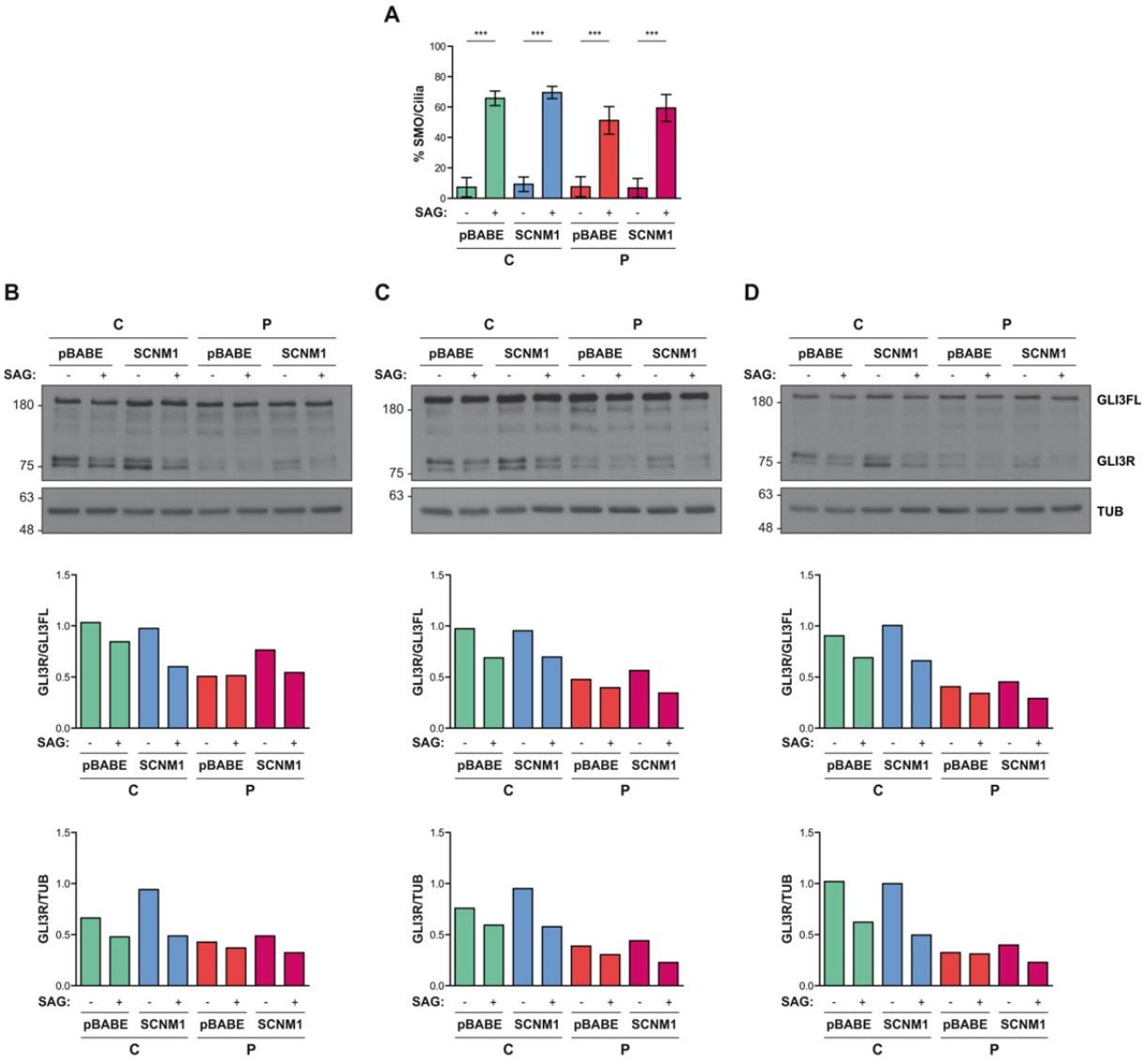

Figure S8. Analysis of the ciliary localization of SMO and protein levels of GLI3 in retrotransduced fibroblasts . A) Percentage of SMO positive cilia upon treatment with SAG (+) or DMSO (-) in control (C) and affected individual (P) fibroblasts retrotransduced with *SCNM1* or the empty retroviral vector pBABE-puro (pBABE). At least 90 cilia per condition were analyzed from three independent experiments. Data show the mean ± SD. ****p*<0.001. One-way ANOVA with Tukey´s multiple comparison test. **B-D)** GLI3 immunoblots corresponding to three independent biological replicates illustrating protein levels of GLI3-full length (GLI3FL) and GLI3-repressor (GLI3R) upon SAG (+) or DMSO (-) incubation in control (C) and affected individual (P) fibroblasts retrotransduced with *SCNM1* or the empty vector (pBABE). Tubulin (TUB) served as a loading control. For each individual blot, the graphs underneath show the ratio between the values of GLI3R and GLI3FL or GLI3R and TUB calculated by band densitometry.

II. Supplemental Tables

Table S1. List of RT-PCR primers used in this work for the analysis of the splicing of U12 intron-containing genes:

a) Exon-exon RT-PCR primers for OFD genes containing U12 introns and *SCN8A*

Primer	Sequence (5'-3')
C2CD3 (Fw)	CAAGCCTTTCCTACCCCTCA
C2CD3 (Rv)	TCCGTGCAATCCTGAGAGTT
TBC1D32 (Fw)	TGCTACTTCCCTTCAGTGGA
TBC1D32 (Rv)	CTCACTCGAAACCCATGCAG
TCTN3 (Fw)	TGGCTGGAAACACTTGTCAG
TCTN3 (Rv)	TGCCCACAATACCTGGATCT
TMEM107 (Fw)	GCATCTGGTGGTCGTCATC
TMEM107 (Rv)	AGGTTCCCGTCATGAAGGTA
TMEM231 (Fw)	TGTGCAGCTCATCCTGACTT
TMEM231 (Rv)	TCTGAGGGAAAAGCACACAA
SCN8A (Fw)	TGCTGAGAGCAAGCTCAAGA
SCN8A (Rv)	ACCCTGAAAGTGCGTAGAGC

b) Exon-intron(U12) RT-PCR primers used for evaluation of the retention levels of U12 introns identified through bioinformatics analysis

Primer	Sequence (5'-3')
AHCTF1 (Fw)	GGGCTTTTACCAGAAGGCATAGAT
AHCTF1-U12 (Rv)	ATATGTCAGGTTTCCACTACTC
ARPC5-U12 (Fw)	ATCTGCCAGCATGTTCTACC
ARPC5 (Rv)	AGCAAGTGCCTTTTCATGCC
ARPC5L-U12 (Fw)	TGACGGCATGGATGTGCTTG
ARPC5L (Rv)	CGCTGCTATTTTCTGTGGGC
C17orf75-U12 (Fw)	CTTGTTGAGAGATGAAATACTGTCTTC
C17orf75 (Rv)	GCTGCTGCAATCGATGACCACAGAC
CYBA-U12 (Fw)	CTGGGCTGTTCCTTAACCAC
CYBA (Rv)	TTGATGGTGCCTCCGATCTG
DCTN3 (Fw)	ATGCTGGACAGTGCTCACATC
DCTN3-U12 (Rv)	CATTAGGGTCCTAGTTGAGC
DERL2-U12 (Fw) ID: 1194	CAGGTGGGTTACTGTTAAGC
DERL2 (Rv)	TGGCCGTTCCTCAGGTAGT
DOCK1 (Fw)	GCCGAGCAGTATGAGAACGA
DOCK1-U12 (Rv)	GGCAGCATAGACATCAACCATG
FAM92A (Fw)	AAAACCGGAACGCTCAAACG
FAM92A-U12 (Rv)	TGAATTAGGTGGCGTAACCTCT
KIFAP3 (Fw)	CGGGGTTCTGCTTTGATCCT
KIFAP3-U12 (Rv) ID: 1287	TTACCTGCTCCTTCTACC

KRTCAP2 (Fw)	TTCAGCTCGCCTTTCTTGG
KRTCAP2-U12 (Rv)	CTCCAACTGCAGGGAGGATAC
MAPK1 (Fw)	ACACCAACCTCTCGTACATCG
MAPK1-U12 (Rv)	CAGGGCACCCTCTTTCTTAG
MARCHF6 (Fw)	TAACCCACCAGCTGAGAACG
MARCHF6-U12 (Rv)	TTGTCCACACTGACCCTTAC
RABL2B (Fw)[#]	GAGTGAGTTGGACCAAGGGA
RABL2B-U12 (Rv)[#]	TGGGTCAATCGGCAAGAAAC
SBNO1 (Fw)	TATGCAGCCCAGCAACATGA
SBNO1-U12 (Rv) ID: 1411	CCTGTTTCCCAAAGGTCTAC
SLC12A4-U12 (Fw) ID: 1283	GGCACGTGGTACTATTCAGG
SLC12A4 (Rv)	AGTGCCCATGTGGGTTCAC
TMEM87B (Fw)	GCAACACTGGACTGTCAACC
TMEM87B-U12 (Rv)	TCCCCTCCGTTCTTCCTAGA
ZC3H8 (Fw)	GCTGGTCACAAGAATGGCAA
ZC3H8-U12 (Rv)	CACTTTGATACTAAGTTGGGCTTGG

c) Exon-exon RT-PCR primers used for evaluation of splicing efficiency of U12 introns identified through bioinformatics analysis

Primer	Sequence (5'-3')
AHCTF1 (Fw)	GGGCTTTTACCAGAAGGCATAGAT
AHCTF1 (Rv)	CGCCCTGACTCATGAATGCC
ARPC5 (Fw)	CGGGATTGGGATGTCGAAGA
ARPC5 (Rv)	AGCAAGTGCCTTTTCATGCC
ARPC5L (Fw)	CGGGTGGACATCGACGAAT
ARPC5L (Rv)	CGCTGCTATTTTCTGTGGGC
C17orf75 (Fw)	CTACCATCCGAAGTGGAGCC
C17orf75 (Rv)	GTCCTTGAAGACTGGCCACA
CYBA (Fw)	CAGATCGAGTGGGCCATGTG
CYBA (Rv)	CAGTAGGTAGATGCCGCTCG
DCTN3 (Fw)	GCTTGGCCCAGATCCACATT
DCTN3 (Rv)	AAAGGCAGGTTGGCATAGGG
DERL2 (Fw)	ATGGCGTACCAGAGCTTGC
DERL2 (Rv)	TCCACCAGGTTGATTGGGAA
DOCK1 (Fw)	GCCGAGCAGTATGAGAACGA
DOCK1 (Rv)	CATTCTCCAGGGGGCTGATT
FAM92A (Fw)	GCTACAGAGACCCCGCATTT
FAM92A (Rv)	TCTGGTAGGCAGCAGTGTAGA
KIFAP3 (Fw)	CGGGGTTCTGCTTTGATCCT
KIFAP3 (Rv)	TCCCGATCAAGGGCTTTCAC
KRTCAP2 (Fw)	TTCAGCTCGCCTTTCTTGG
KRTCAP2 (Rv)	GCAGGTGGTGACACAGACTC
MAPK1 (Fw)	ACACCAACCTCTCGTACATCG
MAPK1 (Rv)	TTGAGCAGCAGGTTGGAAGG
MARCHF6 (Fw)	TAACCCACCAGCTGAGAACG

MARCHF6 (Rv)	AAACCACAGGACACCAGGTC
RABL2B (Fw)[#]	GATCATCTGCCTGGGAGACA
RABL2B (Rv)[#]	TCAAACACCATGATGCAGGC
SBNO1 (Fw)	TATGCAGCCCAGCAACATGA
SBNO1 (Rv)	TACCCTCACCCCATATGCCA
SLC12A4 (Fw)	AAGGAGAGCCTGCCTCTGTA
SLC12A4 (Rv)	AGTTCCTGCTGACACCATCG
TMEM87B (Fw)	GCAACACTGGACTGTCAACC
TMEM87B (Rv)	TGTCTTAGTTGTCCACCCCA
ZC3H8 (Fw)	GCTGGTCACAAGAATGGCAA
ZC3H8 (Rv)	ACAGTCTTGAACACGCATGG
hu*GAPDH* (Fw)*	GAAGGTGAAGGTCGGAGTC
hu*GAPDH* (Rv)*	GAAGATGGTGATGGGATTTC

Asterisks designate primers for *GAPDH*. This gene was used as housekeeping gene for semi-quantitative RT-PCR analysis in a) b) and c). The # symbol indicates that due to high identity between *RABL2B* and *RABL2A* genomic sequences (98%), primers cannot discern between these two genes. Fw: forward primer, Rv: reverse primer.

Table 3b: For those genes containing more than one U12 intron, only U12 introns detected as retained by bioinformatics analysis were tested. In this case, the ID number given by the minor intron database[4] to these introns is indicated. Primers annealing to U12 intron nucleotide sequences are labelled as “gene name–U12”.

Table S2. Genomic coordinates of the region of homozygosity (ROH) on chromosome 1 comprising *SCNM1* of each affected individual (P1-P4).

Individual	ROH (GRCh37/hg19)
P-1	chr1: 88,129,160 - 166,504,879
P-2	chr1: 88,129,160 - 166,504,879
P-3	chr1: 99,416,102 - 157,275,294
P-4	chr1: 111,531,126 - 151,547,397
common ROH	chr1: 111,531,126 - 151,547,397
SCNM1	chr1: 151,138,498 - 151,142,773

Table S3. Genes containing the 25 U12 introns detected as significantly retained in SCNM1-deficient fibroblasts through bioinformatics analysis, and summary of exon-exon RT-PCR results for the 18 analyzed genes.

a) Summary of abnormal alternative splicing events detected by exon-exon RT-PCR in the 18 analyzed genes; and Gene Ontology (GO) terms (biological process) for these genes.

U12 intron containing gene	Defective spliced isoforms detected	Biological process (GO)
AHCTF1	CU12IR	GO0032465: regulation of cytokinesis; GO0051292: nuclear pore complex assembly
ARPC5	-	GO0034314: Arp2/3 complex-mediated actin nucleation; GO0016477: cell migration
ARPC5L	-	GO0034314: Arp2/3 complex-mediated actin nucleation; GO0016477: cell migration
C17orf75	CU12IR	GO0099041: vesicle tethering to Golgi; GO0006886: intracellular protein transport
CYBA	-	GO0014895: smooth muscle hypertrophy; GO0006954: inflammatory response; GO0045087: innate immune response; GO0045730: respiratory burst
DCTN3	CU12IR	GO0061640: cytoskeleton-dependent cytokinesis
DERL2	CU12IR, A5SS, A3SS, ES	GO0030433: ubiquitin-dependent ERAD pathway; GO0030307: positive regulation of cell growth
DOCK1	CU12IR	GO0007229: integrin-mediated signaling pathway; GO0006911: phagocytosis, engulfment; GO0010634: positive regulation of epithelial cell migration; GO0007264: small GTPase mediated signal transduction
FAM92A (*CIBAR1*)	A5SS, ES, CEI	GO0060271: cilium assembly; GO0007007: inner mitochondrial membrane organization; GO0035108: limb morphogenesis; GO0045880: positive regulation of smoothened signaling pathway
KIFAP3	-	GO0044782: cilium organization; GO0072383: plus-end-directed vesicle transport along microtubule; GO0065003: protein-containing complex assembly; GO0007165: signal transduction
KRTCAP2	CU12IR	GO0042543: protein N-linked glycosylation via arginine
MAPK1	-	GO0070371: ERK1 and ERK2 cascade
MARCHF6	CU12IR	GO0030433: ubiquitin-dependent ERAD pathway; GO1904380: endoplasmic reticulum mannose trimming

RABL2B	CU12IR, A5SS, ES	GO0042073: intraciliary transport; GO0060271: cilium assembly
SBNO1	CU12IR	GO0006355: regulation of transcription, DNA-templated
SLC12A4	CU12IR	GO0006884: cell volume homeostasis; GO0007268: chemical synaptic transmission
TMEM87B	-	GO0042147: retrograde transport, endosome to Golgi
ZC3H8	A5SS, A3SS	GO0042795: snRNA transcription by RNA polymerase II; GO0042796: snRNA transcription by RNA polymerase III; GO0043029: T cell homeostasis

b) GO terms (biological process) for the remaining genes identified by bioinformatics analysis with increased U12 intron retention.

U12 intron containing gene	**Biological process (GO)**
MAEA	GO0007155: cell adhesion; GO0033033: negative regulation of myeloid cell apoptotic process; GO0043161: proteasome-mediated ubiquitin-dependent protein catabolic process; GO0007346: regulation of mitotic cell cycle
TMEM161B	GO0098901: regulation of cardiac muscle cell action potential
FIG4	GO0006661: phosphatidylinositol biosynthetic process
MED23	GO0006367: transcription initiation from RNA polymerase II promoter
ERCC5	GO0006289: nucleotide-excision repair; GO0043066: negative regulation of apoptotic process
ALG12	GO0006457: protein folding
SMS	GO0006597: spermine biosynthetic process

Abbreviations in a): CU12IR (Complete U12 Intron Retention), A5SS (Activation of cryptic 5' Splice Site [within U12 intron or adjacent exon]), A3SS (Activation of cryptic 3' Splice Site [within U12 intron or adjacent exon]), ES (Exon Skipping) and CEI (Cryptic Exon Inclusion). For each gene, a selection of manually curated GO terms for biological process listed in UniProt database[5] is included.

III. Supplemental References

1. Kohany, O., Gentles, A.J., Hankus, L., and Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7, 474.

2. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.

3. Oghabian, A., Greco, D., and Frilander, M.J. (2018). IntEREst: intron-exon retention estimator. BMC Bioinformatics 19, 130.

4. Olthof, A.M., Hyatt, K.C., and Kanadia, R.N. (2019). Minor intron splicing revisited: identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. BMC Genomics 20, 686.

5. UniProt, C. (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49, D480-D489.