



UNIVERSITY OF LEEDS

This is a repository copy of *An siRNA-based functional genomics screen for the identification of regulators of ciliogenesis and ciliopathy genes*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/88332/>

Version: Accepted Version

Article:

Wheway, G, Schmidts, M, Mans, DA et al. (75 more authors) (2015) An siRNA-based functional genomics screen for the identification of regulators of ciliogenesis and ciliopathy genes. *Nature Cell Biology*, 17 (8). pp. 1074-1087. ISSN 1465-7392

<https://doi.org/10.1038/ncb3201>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

An siRNA-based functional genomics screen for the identification of regulators of ciliogenesis and ciliopathy genes

Gabrielle Wheway*¹, Miriam Schmidts*^{2,3,4,5}, Dorus A. Mans*^{3,4}, Katarzyna Szymanska*¹, Thanh-Minh T. Nguyen*^{3,4}, Hilary Racher⁶, Ian G. Phelps⁷, Grischa Toedt⁸, Julie Kennedy⁹, Kirsten A. Wunderlich¹⁰, Nasrin Soroush¹⁰, Zakia A. Abdelhamed¹, Subaashini Natarajan¹, Warren Herridge¹, Jeroen van Reeuwijk^{3,4}, Nicola Horn¹¹, Karsten Boldt¹¹, David A. Parry¹², Stef J.F. Letteboer^{3,4}, Susanne Roosing¹³, Matthew Adams¹⁴, Sandra M. Bell¹⁴, Jacquelyn Bond¹⁴, Julie Higgins¹⁴, Ewan E. Morrison¹⁴, Darren C. Tomlinson¹⁴, Gisela G. Slaats¹⁵, Teunis J. P. van Dam¹⁶, Lijia Huang¹⁷, Kristin Kessler¹⁸, Andreas Giessl¹⁹, Clare V. Logan¹, Evan A. Boyle²⁰, Jay Shendure²⁰, Shamsa Anazi²¹, Mohammed Aldahmesh²¹, Selwa Al Hazzaa^{22,23}, Robert A. Hegele²⁴, Carole Ober²⁵, Patrick Frosk²⁶, Aizeddin A. Mhanni²⁶, Bernard N. Chodirker²⁶, Albert E. Chudley²⁶, Ryan Lamont⁶, Francois P. Bernier⁶, Chandree L. Beaulieu¹⁷, Paul Gordon⁶, Richard T. Pon⁶, Clem Donahue²⁷, A. James Barkovich²⁸, Louis Wolf²⁹, Carmel Toomes¹, Christian T. Thiel¹⁸, Kym M. Boycott¹⁷, Martin McKibbin³⁰, Chris F. Inglehearn¹, UK10K Consortium³¹, University of Washington Center for Mendelian Genomics³², Fiona Stewart³³, Heymut Omran³⁴, Martijn A. Huynen¹⁶, Panagiotis I. Sergouniotis^{35,36}, Fowzan S. Alkuraya^{21,37}, Jillian S. Parboosingh⁶, A Micheil Innes⁶, Colin E. Willoughby³⁸, Rachel H. Giles¹⁵, Andrew R. Webster^{35,36}, Marius Ueffing^{11,39}, Oliver Blacque⁹, Joseph G. Gleeson¹³, Uwe Wolfrum¹⁰, Philip L. Beales², Toby Gibson⁸, Dan Doherty^{7,40}, Hannah M. Mitchison², Ronald Roepman^{3,4}, Colin A. Johnson¹

* contributed equally

+ corresponding authors

¹ Section of Ophthalmology and Neuroscience, Leeds Institutes of Molecular Medicine, University of Leeds, Leeds, LS9 7TF, UK

² Genetics and Genomic Medicine and Birth Defects Research Centre, Institute of Child Health, University College London, London, WC1N 1EH, UK

³ Department of Human Genetics, Radboud University Medical Center, Nijmegen, 6525 GA, The Netherlands

⁴ Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, 6525 GA, The Netherlands

⁵ Pediatric Genetics Section, Center for Pediatrics and Adolescent Medicine, University Hospital Freiburg, Freiburg 79112, Germany

⁶ Department of Medical Genetics and Alberta Children's Hospital Research Institute for Child and Maternal Health, Calgary, T3B 6A8, AB, Canada

⁷ Department of Pediatrics, University of Washington, Seattle, WA 98195, USA

⁸ Structural and Computational Biology, European Molecular Biology Laboratory, 69117 Heidelberg, Germany

- ⁹ School of Biomolecular and Biomedical Science, University College Dublin, Dublin 4, Ireland
- ¹⁰ Department of Cell and Matrix Biology, Institute of Zoology, Johannes Gutenberg University of Mainz, 55122 Mainz, Germany
- ¹¹ Division of Experimental Ophthalmology and Medical Proteome Center, Center of Ophthalmology, University of Tübingen, 72074 Tübingen, Germany
- ¹² Section of Genetics, Leeds Institutes of Molecular Medicine, University of Leeds, Leeds, LS9 7TF, UK
- ¹³ Laboratory for Pediatric Brain Disease, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Ave, Box 268, New York, NY 10065, USA
- ¹⁴ BioScreening Technology Group, Biomedical Health Research Centre, St. James's University Hospital, Leeds, LS9 7TF, UK
- ¹⁵ Department of Nephrology and Hypertension, University Medical Centre Utrecht, Utrecht, 3584 CX, The Netherlands
- ¹⁶ Centre for Molecular and Biomolecular Informatics, Radboud University Medical Center, Nijmegen, 6525 GA, The Netherlands
- ¹⁷ Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, K1H 8L1, ON, Canada
- ¹⁸ Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany
- ¹⁹ Animal Physiology, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany
- ²⁰ Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
- ²¹ Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia
- ²² Department of Ophthalmology, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia
- ²³ Department of Ophthalmology, College of Medicine, Alfaisal University, Riyadh, 11533, Saudi Arabia
- ²⁴ Robarts Research Institute, University of Western Ontario, London, N6G 2V4, ON, Canada
- ²⁵ Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
- ²⁶ Department of Pediatrics and Child Health & Department of Biochemistry and Medical Genetics, Faculty of Medicine, University of Manitoba, Winnipeg, R3E 3P5, MB, Canada
- ²⁷ Department of Pediatrics, University of California San Francisco, San Francisco, CA 94143, USA

²⁸ Department of Radiology, University of California San Francisco, San Francisco, CA 92093, USA

²⁹ Department of Pathology, Radboud University Medical Center, Nijmegen, 6525 GA, The Netherlands

³⁰ Department of Ophthalmology, Leeds Teaching Hospitals NHS Trust, St. James's University Hospital, Leeds, LS9 7TF, UK

³¹ UK10K Consortium, Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK

³² University of Washington Center for Mendelian Genomics, University of Washington, Seattle, WA 98195, USA

³³ Department of Medical Genetics, Belfast City Hospital and Queens University, Belfast, BT12 6BA, UK

³⁴ Department of Pediatrics and Adolescent Medicine, University Hospital Muenster, 48149 Muenster, Germany

³⁵ Moorfields Eye Hospital NHS Foundation Trust and NIHR Ophthalmology Biomedical Research Centre, London, EC1V 2PD, UK

³⁶ UCL Institute of Ophthalmology, University College London, London, EC1V 9EL, UK

³⁷ Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh 11533, Saudi Arabia

³⁸ Department of Eye and Vision Science, Institute of Ageing and Chronic Disease, Faculty of Health & Life Sciences, University of Liverpool, Liverpool, L69 3BX, UK

³⁹ Research Unit of Protein Science, Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt, 85764 Neuherberg, Germany

⁴⁰ Divisions of Developmental Medicine and Genetic Medicine, Seattle Children's Research Institute, University of Washington, Seattle, WA 98105, USA

Keywords: cilia, ciliopathies, reverse genetics, whole-genome siRNA screen, Jeune syndrome, Joubert syndrome

ABSTRACT

Defects in primary cilium biogenesis underlie the ciliopathies, a growing group of genetic disorders. We describe a whole genome siRNA-based reverse genetics screen for defects in biogenesis and/or maintenance of the primary cilium, obtaining a global resource for investigation into essential ciliary processes. We identify 112 candidate ciliogenesis and ciliopathy genes, including 44 components of the ubiquitin-proteasome system, 12 G-protein-coupled receptors, and three pre-mRNA processing factors (PRPF6, PRPF8 and PRPF31) mutated in autosomal dominant retinitis pigmentosa. The PRPFs localise to the connecting cilium, and *PRPF8*- and *PRPF31*-mutated cells have ciliary defects. Combining the screen with exome sequencing data also identified recessive mutations in two screen candidate genes (*PIBF1/CEP90* and *C21orf2/LRRC76*) as causes of the ciliopathies Joubert and Jeune syndromes. Affinity proteomics and protein-protein interaction studies place *C21orf2* within key ciliopathy-associated protein modules, offering an explanation for the skeletal and retinal involvement observed in individuals with *C21orf2*-variants. Our global, unbiased approaches provide insights into ciliogenesis complexity and identify roles for unanticipated pathways in human genetic disease.

The primary cilium is a microtubule-based organelle present on the apical surface of most vertebrate cells. Primary cilia are regarded as cellular signalling hubs, regulating diverse signalling pathways, with particularly important roles during embryonic development and the patterning of the developing neural tube¹. Defects in primary cilia are associated with heterogeneous inherited developmental conditions known as the ciliopathies that often present with cystic kidney disease, and other diverse multi-organ phenotypes that affect the central nervous system, eye and skeleton². As a group, ciliopathies are comparatively common Mendelian inherited conditions with an overall estimated prevalence of 1 in 2000³.

The importance of the primary cilium has only become apparent in the last decade, and therefore processes involved in cilia formation and maintenance remain poorly characterised. To address this in the most comprehensive way, we have carried out a high-throughput cilia characterisation assay to interrogate the involvement of every gene in the formation of the primary cilium. This represents a global, hypothesis-neutral approach to identify genes involved in ciliogenesis. It also provides a complementary approach to traditional linkage analysis, candidate gene, or whole exome sequencing (WES) approaches for gene discovery. Previously, several cell-based medium- or small-scale screens of ciliogenesis have been published, performed on subsets of target genes such as a druggable library (7784 targets)⁴ and uncharacterised components of the cilia proteome (40 targets)⁵. Here, we present a whole genome knockdown screen in a ciliated cell line that provides a comprehensive view of critical factors involved in ciliogenesis and offers numerous candidate genes for ciliopathies.

RESULTS

Genome-wide siRNA screen for ciliogenesis

We used small interfering RNA (siRNA) to knockdown and ascertain the function of 19059 Entrez RefSeq (B37.1) mouse genes across the genome. The siRNA screen was carried out in the mouse inner medullary collecting duct (mIMCD3) ciliated cell-line using a Thermo Fisher siGENOME library containing pools of four siRNA duplexes per gene (see **On-line Methods** and **Supplementary Note** for further details). We assayed for a loss-of-cilium phenotype 72 hours after transfection by imaging at three focal planes to detect the nucleus (DAPI), cytoplasm (TOTO3) and cilium (acetylated α -tubulin) (**Figure 1a**). The cilia could be identified as single spots above the nucleus of the cell (**Figure 1b,c**), allowing the use of simple cilia recognition protocols based on optimised Perkin-Elmer spot detection algorithms (**Figure 1b**).

The statistical significance of effects on cilia and cell number after siRNA knockdown was tested by calculating robust z scores⁶ (**Figure 1d**, **Suppl. Table 1**). To ensure normalisation of data and exclusion of batch-specific effects, data were analysed within processed batches (**Figure 1e,f**). Duplicate assays of batches resulted in little variation, with a median Pearson's correlation coefficient between replicates of 0.71 (**Figure 2a**), and an average strictly standardised mean difference (SSMD) value for all batches of 1.717 (**Figure 1e**). Robust z scores for cell number (z_{cell}) and the percentage of cells with a single cilium (z_{cilia}) were calculated for all results, compared to the median and median absolute deviation (MAD) of all positive and negative controls in a processed batch (**Figure 2a,b**, **Suppl. Table 1**). This allowed application of meaningful statistical cut-off values ($z_{\text{cilia cutoff}}$ and $z_{\text{cell cutoff}}$) for the identification of significant "hits" affecting ciliogenesis based on the median and MAD of positive and negative controls per 10-plate batch⁷ (**Figure 1f**).

Filtering strategy, bioinformatics analysis, and validation screens

In total, 2174 siRNA pools significantly decreased the percentage of cells with a single primary cilium ($z_{\text{cilia}} < z_{\text{cilia cutoff}}$ in a processed batch for both replicates, where $z_{\text{cilia cutoff}} =$

median z of all positive controls), with 1956 hits targeting a gene with a human orthologue (**Figure 2c, Suppl. Table 2**). We filtered out potential non-specific siRNAs comprising those with predicted off-target effects or with microRNA-like effects (see **Supplementary Note**) leaving a total of 1829 mouse genes with a human orthologue (**Suppl. Table 2**). The list of 1829 genes was significantly enriched in known ciliary components (the SYSCILIA Gold Standard⁸; $p=6.7 \times 10^{-4}$ hypergeometric test, observed 48, expected 31.2). Functional annotation clustering with enrichment analysis using DAVID^{9,10} and relevant literature information was used to select representative functional candidate hits for further validation in a secondary screen. We selected a total of $n=40$ hits that were components of significantly enriched pathways (**Figure 2d** left panel, **Suppl. Table 3**) including the spliceosome, the proteasome, and the ubiquitin-proteasome system (UPS).

To exclude the effect of proliferation on ciliogenesis, we filtered the 1829 primary screen hits to prioritise those that significantly affected the percentage of cells with a single cilium but did not have a significant effect on cell number ($Z_{\text{cell}} > Z_{\text{cell cutoff}}$) in a processing batch for both replicates, where $Z_{\text{cell cutoff}} = \text{median } z - 2\text{MAD}$ for all negative controls; see **On-line Methods**). This filtering gave a total of $n=154$ hits (**Suppl. Table 3**) that were significantly enriched in known ciliary components ($p=7.7 \times 10^{-4}$ hypergeometric test, observed 9, expected 2.5), and were taken forward for validation. DAVID analysis of this list identified enrichment of GPCRs, and components of the non-motile primary cilium and the photoreceptor outer segment (**Figure 2d** right panel). In total, 194 functional candidates were taken forward for validation in a secondary screen.

Secondary screening in mIMCD3 cells validated a total of $n=68/194$ mouse genes that had a significant effect on cilia number ($Z_{\text{cilia}} < Z_{\text{cilia cutoff}}$) for $n \geq 1$ out of 4 individual siRNAs in duplicate, and $n=29/194$ for $n \geq 2$ out of 4 individual siRNAs in duplicate (**Table 1, Suppl. Table 3**). Validated hits included genes identified previously in functional genomics of centriole biogenesis in human cells¹¹ including *PLK4* and *CEP120*, and other hits with known ciliary roles including *CCDC41*¹² and *OFD1*^{11,13}. *PLK4* is implicated in the regulation of both

centriole biogenesis and cilia-dependent processes, with mutations causing microcephaly, growth failure and retinopathy^{14,15}. In addition, mutations in *CEP120* are a cause of the skeletal ciliopathy Jeune asphyxiating thoracic dystrophy¹⁶, suggesting that our screen has a high predictive value to identify genes involved in ciliary processes. Functional classifications for a selection of these validated genes are shown in **Figure 3a**. Interestingly, the two hits PRPF8 and PRPF38A have also been implicated in the process of centriolar under-duplication¹¹.

Tertiary screening in hTERT-RPE1 cells using pooled siRNAs enabled the assessment of increase or decrease in both cilia number and/or cilia length. From the hits that were validated by the secondary screen, n=37/68 human genes had defects in cilia number and/or length (using conventional cut-offs of $z < -2$ or $z > +2$) including: *C21orf2*, *CCDC41*, *OFD1*, *PIBF1* and *PLK4*, and several PRPFs (*PRPF31*, *PRPF6* and *PRPF8*; **Table 1, Suppl. Table 3**). Since the tertiary screen validated representative UPS hits (*HUWE1* and *USP39*; **Table 1, Suppl. Table 3**), we used hTERT-RPE1 cells and individual siRNAs in a secondary screen to investigate all enriched human UPS components from the primary screen (n=57/1829). This validated a total of n=44/57 genes that had a significant effect on cilia number and/or cilia length (using cut-offs of $z < -2$ or $z > +2$) for $n \geq 2$ out of 3 individual duplexes (**Suppl. Table 4**; see **Supplementary Note** for further details). Interestingly, out of the 24 proteasome subunits represented in this secondary screen, knockdown of the majority (n=17/24) caused a significant increase in cilia number (**Suppl. Table 4**).

Validation screens identify PRPFs and GPCRs as components of ciliogenesis pathways

Based on our significance analyses, functional categorisation, relevant literature information, and availability of validated reagents, we chose a total of 15 genes encoding GPCRs, PRPFs, and predicted centrosomal proteins for further study. For each gene we measured a reduction in mRNA levels by quantitative PCR (**Figure 3b**), and, when possible, decrease in protein levels by western immunoblot analysis (**Figure 3c**). We confirmed the

decrease in cilia number was accompanied by a reduction in immunostaining of the knocked-down protein by immunofluorescence (IF) microscopy (**Figure 3d, Suppl. Figure 1a**). 3D spheroid culture of mIMCD3 cells was also used as a more physiologically relevant model to confirm the effect of *Plk4* knockdown on cilia number (**Figure 3e**).

PRPFs were selected for further analysis since *PRPF6*, *PRPF8* and *PRPF31* are all mutated in autosomal dominant retinitis pigmentosa (RP types 60, 13 and 11, respectively). The pathogenic mechanism for these forms of RP remains poorly understood, and none have been characterised as non-syndromic retinal ciliopathies. Although *PRPF6*, *PRPF8* and *PRPF31* predominantly localised to the nuclear speckles as expected (**Figure 3d, 4a-c**), we also detected co-localisation of these proteins to the base of the cilium in diverse human and mouse ciliated cell-lines (**Figure 4a**) and to the cilium of photoreceptor cells in adult mouse retina (**Figure 4b**). Immunoelectron microscopy staining showed that *PRPF6* and *PRPF8* localised to the apical inner segment, basal body complex, apical connecting cilium of photoreceptor cells (**Figure 4d**) and post synapse of secondary retinal neurons (data not shown).

We obtained adult dermal fibroblasts from three RP11 families carrying the heterozygous *PRPF31* frame-shift mutation c.1115_1125del¹⁷. Fibroblast lines from individuals either mildly or severely affected with RP had statistically significant decreases in the length and/or number of cilia, compared to an age-matched disease-control individual with age-related macular degeneration (ARMD) and healthy control individuals (**Figure 4e**). Furthermore, a *Caenorhabditis elegans* strain (MR247; see **On-line Methods**) containing a homozygous splice-site mutation (*rr40*) in the *PRPF8* orthologue (*prp-8*) previously associated with reduced spliceosome function in the germline¹⁸, had a partial ciliogenesis defect (**Figure 4f**). The microtubule singlet (A-tubule)-containing distal segments (DS) of at least 4 (of 10) amphid channel sensory cilia were truncated or missing entirely, and in those distal segments that remained, A-tubule numbers were abnormally low (**Figure 4f**). A few axonemes in the distal segment/middle segment (MS; microtubule doublets) boundary also had reduced microtubule numbers (**Figure 4f**).

Many of the human GPCRs identified as hits in our validation screen were neuroactive GPCRs. We therefore investigated the localisation of these proteins in a human ciliated neuroblastoma cell-line (SH-SY5Y) with neuronal-like characteristics that differentiate by extending neurites. We identified the localisation of most of these GPCRs (HTR1B, P2RY14, MAS1, OPRL1, and CRHR2) to the base of the cilium in differentiated SH-SY5Y cells (**Figure 5a, Suppl. Figure 1a**) and to the ciliary region of photoreceptors of murine retinas (**Figure 5b**). Higher resolution fluorescence and immunoelectron microscopy of a sub-set revealed localisation at the basal body and adjacent centriole of the connecting cilium, at the apical part of the connecting cilium (outer segment base), and in the axoneme of the outer segment of photoreceptor cells (**Figure 5c,d**). We confirmed the localisation of several GPCRs (SREB3, MAS1, P2RY14 and CRHR2) to the base of primary cilia in the marginal zone (layer I), sub-plate zone (layer IV) and ventricular zone (layer VI) of embryonic mouse neocortex (**Suppl. Figure 1b**). In knock-out mouse embryos mutated for the transition zone orphan receptor TMEM67 (meckelin)¹⁹, the ciliary localisation of these GPCRs was lost (**Suppl. Figure 1b**) suggesting that they require the structural or functional integrity of the transition zone to maintain their ciliary localisation. We hypothesised that ligand binding to GPCRs may promote the earliest stages of ciliogenesis. However, treatment of proliferating mIMCD3 cells and hTERT-RPE1 cells with a range of concentrations of antagonists against CRHR2, HTR1B and OPRL1 over 6-72 hours had no effect on ciliogenesis (data not shown). Inhibition of ligand binding was therefore not sufficient to inhibit cilia formation, suggesting that other mechanisms regulate ciliogenesis through GPCRs *e.g.* endocytic membrane trafficking²⁰.

Validated screen hits PIBF1 and C21orf2 predict new ciliopathy disease genes for Joubert syndrome and Jeune syndrome

We next investigated whether our list of validated ciliogenesis effectors could be used to prioritise predicted pathogenic variants identified from WES of ciliopathy patients including Joubert syndrome and Jeune syndrome. Joubert syndrome (JBTS; OMIM #213300)

represents a classic ciliopathy characterised by hypotonia, ataxia, cognitive impairment, and a distinctive brain malformation (the so-called “molar tooth sign”), with retinal dystrophy, cystic kidney disease, liver fibrosis and polydactyly occurring in subsets of patients²¹. Jeune asphyxiating thoracic dystrophy (JATD; Jeune syndrome, OMIM #611263) is a chondrodysplasia within the short-rib polydactyly syndrome ciliopathy spectrum, characterised by shortened limbs and ribs and a narrowed chest. Additional features include polydactyly, kidney cysts and renal failure, retinal degeneration and liver disease²².

A rare, homozygous missense variant (NM_006346.2: c.1910A>C, p.Asp637Ala) in *PIBF1* (also known as *C13orf24* or *CEP90*) was identified by WES in an affected JBTS individual of Schmiedeleut Hutterite descent (H1-3) and a similarly affected brother (H1-4; **Figure 6a-c, Suppl. Figure 2a-c, Suppl. Table 5**), in the absence of pathogenic mutations in known JBTS genes (**Suppl. Table 5**). The c.1910A>C variant segregated with the phenotype (**Suppl. Figure 2b**), was absent in the Exome Variant Server (EVS) and Exome Aggregation Consortium (ExAC) datasets, and was identified as homozygous in the affected individuals from a further three Hutterite families (H2, H3 and H4; **Suppl. Figure 2b**) suggesting a founder effect. Imaging findings for *PIBF1*-mutated individuals ranged from the classic molar tooth sign to moderate vermis hypoplasia with mildly thick superior cerebellar peduncles and characteristic superior cerebellar dysplasia (**Figure 6a-c, Suppl. Figure 3, Suppl. Table 6**). All individuals for whom information was available had ataxia and developmental delay, ranging from mild to moderate (**Suppl. Table 7**). Targeted molecular inversion probe sequencing (MIPS) of known JBTS genes and WES (**Suppl. Table 5**) identified a further seven families in which the affected individuals carried heterozygous truncating *PIBF1* variants (out of 643 additional families sequenced) in the absence of pathogenic mutations in known JBTS genes. Although exogenous expression of human wild-type *PIBF1* rescued ciliogenesis in mIMCD3 cells following siRNA knockdown of endogenous *Pibf1*, expression of *PIBF1* containing the p.Asp637Ala Hutterite variant was unable to rescue normal ciliogenesis suggesting that it is a pathogenic missense mutation (**Suppl. Figure 4**).

We next cross-compared WES data from patients with JATD to validate ciliogenesis candidate genes from the screen. Compound heterozygous missense changes in *C21orf2* (NM_004928.2: c.218G>C, p.Arg73Pro and c.671T>C, p.Leu224Pro) were identified in two affected siblings of non-consanguineous Caucasian northern European descent (UCL-111.1 and UCL111.2; **Suppl. Table 5, Suppl. Figure 5a**) with a clinical diagnosis of JATD. Compound heterozygous mutations (p.Leu161Serfs*9 and predicted intronic splice site mutation c.96+6T>A) were also found in a non-consanguineous Caucasian northern European individual UCL78.1 (**Suppl. Table 7, Suppl. Figure 5f**). All variants were either very rare or absent from the EVS and ExAC datasets, predicted damaging by Polyphen2 and/or Mutationtaster, and/or were non-conservative changes in highly conserved regions (see **On-line Methods** for further details). The three affected individuals presented with narrow thorax (**Figure 6d-g,i,j, Suppl. Table 7**) and pelvic bone malformation (**Figure 6h**) in skeletal radiological surveys. All developed retinal degeneration due to cone-rod dystrophy in childhood but none had impairment of renal function (**Figure 6k,l, Suppl. Table 7**). Further screening of *C21orf2* identified a homozygous missense change shared with the UCL-111 family (p.Arg73Pro) in individual GC4693.1 and four siblings, all of whom have cone-rod dystrophy, severe scoliosis and hip dysplasia (**Suppl. Figure 5b, Suppl. Table 7**). Sanger sequencing confirmed that all variants segregate with the disease (**Suppl. Figure 5a-f**). Exogenous expression of both the p.Arg73Pro and p.Leu224Pro variants in *C21orf2* partially rescued ciliogenesis in mIMCD3 cells following siRNA knockdown of endogenous *C21orf2*, suggesting that they are hypomorphic mutations (**Suppl. Figure 4**).

Homozygous mutations in *C21orf2* were also found in patients that initially presented with isolated cone-rod dystrophy (**Figure 6l**), consistent with a previous report that *C21orf2* is mutated in non-syndromic retinal degeneration²³. Re-assessment of these individuals revealed previously unidentified skeletal involvement in CR-F024.1 (c.103delA, p.Ile35Phefs*10; **Figure 6l,j, Suppl. Table 7, Suppl. Figure 5c**). These findings indicate that mutations in *C21orf2* (**Figure 6m**) cause both syndromic and non-syndromic forms of retinal dystrophy, with a variable spectrum of skeletal involvement.

The validated screen hit C21orf2 forms a ciliary functional module with NEK1 and SPATA7

C21orf2 (also known as LRRC76) is an LRR-containing protein conserved from *C. elegans* to human (**Figure 6m**). Although a small-scale siRNA screen suggested that knockdown impairs ciliogenesis and Hedgehog (Hh) signalling in mammalian cell cultures⁵, the exact role of this protein in ciliogenesis is unclear. C21orf2 localises to the basal body in mIMCD3 and hTERT-RPE1 cells and to the base of the connecting cilium in mouse photoreceptors (**Figure 3d, Suppl. Figure 6a**). Immunostaining confirms expression in tissues consistent with the phenotype of JATD, including developing ribs and hip joint in mouse embryos (**Suppl. Figure 6b**). Tandem affinity purification followed by mass spectrometry (MS) analysis identified a total of 88 unique proteins co-purifying with Strep-tag II/FLAG (SF-TAP) tagged C21orf2 in $n \geq 2/3$ experiments (**Suppl. Table 8**). Affinity proteomics (**Figure 7a**) and co-immunoprecipitation assays (**Figure 7b,c**) revealed that C21orf2, NEK1, and a retinal ciliopathy protein, SPATA7²⁴, are part of the same protein complex. Furthermore, co-immunoprecipitations with anti-NEK1 (**Suppl. Figure 7a**) followed by MS analysis of unique peptides (for three concurrent repeats of the experiment) showed that NEK1 pulled down endogenous C21orf2 from bovine retinal lysates (**Suppl. Table 9**). Immunofluorescence studies showed that C21orf2 co-localises with NEK1 and SPATA7 at the basal body in hTERT-RPE1 cells (**Suppl. Figure 6c**), and GST pull-down of SPATA7 from bovine retinal lysates efficiently recovered endogenous C21orf2 and NEK1 (**Suppl. Figure 7b, Suppl. Table 9**). Collectively, these data indicate the existence of a ciliary functional module that includes NEK1-C21orf2-SPATA7, which occurs endogenously in ciliated cells (**Suppl. Table 9**). Importantly, both the p.Arg73Pro and p.Leu224Pro mutations in C21orf2 abolished the interaction with NEK1 (**Figure 7c**).

The association of C21orf2 with NEK1 could provide an explanation for the variable JATD skeletal phenotype observed in *C21orf2*-mutated individuals (**Figure 6d-j**), since mutations in *NEK1* cause short-rib polydactyly syndrome type II (SRPS type II; Majewski

syndrome, OMIM #263520) in humans²⁵, a ciliary chondrodysplasia related to JATD. Furthermore, the association of C21orf2 with SPATA7 could explain the prominent retinal phenotype (which tends to be only an occasional feature of typical JATD) in individuals with C21orf2 mutations, since mutations in SPATA7 cause the retinal dystrophies Leber congenital amaurosis type 3 (LCA3) and juvenile RP (OMIM #604232). Knockdown of *nek1* in zebrafish using antisense morpholino oligonucleotides (MO; verified in **Suppl. Figure 7c**) produced classical ciliopathy phenotypes, with a ventrally curved body axis, hydrocephalus, otolith abnormalities and abnormal craniofacial cartilage development (**Figure 7d**). Knockdown of *nek1* also resulted in retinal defects including loss of photoreceptors (**Suppl. Figure 7fd-g**) and cilia length defects in the pronephros (**Suppl. Figure 7h,i**). Co-injection of human wild-type (WT) C21orf2 RNA with the *nek1* MO partially rescued these phenotypes (**Figure 7e**), consistent with the functional interaction between C21orf2 and NEK1. RNA expressing C21orf2 with the missense mutation p.Arg73Pro rescued the *nek1* morphant phenotype less effectively than wild-type, whereas C21orf2-p.Leu224Pro had little effect (**Figure 7e**), confirming the predicted hypomorphic effect of both variants observed *in vitro* (**Suppl. Figure 4**).

DISCUSSION

Here we present a whole genome siRNA-based functional genomics screen to identify genes whose activity is required for ciliogenesis and/or cilia maintenance. By interrogating every gene for a role in these processes, we obtained a global resource for investigation and interventions into the processes that are critical for the ciliary system. Our visual assay was primarily designed to test the presence or absence of cilia in the primary screen, and we provide all numerical features for this cellular phenotype as a functional resource to investigate aspects of cilia biology (**Suppl. Table 1**). In addition, we also imaged other cellular phenotypes during data acquisition (cilia intensity and nuclear morphology) and include these other numerical features (**Suppl. Table 1**) to enable the scientific community to interrogate and further annotate our screen data-set.

To minimise the false positive rate, we extensively characterised a widely used cellular model system (ciliated mIMCD3 cells) by arrayCGH, and our siRNA library was filtered for off-target effects, partial on-target effects and microRNA-like effects (**Figure 2c, Suppl. Table 1**). Using stringent filters and robust statistical cut-offs, we identified 194 genes which significantly affected ciliogenesis and were of particular functional interest, and took these on for validation by further knock down experiments in mIMCD3 and hTERT-RPE1 cells.

In total, we present a high confidence list of 68 validated genes that play significant roles in ciliogenesis in mIMCD3 cells, of which 37 were further found to affect ciliogenesis in hTERT-RPE1 cells (**Table 1, Suppl. Table 3**). A parallel secondary screen validated a further 44 genes encoding components of the UPS (**Suppl. Table 4**). Validated hits include *CEP120*, *KCNQ1*, several GPCRs (*CRHR2*, *HTR1B* and *OPRL1*) and PRPFs (*PRPF6*, *PRPF8*, *PRPF31* and *PRPF38A*) (**Figure 3a-b, Table 1, Suppl. Table 3**). The specificity of the screen for targeting mediators of ciliogenesis is demonstrated by the number of genes that have been independently implicated in this process. These include *CCDC41*¹², *OFD1*¹³, *CEP120*¹⁶ and *PIBF1*²⁶. *C21orf2* has been suggested to have a role in ciliogenesis from a previous small-scale siRNA screen⁵. In a complementary functional genomics study, *PLK4*, *CEP120*, *PRPF8* and *PRPF38A* have all been implicated in the process of centriolar under-duplication¹¹. Furthermore, *PLK4*, *OFD1*, *CDC27*, *PRPF6* and *PRPF8* were identified in the Mitochek screen as important effectors of cell division²⁷. Shared hits across multiple independent functional genomics studies indicate that our screen is a useful resource for the identification of proteins and functional modules that are essential for cilia biology and related cell biological processes.

This study also demonstrated the power of an unbiased screen to discover cellular or molecular processes that play a role in cilium biology. To illustrate this point, our screen unveiled neuroactive GPCRs as one of the functionally enriched groups of proteins affecting ciliogenesis (**Figure 3a, Table 1, Suppl. Figure 1**). GPCRs may play sensory or signalling roles in the ciliary vesicle or procilium at the earliest stages of ciliogenesis both in the

developing brain and in other tissues²⁸. In support of this hypothesis, adenylyl cyclase III co-localises with GPCRs at proximal ciliary regions in the developing neocortex, and over-expression of specific ciliary GPCRs, including the 5-HT₆ serotonin receptor, in cortical neurons causes cilia elongation²⁹.

Although the role of the PRPFs in the spliceosome is well-known and well-studied, we identified a group of seven splicing factors including PRPF6, PRPF8, PRPF31 and PRPF38a (**Table 1**) that are required for ciliogenesis. We demonstrate that PRPFs mutated in retinitis pigmentosa localise specifically to the base of the primary cilium in several different cell-types (**Figure 3d, Figure 4a**) and to the basal body complex and connecting cilium in photoreceptors (**Figure 4b,d**). We speculate that some of the splicing factors fulfil an additional ciliary function independent of their nuclear role in splicing. In support of this notion, two previous studies of the centrosomal proteome have suggested that splicing factors, including two of the splicing factors identified in our screen (PRPF6 and PRPF8), may be true centrosomal proteins^{30,31}. However, it seems less speculative to hypothesise that these proteins may be required for the correct splicing of an unidentified subset of genes that are involved in cilia formation. In support of the possible involvement of pre-mRNA processing factors in ciliogenesis, previous screens have identified that several splicing factor hits are important in related cellular processes of microtubule formation and regulation, namely centriolar biogenesis (PRPF8, PRPF38A)¹¹ and cell division (PRPF6, PRPF8)²⁷.

Several hits in our screen have a known function in the ubiquitin proteasome system (UPS), reflecting an importance of specific proteostasis mechanisms in mediating or regulating ciliogenesis which could be mediated through, for example, ubiquitination or SUMOylation of ciliary transcription factors or chaperoning of ciliary proteins to the base of the cilium. Remarkably, 5/7 of our splicing factor hits (LSM2, PRPF6, PRPF8, PRPF31 and USP39) are implicated in the ubiquitin-dependent regulation of the spliceosome³². It is interesting to note that the interaction of PRPF8 with ubiquitinated PRPF3 is regulated by the deubiquitinating enzyme USP4, and that loss of USP4 prevents the correct splicing of mRNAs including those for α -tubulin³². In support of this mechanism, we observe that a

viable *C. elegans* strain possessing a splice site mutation in *PRPF8/prp-8* has structural and numerical defects of axonemal microtubules in the primary cilia of amphid sensory neurons (**Figure 4f**). Thus, UPS and/or PRPF proteins could act as multifunctional “nexus molecules” that are involved in multiple aspects of proteostasis of ciliary proteins or their trafficking (for example, the ubiquitin-mediated internalisation and endosomal sorting of GPCRs). Alternatively, or additionally, these specific PRPFs could also ensure the correct splicing of transcripts encoding proteins important for ciliogenesis, including structural components of the cilium such as α -tubulin^{32,33}.

We also show the specificity and clinical utility of the screen as a tool for disease-gene discovery. When combined with large variant datasets, for example WES of ciliopathy patients, the functional data from our siRNA screen allowed the filtering and prioritisation of variants to identify pathogenic mutations. Although previous studies have indicated both *PIBF1* and *C21orf2* as functional candidate genes for ciliopathies^{5,23}, we formally demonstrate the utility and validity of a systems biology approach by the identification of mutations in these genes in JBTS and JATD patients, respectively (**Figure 6, Suppl. Table 7**). Furthermore, our affinity proteomics data provides an explanation of the mechanism for the variable phenotype in this form of JATD. We propose that the retinal phenotype observed in *C21orf2*-mutated patients may result from a dysfunctional SPATA7-C21orf2-containing module in the photoreceptors of these patients. Conversely, the skeletal phenotype observed in both *C21orf2*- and *NEK1*-mutated human patients probably has a common origin in the disruption of the C21orf2-NEK1 interaction. NEK1 is a serine-threonine kinase and two serines (positions 136 and 177) in C21orf2 have been previously found to be phosphorylated³⁴. In light of this, our affinity proteomics findings, and the incomplete rescue of NEK1 knockdown effects in zebrafish by C21orf2 overexpression, C21orf2 might represent a substrate of NEK1.

In conclusion, our findings highlight the utility of combining different systems biology approaches, namely high-content functional genomic screening, WES, and affinity proteomics. Our approach has identified ciliary roles for well-studied proteins, identified

disease-causing genes, allowed the refinement of patient phenotypes, and highlighted potential disease pathways that could provide deeper insights into cilium biology. This work demonstrates the utility of a systems biology approach in providing a compendium of candidate genes and offering insights into pathogenic mechanisms for an increasingly important group of Mendelian conditions, the ciliopathies.

Accession numbers: PIBF1: NM_006346.2 and NP_006337.2; C21orf2: NM_004928.2 and NP_004919.1. Whole exome sequencing data is available from the UK10k Consortium (UK10kCIL5236549 and UK10kCIL5236550) and dbGaP (phs000205.v5.p2)

Web Resources

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

University of California, Santa Cruz (UCSC) Genome Browser (<http://www.genome.ucsc.edu>)

Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (<http://evs.gs.washington.edu/EVS/>) date accessed Dec. 2014

Exome Aggregation Consortium (ExAC), Cambridge, MA, USA (<http://exac.broadinstitute.org>) date accessed Dec. 2014

The Consensus CDS (CCDS) project (<http://www.ncbi.nlm.nih.gov/projects/CCDS/>)

Copy Number Analyzer for Affymetrix GeneChip (<http://www.genome.umin.jp/>)

GraphPad, Quick Calcs (<http://www.graphpad.com/quickcalcs/chisquared2.cfm>)

Polyphen2 (<http://genetics.bwh.harvard.edu/pph2/>)

Mutationtaster (<http://www.mutationtaster.org/>)

Sorting intolerant from tolerant (SIFT) (<http://sift.jcvi.org/>)

Genomic evolutionary rate profiling (GERP)

(<http://mendel.stanford.edu/SidowLab/downloads/gerp>)

PhyloP (<http://genome.ucsc.edu>)

Alamut (<http://www.interactive-biosoftware.com/software/alamut/features>)

Acknowledgements

Our deepest thanks go to all of the families and individuals with Joubert syndrome, Jeune syndrome and retinitis pigmentosa. We thank Paul Barker, Martin Jackson, Kieran Roberts and Amanda Jones from PerkinElmer Inc. for technical support. The research received funding from the European Community's Seventh Framework Programme FP7/2009 under grant agreement no: 241955 SYSCILIA towards C.A.J., R.R., M.U., T.G., P.L.B, O.E.B, U.W., R.H.G., M.A.H. and H.O. This work was supported by a Sir Jules Thorn Award for Biomedical

Research (JTA/09, to C.F.I. and C.A.J.) and a Medical Research Council grant (MR/K011154/1, to C.A.J.) M.S. was funded by an Action Medical Research Clinical Training Fellowship (RTF-1411), a Radboud University Excellence fellowship, a Radboud UMC Hypatia Tenure Track fellowship and acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG; grant SFB 1140 (KIDGEM)). R.R. is funded by the Netherlands Organization for Scientific Research (NWO Vici-865.12.005), and by the Foundation Fighting Blindness (C-CMM-0811-0546-RAD02). P.L.B. is a Wellcome Trust Senior Research Fellow, and P.L.B., M.S., R.H.G. and R.R. acknowledge funding from the Dutch Kidney Foundation (CP11.18, "Kouncil"). C.T.T. is funded by DFG grant TH 896/3-3 and IZKF (Interdisciplinary Centre for Clinical Research of the Universität of Erlangen-Nürnberg) grant F4. U.W. is funded by the FAUN foundation, Nuernberg. Z.A.A. and W.H. are supported by grants from the Rosetree's Trust (A210 and A465). H.M.M. is supported by the Great Ormond Street Hospital Children's Charity and received grants from the Milena Carvajal Pro-Kartagener Foundation and Action Medical Research (GN2101). H.R. is funded by the Canadian Institutes of Health Research (CIHR) Training Program in Genetics, Child Development and Health, Alberta Children's Hospital Research Institute (ACHRI) and Alberta Children's Hospital Foundation. A.R.W. is funded by RP Fighting Blindness, Moorfields Eye Hospital Special Trustees, the National Institute for Health Research (NIHR; Moorfields Eye Hospital and Institute of Ophthalmology, London, U.K.) and the Foundation Fighting Blindness (U.S.) D.D. is funded by NINDS grant R01NS064077, the University of Washington Intellectual and Developmental Disabilities Research Center Genetics Core P30HD002274, and private donations from families of individuals with Joubert syndrome. J.S. is funded by NCI grant R21CA160080. Access to the B6;129P2-*Tmem67*^{tm1Dgen/H} line was funded by the Wellcome Trust Knock-out Mouse Resource scheme (C.A.J. and C.F.I.; grant ME041596). The UK10K project is funded by the Wellcome Trust under grant agreement WT091310. A full list of UK10k investigators can be accessed at <http://www.uk10k.org>. The University of Washington Center for Mendelian Genomics is supported by NHGRI grant U54HG006493.

Competing financial interests

The authors declare that that they do not have competing financial interests.

Author contributions

K.S., G.W., S.N., W.H., M.A. and J.H. performed the primary and secondary siRNA screens. K.S., G.W., D.A.M., T-M.T.N. and S.J.F.L. performed and analysed the tertiary and other validation screens. L.W. wrote scripts for UPS validation screen analysis. C.T., S.M.B., J.B., E.E.M., D.C.T. and C.F.I supervised and managed the BioScreening Technology Group and edited the manuscript. S.N. did qRT-PCR analyses, and K.S., G.W. and W.H. did western blot analyses in validation screens. G.W. and Z.A.A. performed and interpreted IF and IHC staining of mouse tissue sections. D.A.P. wrote Perl scripts for automated analysis of siRNA screen data and re-analysed UK10k WES data. K.S. and G.W. performed confocal IF microscopy of primary fibroblasts and cell-lines. G.T. and T.G. developed software and analysed array CGH data, and G.T. analysed siRNA sequences. G.G.S. and R.H.G. performed 3D spheroid cultures of mIMCD3 cells. J.K. and O.B. performed and interpreted all experiments with *C. elegans*. N.S., U.W., and K.A.W. analysed mouse photoreceptor cells by high magnification IF and immunoelectron microscopy. K.K., A.G. and C.T.T. developed and characterised the NEK1 antibody. H.R., C.L.B., P.G., R.T.P., R.L. and F.P.B ascertained and recruited subjects, performed linkage analysis, WES and Sanger sequencing, and analysed data for Hutterite participants. J.S.P. and A.M.I. diagnosed and enrolled participants, collated clinical documentation and supervised research on Hutterite participants. C.O. analysed and provided data for Schmiedeleut controls. R.H. analysed and provided data for Dariusleut and Leherleut controls. L.H. and K.M.B. provided reagents for work on PIBF1. I.G.P. analysed the Joubert syndrome exomes, designed the MIPS capture and performed the MIPS sequencing and analysis. J.S. conceived of the MIPS method and E.A.B. contributed to MIPS method development. D.D. diagnosed and enrolled participants, collated clinical documentation, supervised the sequencing and edited the manuscript. A.M,

B.N.C, A.E.C., C.E.W., K.M.B., P.L.B., F.S.A., M.M., C.T., C.F.I, H.O., C.W., F.S., S.A.H. and P.F. diagnosed and ascertained participants and collated clinical information. A.R.W. diagnosed and ascertained participants, collated clinical information and analysed SNP genotyping data. P.I.S. collated clinical information, prepared samples for WES and analysed WES data. M.A. and S.A. performed RT-PCR experiments. T-M.T.N., D.A.M. and J.v.R. performed and analysed C21orf2 PPI experiments and IF microscopy. N.H., K.B. and M.U. performed, analysed and supervised TAP MS experiments. M.S. performed WES analysis, zebrafish experiments, performed and supervised Sanger sequencing analysis and analysed the data. M.S. and H.M.M. conceived the human genetics and zebrafish experiments and wrote the manuscript. R.R. and C.A.J. conceived the study, directed and supervised the research, analysed and collated data, and wrote the manuscript.

REFERENCES

1. Goetz, S.C. & Anderson, K.V. The primary cilium: a signalling centre during vertebrate development. *Nat Rev Genet* **11**, 331-44 (2010).
2. Adams, M., Smith, U.M., Logan, C.V. & Johnson, C.A. Recent advances in the molecular pathology, cell biology and genetics of ciliopathies. *Journal of Medical Genetics* **45**, 257-267 (2008).
3. Quinlan, R.J., Tobin, J.L. & Beales, P.L. Modeling ciliopathies: Primary cilia in development and disease. *Curr Top Dev Biol* **84**, 249-310 (2008).
4. Kim, J. *et al.* Functional genomic screen for modulators of ciliogenesis and cilium length. *Nature* **464**, 1048-51 (2010).
5. Lai, C.K. *et al.* Functional characterization of putative cilia genes by high-content analysis. *Mol Biol Cell* **22**, 1104-19 (2011).
6. Birmingham, A. *et al.* Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods* **6**, 569-75 (2009).
7. Zhang, X.D. A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. *Genomics* **89**, 552-61 (2007).
8. van Dam, T.J. *et al.* The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. *Cilia* **2**, 7 (2013).
9. Huang da, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).
10. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
11. Balestra, F.R., Strnad, P., Fluckiger, I. & Gonczy, P. Discovering regulators of centriole biogenesis through siRNA-based functional genomics in human cells. *Dev Cell* **25**, 555-71 (2013).
12. Joo, K. *et al.* CCDC41 is required for ciliary vesicle docking to the mother centriole. *Proc Natl Acad Sci U S A* **110**, 5987-92 (2013).
13. Ferrante, M.I. *et al.* Oral-facial-digital type I protein is required for primary cilia formation and left-right axis specification. *Nat Genet* **38**, 112-7 (2006).
14. Martin, C.A., *et al.* Mutations in PLK4, encoding a master regulator of centriole biogenesis, cause microcephaly, growth failure and retinopathy. *Nat Genet* **46**, 1283-92 (2014).

15. Shaheen, R., Al Tala, S., Almoisheer, A. & Alkuraya, F.S. Mutation in PLK4, encoding a master regulator of centriole formation, defines a novel locus for primordial dwarfism. *J Med Genet* **51**, 814-6 (2014).
16. Shaheen, R. *et al.* A founder CEP120 mutation in Jeune asphyxiating thoracic dystrophy expands the role of centriolar proteins in skeletal ciliopathies. *Hum Mol Genet* **24**, 1410-9 (2015).
17. Vithana, E.N. *et al.* A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol Cell* **8**, 375-81 (2001).
18. Hebeisen, M., Drysdale, J. & Roy, R. Suppressors of the cdc-25.1(gf)-associated intestinal hyperplasia reveal important maternal roles for prp-8 and a subset of splicing factors in *C. elegans*. *RNA* **14**, 2618-33 (2008).
19. Abdelhamed, Z.A., Wheway, G., Szymanska, K., Natarajan, S., Toomes, C., Inglehearn, C. & Johnson, C.A. Variable expressivity of ciliopathy neurological phenotypes that encompass Meckel-Gruber syndrome and Joubert syndrome is caused by complex de-regulated ciliogenesis, Shh and Wnt signalling defects. *Hum Mol Genet* **22**, 1358-72 (2013).
20. Mukhopadhyay, S. *et al.* The ciliary G-protein-coupled receptor Gpr161 negatively regulates the Sonic hedgehog pathway via cAMP signaling. *Cell* **152**, 210-23 (2013).
21. Romani, M., Micalizzi, A. & Valente, E.M. Joubert syndrome: congenital cerebellar ataxia with the molar tooth. *Lancet Neurol* **12**, 894-905 (2013).
22. Huber, C. & Cormier-Daire, V. Ciliary disorder of the skeleton. *Am J Med Genet C Semin Med Genet* **160C**, 165-74 (2012).
23. Abu-Safieh, L. *et al.* Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. *Genome Res* **23**, 236-47 (2013).
24. Eblimit, A., *et al.* Spata7 is a retinal ciliopathy gene critical for correct RPGRIP1 localization and protein trafficking in the retina. *Hum Mol Genet* in press (2014) pii: ddu573.
25. Thiel, C. *et al.* NEK1 mutations cause short-rib polydactyly syndrome type Majewski. *Am J Hum Genet* **88**, 106-14 (2011).
26. Kim, K., Lee, K. & Rhee, K. CEP90 is required for the assembly and centrosomal accumulation of centriolar satellites, which is essential for primary cilia formation. *PLoS One* **7**, e48196 (2012).
27. Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464**, 721-7 (2010).

28. Arellano, J.I., Guadiana, S.M., Breunig, J.J., Rakic, P. & Sarkisian, M.R. Development and distribution of neuronal cilia in mouse neocortex. *J Comp Neurol* **520**, 848-73 (2012).
29. Guadiana, S.M. *et al.* Arborization of dendrites by developing neocortical neurons is dependent on primary cilia and type 3 adenylyl cyclase. *J Neurosci* **33**, 2626-38 (2013).
30. Andersen, J.S. *et al.* Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**, 570-4 (2003).
31. Jakobsen, L. *et al.* Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. *EMBO J* **30**, 1520-35 (2011).
32. Song, E.J. *et al.* The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. *Genes Dev* **24**, 1434-47 (2010).
33. Pelisch, F., Risso, G. & Srebrow, A. RNA metabolism and ubiquitin/ubiquitin-like modifications collide. *Brief Funct Genomics* **12**, 66-71 (2013).
34. Oppermann, F.S. *et al.* Large-scale proteomics analysis of the human kinome. *Mol Cell Proteomics* **8**, 1751-64 (2009).

FIGURE LEGENDS

Figure 1. Automated high content imaging, validation of siRNA screen controls and calculation of cut-off values.

(a) Schematic of a polarised mIMCD3 cell showing the focal planes used to image nuclei (blue), cytoplasm (pink) and ciliary axonemes (green). **(b)** mIMCD3 cells imaged using an Operetta high-content imaging system, with representative images from Harmony/Columbus software of cilia recognition (“find spots”). Scale bar = 10 μm . **(c)** Upper panels: significant effect on ciliogenesis following reverse transfection with positive control siRNA pool against *Ift88* compared to non-targeting scrambled siRNA, imaged by immunostaining for acetylated (Ac) α -tubulin (green). Bar graph (green) quantitates the effect on ciliogenesis (% cells with a single cilium) for positive controls (*Plk1*, *Mks1*, *Rpgrip11*, *Ift88*) and negative (-ve) controls including *MLNR*, scrambled (scr.) siRNAs and mock transfection. Lower panels: positive transfection control siRNA pool against *Plk1* has a significant effect on cell number. Images show cells stained for acetylated α -tubulin (green), TOTO3 (pink) and DAPI (blue). Bar graph (blue) quantitates the effect on cell number following knock-down with control siRNAs. Scale bar = 20 μm . Significance of pairwise comparisons with all negative controls (#): ns, not significant; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ (paired two-tailed Student’s t-test), for $n = 32$ independent replicates. **(d)** Bar graphs showing mean robust z score for % cells with a single cilium (Z_{cilia}) and mean robust z scores for cell number (Z_{cell}) for positive and negative siRNA controls. **(e)** Strictly standardised mean difference (SSMD) values for each replicate per batch in the screen. The average SSMD value for all batches was 1.717. Each bar is colour-coded by batch for sub-libraries in the ThermoFisher mouse siGENOME library (m, mouse; GPCR, G protein-coupled receptor; UC, ubiquitin conjugation; IC, ion channel; PP, protein phosphatase; P, protease; PK, protein kinase) or set of 10 plates (DT, druggable targets; G, genome libraries). **(f)** Robust z score cut-off values for cilia number ($Z_{\text{cilia cutoff}}$) calculated for each replicate per batch with error bars indicating median absolute deviation, where $Z_{\text{cilia cutoff}}$ is the median z score for all positive controls. Dots are colour-coded by batch for sub-libraries as in (e).

Figure 2. Analysis and data filtering strategy for the whole genome siRNA screen. (a) Scatter plot of z_{cilia} values of two screen replicates (rep. 1 and rep. 2). Each dot represents one targeted gene in the Thermo Fisher mouse siGENOME library, colour-coded by batch for either sub-library (m, mouse; GPCR, G protein-coupled receptor; UC, ubiquitin conjugation; IC, ion channel; PP, protein phosphatase; P, protease; PK, protein kinase) or set of 10 plates (DT, druggable targets; G, genome libraries). The median Pearson's correlation coefficient (R^2) between replicates was 0.71, if failed measurements in batch mDT51-60 were excluded (grey region). **(b)** Scatter plot of mean z_{cilia} values plotted against robust z scores for cell number (z_{cell}) for each targeted gene in the screen with each batch colour-coded as for (a); candidate hit genes are indicated. **(c)** Schematic to summarise data filtering steps in the primary, secondary and tertiary screens (complete datasets shown in **Suppl. Tables 1-3**, respectively). The number of genes filtered at each step are shown in bold, with the filter applied highlighted in grey boxes (see main text and **On-line Methods** for further details). Validated hits are listed in **Table 1**. **(d)** Classification enrichment for selected GO terms, or KEGG and BioCarta pathways/processes as determined using DAVID. Left bar graph: $n=1829$ screen hits with a significant effect on cilia number ($z_{\text{cilia}} < z_{\text{cilia cutoff}}$), without filtering for any effect on cell number, that had a human homologue. Right bar graph: 154 filtered screen hits with a human homologue that showed a significant effect on cilia number ($z_{\text{cilia}} < z_{\text{cilia cutoff}}$) but no significant effect on cell number ($z_{\text{cell}} > z_{\text{cell cutoff}}$), where $z_{\text{cell cutoff}}$ is the median z - 2MAD of all negative controls. The significance for terms or pathways was determined by a Benjamini-Hochberg q threshold value <0.05 (red dotted line) to correct for the false discovery rate, with the number of associated human genes shown above each bar.

Figure 3. Validation screens of ciliogenesis genes. (a) Robust z scores for % cells with a single cilium (z_{cilia}) for selected validated genes from the secondary screen, each assayed in two replicates with four individual siRNAs in mouse mIMCD3 cells. Error bars indicate the range in values for the replicates. Cut-off value of $z_{\text{cilia cutoff}} = -1.34$ is indicated (red dotted

line), calculated as the median z of positive controls. Colours indicate selected functional annotations. **(b)** Bar graphs for quantitative PCR values for knock-down of the indicated target genes, for all assays when negative control (siScr.) cycle threshold (Ct) values <30. Transcript levels expressed as relative quantities, with the statistical significance of the indicated pair-wise comparisons: * $p < 0.05$, ** $p < 0.01$ and **** $p < 0.0001$ (unpaired two-tailed Student's t -test), for $n=3$ independent replicates. Error bars indicate s.d. **(c)** Western immunoblots (IB) showing knockdown of expression for the indicated proteins. Loading control is β -actin (act.), with the ratio of band intensities shown below. **(d)** IF confocal microscopy of mIMCD3 cells knocked-down with siRNAs for the indicated validated genes from the screen and immunostained for the cognate protein (green) and acetylated (Ac) α -tubulin (red). Magnified insets for selected cells (white arrowheads) are shown in white frames. All knockdowns cause loss of ciliary-associated protein (yellow arrowheads) and cilia. Scale bars = $10\mu\text{m}$. **(e)** Knock-down of *Plk4* in three-dimensional spheroid cultures of mIMCD3 cells causes statistically significant ciliogenesis defects (visualised by acetylated (Ac) α -tubulin in white); ** $p < 0.01$ (paired two-tailed Student's t -test), for $n=3$ independent experiments. The number of cells analysed were control: 204, 195 and 190; and *Plk4* knockdown: 198, 106 and 212, respectively, for each experiment. Error bars indicate s.d. Scale bars = $10\mu\text{m}$.

Figure 4. Ciliary localisation and functional effect on ciliary axonemal formation of pre-mRNA processing factors. **(a)** PRPF6, PRPF8 and PRPF31 (green) localise to proximal/basal regions of primary cilia (green arrowheads) and nuclear speckles in the indicated cell-lines. Selected cells (white arrowheads) are magnified (insets). **(b)** PRPF6, PRPF8 and PRPF31 (green) localise to the ciliary regions of photoreceptor cells (connecting cilium, CC) and nuclei of inner nuclear layer (INL) in longitudinal sections of adult murine retinas. Retinal layers are depicted schematically: photoreceptor outer segment (OS), connecting cilium (CC), and inner segment (IS); secondary neurons (2^{nd}); ganglion cell layer (GC); synaptic region (S) in the outer and inner plexiform layer (OPL and IPL). **(c)**

Immunoelectron microscopy confirms localisation of PRPF6 and PRPF8 in nuclear speckles in the INL (arrowheads). Frames indicate the magnified insets. **(d)** Immunofluorescence of PRPFs (green) and the ciliary marker centrin-3 (red), and immunoelectron microscopy, reveal localisation of PRPF6 and PRPF8 at the basal body (BB) and adjacent centriole (CE) of the CC (arrowheads) and apical CC (arrow). **(e)** Primary cilia length and number measurements for dermal fibroblasts from normal healthy controls, age-matched disease-control (ARMD), and four patients with RP type 11 of variable severity carrying heterozygous (+/-) *PRPF31* frame-shift mutation c.1115_1125delGAAGCAGGCCA. Significance of pairwise comparisons with the disease negative control (#): ns, not significant; * $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$ (paired two-tailed Student's t-test), for $n = 4$ independent experiments. Error bars indicate s.d. **(f)** Transmission EM images of sensory cilia from sequential cross sections (numbered arrows in schematics) of *C. elegans* amphid pore in wild-type (N2) and *prp-8(rr40)* mutants. Wild-type amphid pores contain 10 ciliary axonemes, each consisting of a distal segment (DS; singlet A microtubules), middle segment (MS; doublet A/B microtubules), transition zone (TZ) and periciliary membrane compartment (PCMC). In *prp-8(rr40)* mutants, axonemes are missing in DS and DS/MS boundary, and microtubule (MT) number is reduced. Images are representative of four analysed amphid pores for each strain. Scale bars: (a) 20 μm and 5 μm ; (b) 50 μm ; (c) 800 nm; (d) 1 μm and 0.4 μm ; (f) 200 nm (low magnification images), 100 nm (high magnification images).

Figure 5. Ciliary localisation of G protein-coupled receptors. **(a)** Localisation of selected indicated GPCRs (green) to proximal or basal regions of primary cilia (polyglutamylated α -tubulin; red) in differentiated SH-SY5Y neuronal cells. Magnified insets for selected cells (white arrowheads) are shown in white frames. Scale bar = 10 μm . **(b)** Immunofluorescence (IF) labelling of MAS1 and OPRL1 (green) in the ciliary regions of photoreceptor cells (CR) (red; yellow arrowheads in magnified insets) in adult mouse retina. Abbreviations: GCL, ganglion cell layer; INL, inner nuclear layer; OPL, outer plexiform layer; ONL, outer nuclear layer. Scale bar = 50 μm **(c)** High magnification immunofluorescence for GPCRs (green) and

centrin-3 (red) reveals localisation of GPR20, SREB3, and MAS1 at the adjacent centriole and the basal body (arrowheads) of the connecting cilium. Scale bars = 1µm **(d)** IEM reveals SREB3, MAS1, and HTR1B at the basal body (BB) of the connecting cilium (CC) (arrowheads). Immunoelectron microscopy of MAS1 shows additional labelling throughout the CC and at the ciliary tip (arrow; two different images of photoreceptors are merged at the dashed line). Abbreviations: CE, centriole; IS, inner segment; OS, outer segment. Scale bars = 400nm.

Figure 6. Clinical features of ciliopathy patients with mutations in validated hits *PIBF1* or *C21orf2*. **(a-c)** Brain MRI findings for *PIBF1*-related Joubert syndrome in individual H1-4. Molar tooth sign with moderate vermis hypoplasia, elevated and thickened superior cerebellar peduncles (arrowhead), and superior cerebellar dysplasia (arrow) indicated. **(a)** is a T1-weighted and **(b-c)** are T2-weighted images. **(d-l)** Clinical features of individuals with *C21orf2* mutations including narrow and deformed thorax in UCL 78.1 **(d-f)**; child), UCL-111.1 **(g)**; child) and CR-F024.1 **(i-j)**; adult). UCL78.1 presented with a slightly dysplastic pelvis **(h)** in later childhood. Examples of funduscopy images revealing mild pigmentary depositions and some mottling shown for CR-F024.1 **(k)** and 111.2 **(l)**; both patients were clinically blind at the time of examination. **(m)** Upper panel: schematic of the *PIBF1* protein (NP_006337.2) showing the approximate positions of coiled coil domains (green boxes) and the p.Asp637Ala mutation. Lower panel: schematic of the *C21orf2* protein (NP_004919.1) showing the positions of three leucine-rich repeats (LRR1-3, blue boxes), an LRRCT motif occurring C-terminal to LRRs (green box) and mutations.

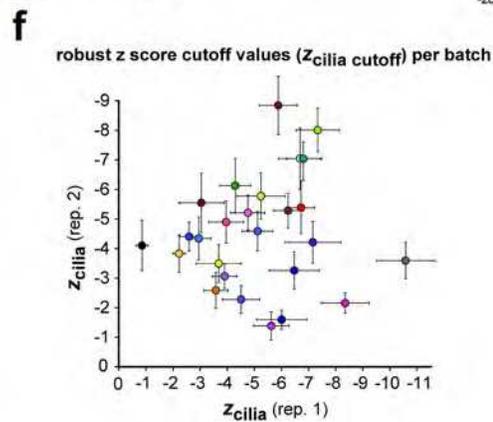
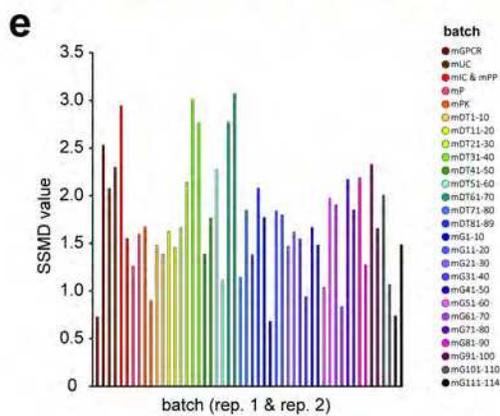
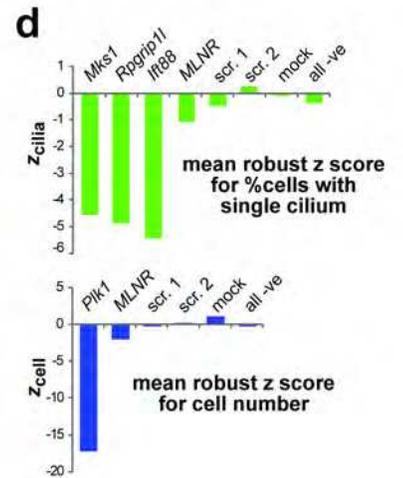
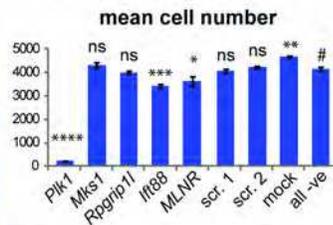
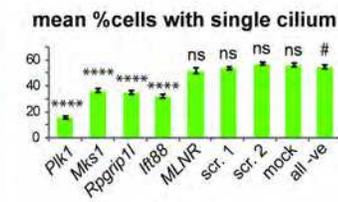
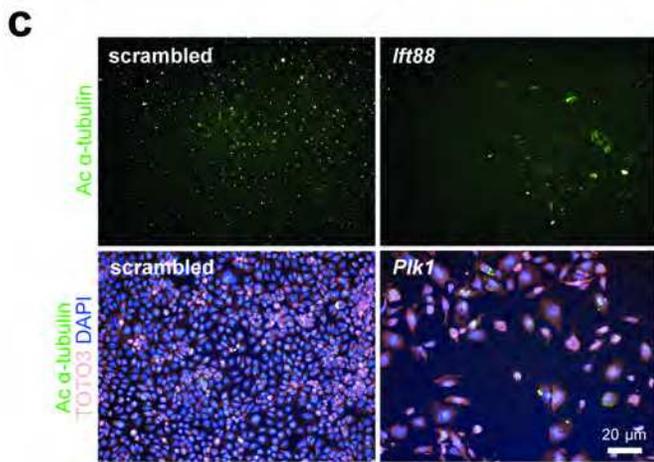
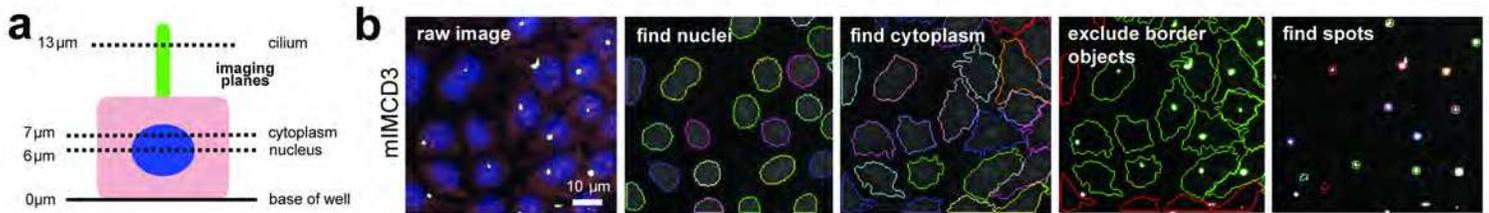
Figure 7. Validated hit *C21orf2* forms a ciliary functional module with *NEK1* and *SPATA7*. **(a)** Interactions between *C21orf2*, *NEK1* and *SPATA7* were identified by tandem affinity purification. The values indicate sequence coverage of identified prey peptides (horizontal) co-purified with specific bait proteins (vertical); see **Suppl. Table 8** for details. **(b)** Reciprocal coimmunoprecipitation of HA-*C21orf2* and both NTAP-*NEK1*, and NTAP-*SPATA7* (5% input: panels 1 and 2; immunoprecipitation (IP) assays in panels 3 to 6), but not the

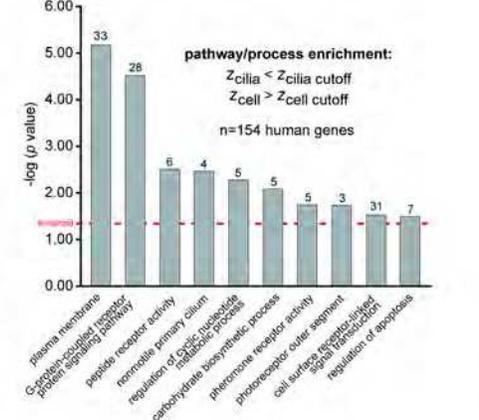
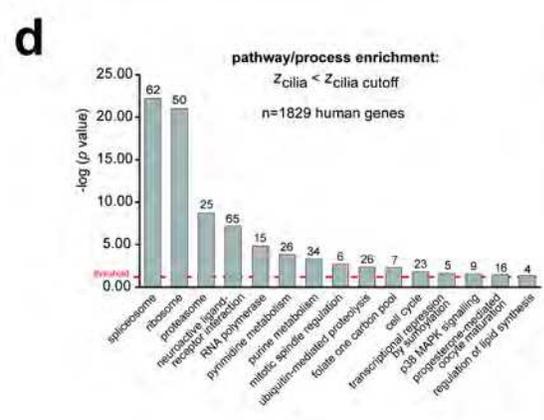
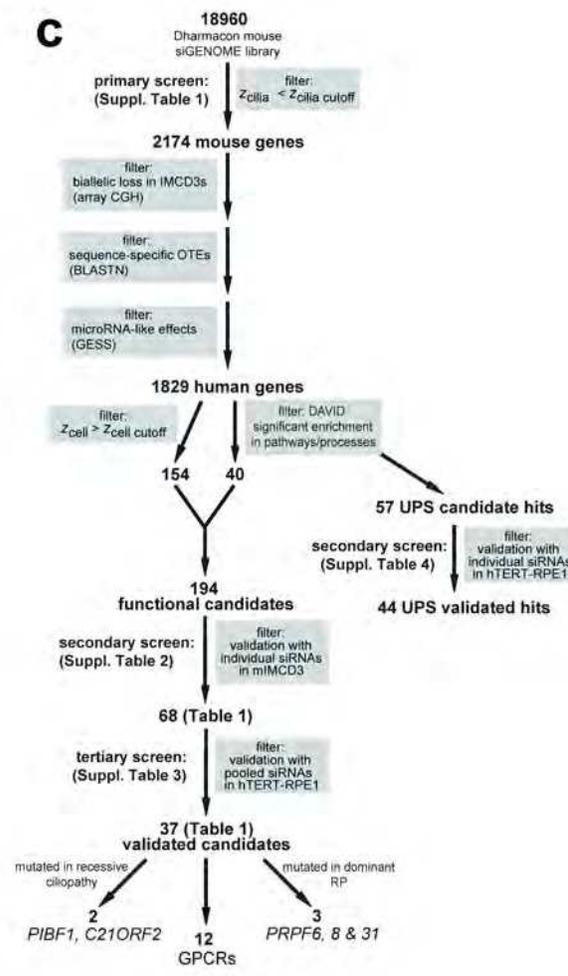
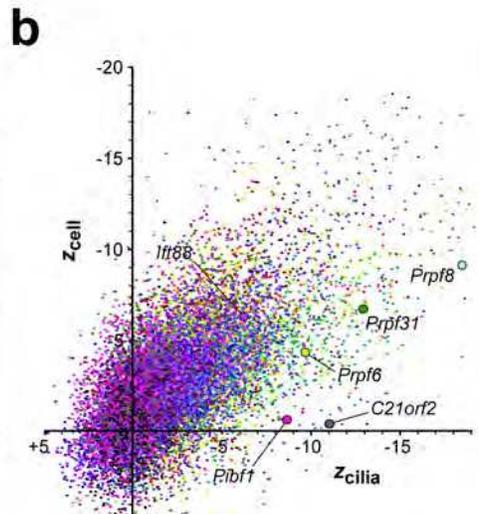
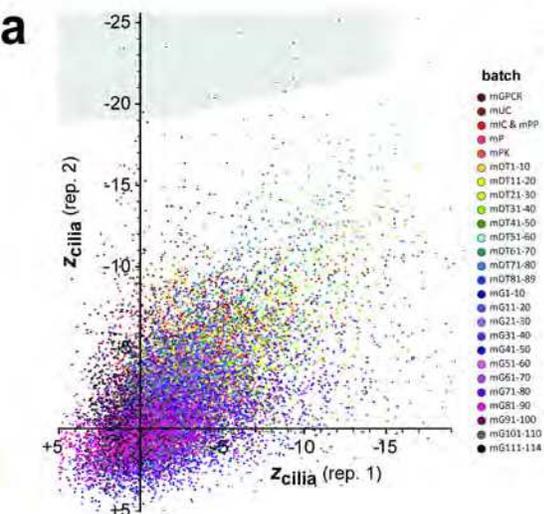
negative control HA-LRRK2 (LRR; panels 4 and 6: lanes 2 and 4) following immunoblot (IB) detection of either the HA or FLAG (Strep-tag II-FLAG, NTAP) epitopes. **(c)** FLAG-NEK1 coimmunoprecipitates wild-type (WT) C21orf2 (panel 4, lane 1; indicated by arrow), but not p.Arg73Pro or p.Leu224Pro mutant C21orf2 (panel 4, lanes 2 & 3) or the negative control HA-LRRK2 (LRR; lanes 4 to 6). Positive control co-IP (NPHP4-RPGRIP1L) is in lane 7, 10% input is shown in panels 1 and 2 and IP assays in panels 3 to 6. IgG heavy chain (HC) is indicated in panel 4. **(d)** Antisense morpholino oligonucleotide (MO1) injection against zebrafish *nek1* resulted in ciliopathy phenotypes (ventrally curved body axis, hydrocephalus, smaller eyes and otolith defects at 3-4 dpf; upper and middle panels, otoliths indicated by arrowheads; scale bars = 500 or 50 μ m, respectively). Lower panels: cartilage visualisation using Alcian Blue staining 4.5 dpf revealed impaired craniofacial cartilage development in *nek1* morphants compared to control MO-injected embryos. These defects resembled, but were less severe, than those for zebrafish *smo* mutant embryos. Scale bar = 200 μ m. **(e)** Co-injection of 1.5ng *nek1* MO1 with 150pg human wild-type (WT) *C21orf2* RNA (three independent experiments, total n=76 embryos) caused significant but incomplete phenotypic rescue, whereas RNA expressing the *C21orf2* missense mutation p.Arg73Pro (C21orf2-p.Arg73Pro; three experiments, total n=72 embryos) rescued less effectively compared to negative control (#). Co-injection of 150pg C21orf2-p.Leu224Pro (three experiments, total n=91 embryos) had a marginal effect on rescue. Significance of pair-wise comparisons with the negative *nek1* MO1-only control (#), or other comparisons indicated by braces, are: n.s., not significant; * p<0.05, ** p<0.01, **** p<0.0001 (unpaired two-tailed Student's t-test).

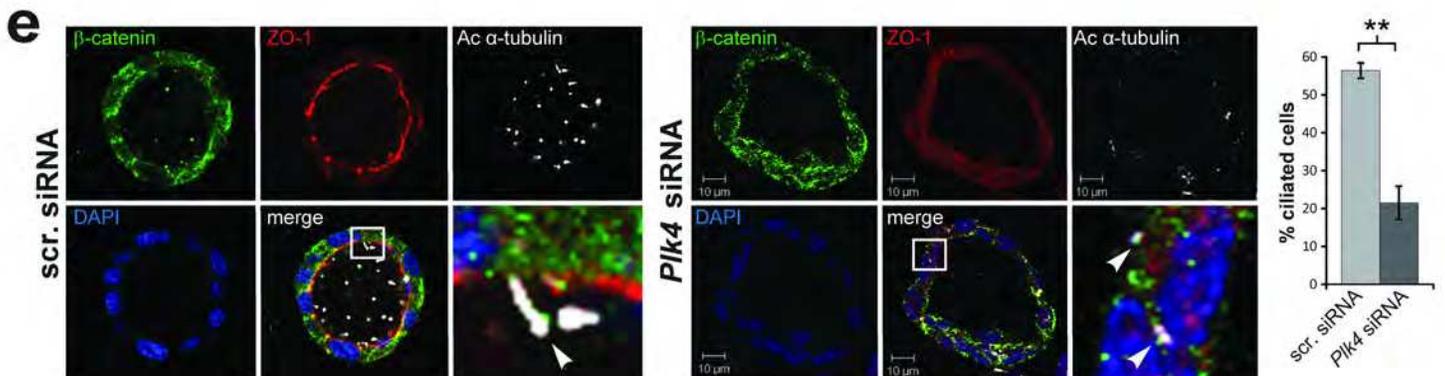
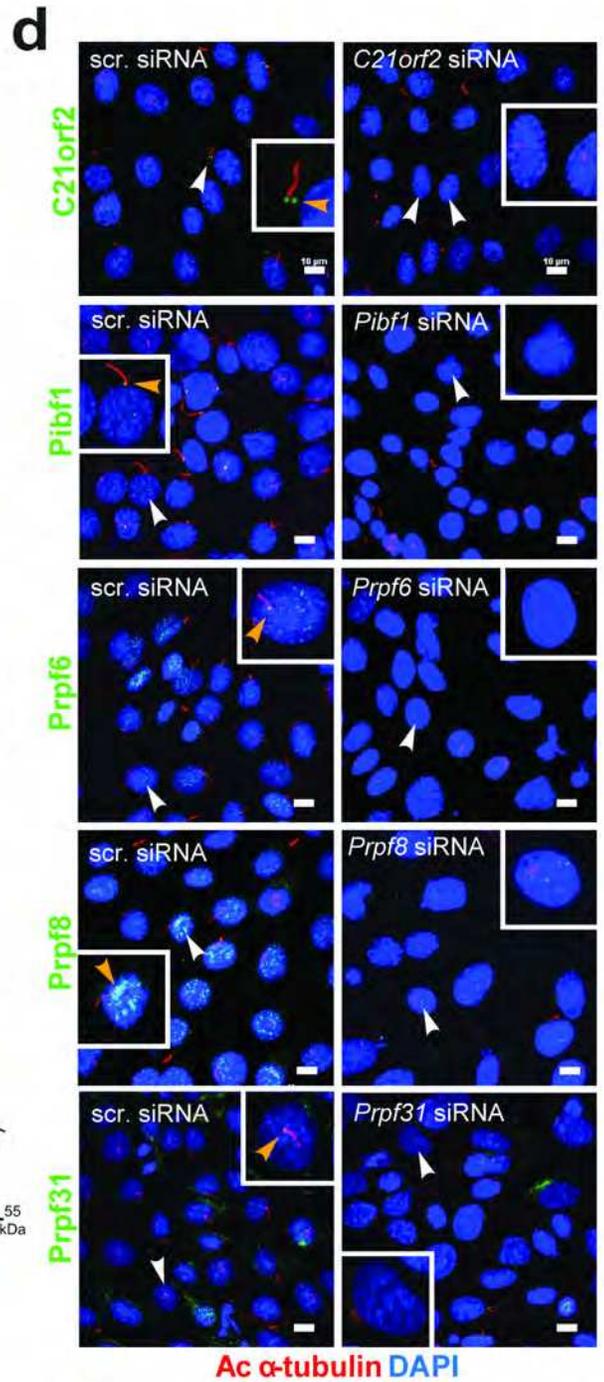
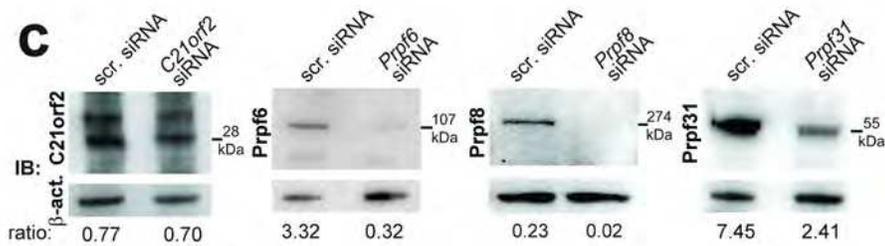
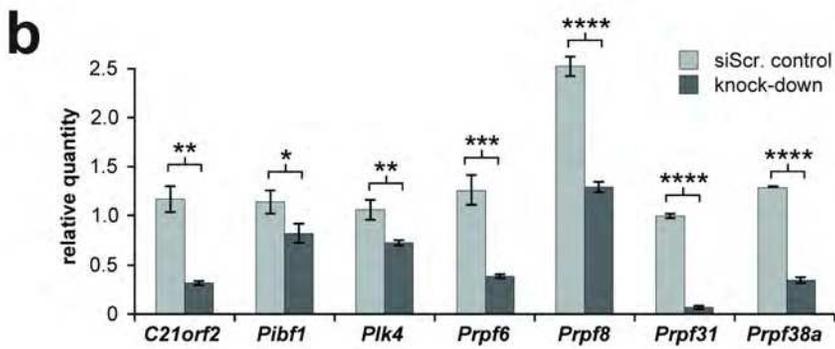
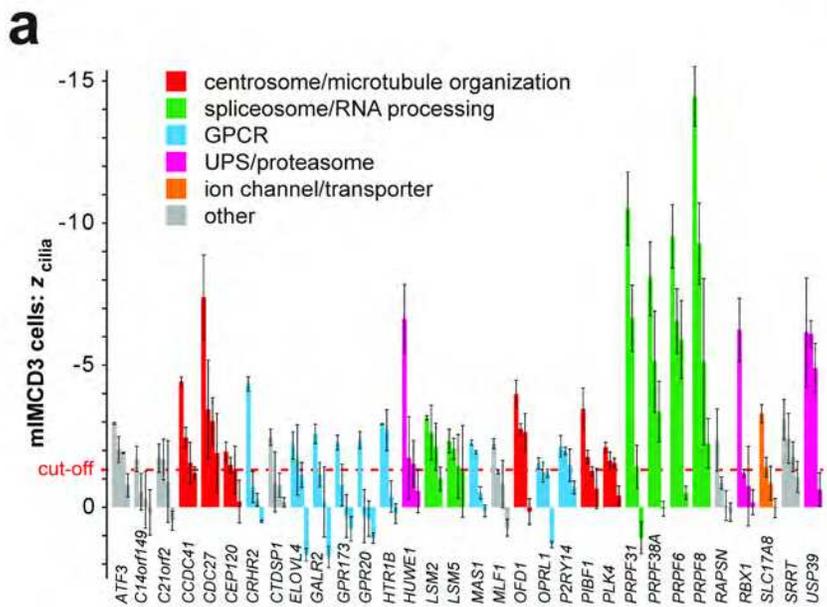
TABLE LEGENDS

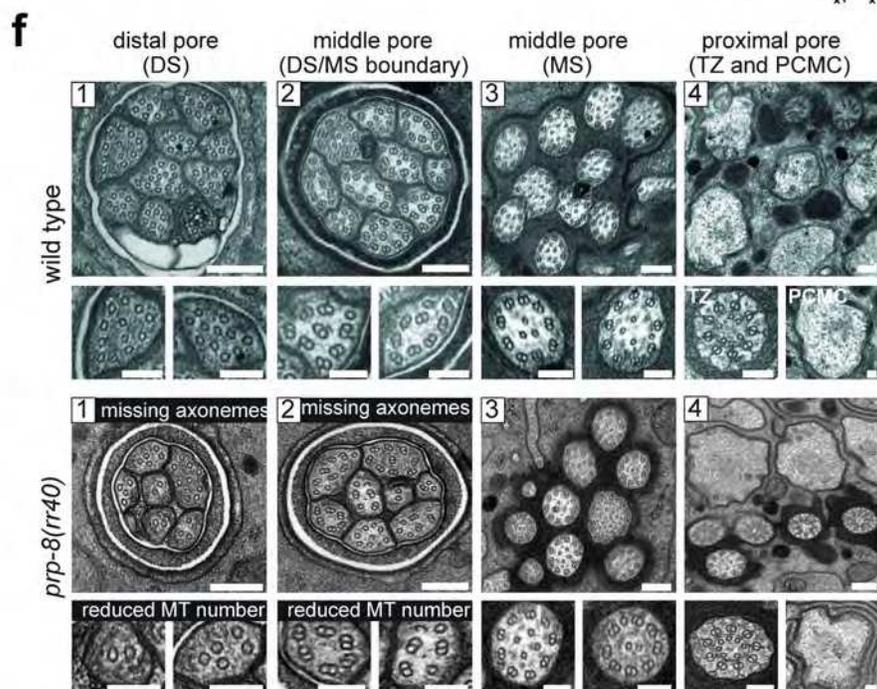
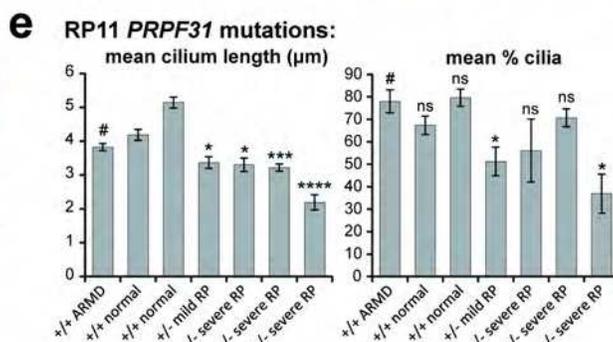
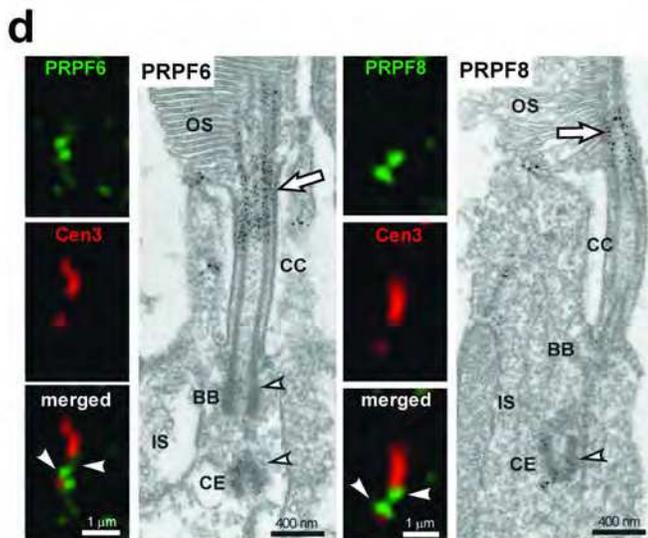
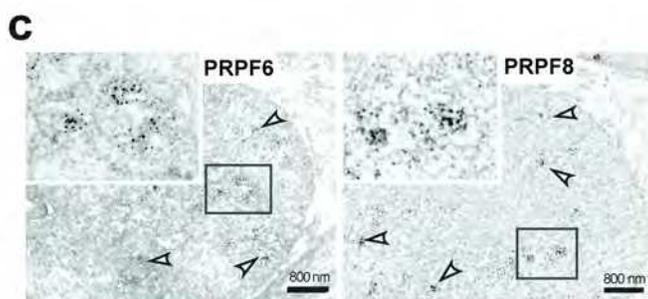
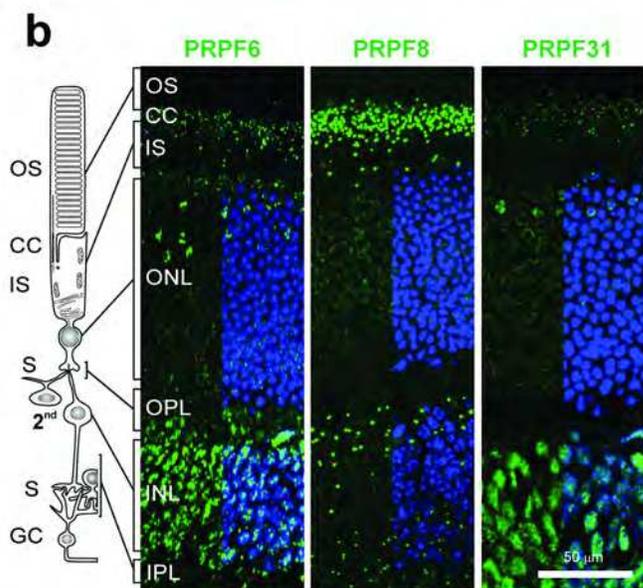
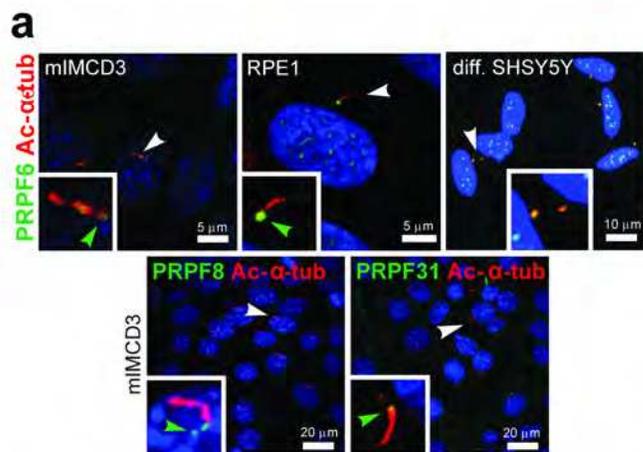
Table 1. Validated hits from secondary and tertiary screens of ciliogenesis. Secondary screen hits are listed (n=68, validated from a total of 194 functional candidates) that had a significant effect on ciliogenesis in mouse mIMCD3 cells (given by robust z scores if $Z_{\text{cilia}} < Z_{\text{cilia cutoff}} = -1.34$, calculated as the median z of positive controls) for $n \geq 1/4$ individual siRNAs duplexes in duplicate. Tertiary screen hits in human hTERT-RPE1 cells for defects in cilia

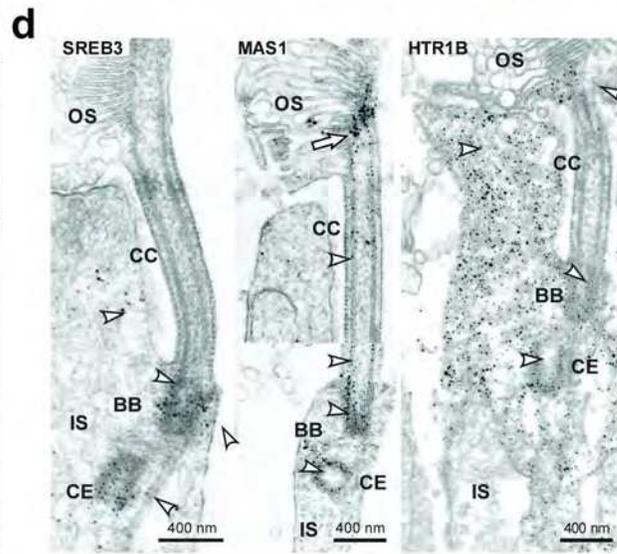
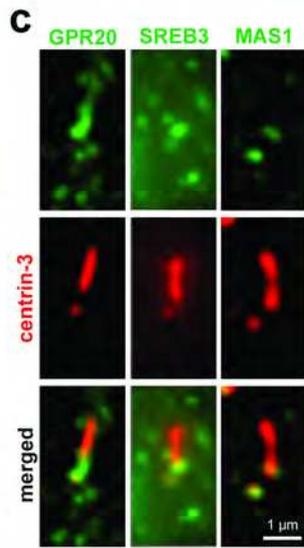
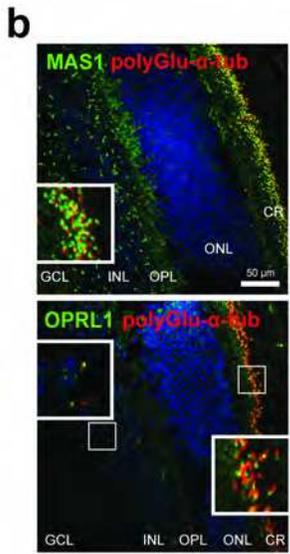
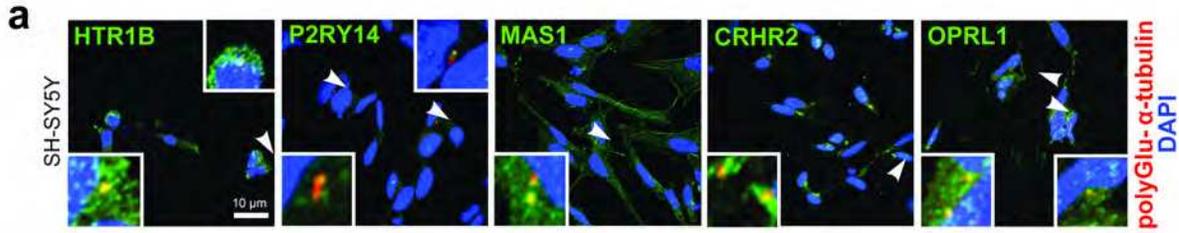
number and/or length (n=37/68, using cut-offs of $z < -2$ or $z > +2$) for siRNA pools in duplicate. Human and mouse gene symbols, mouse gene IDs, known or inferred protein function and selected OMIM annotations are also listed.

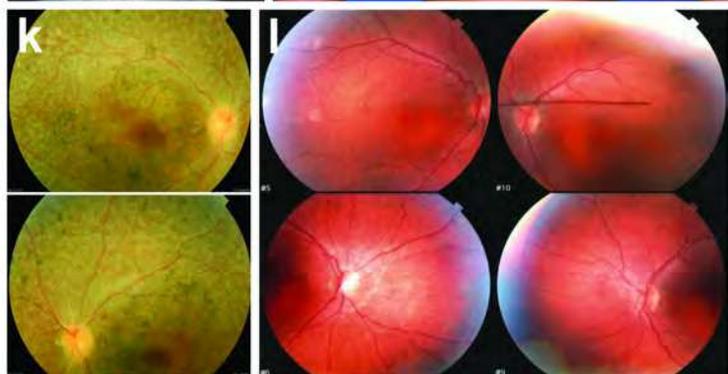
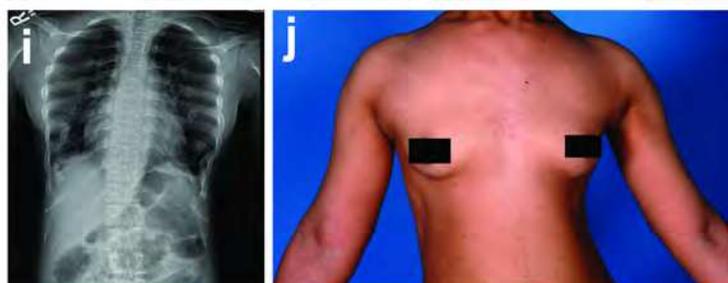
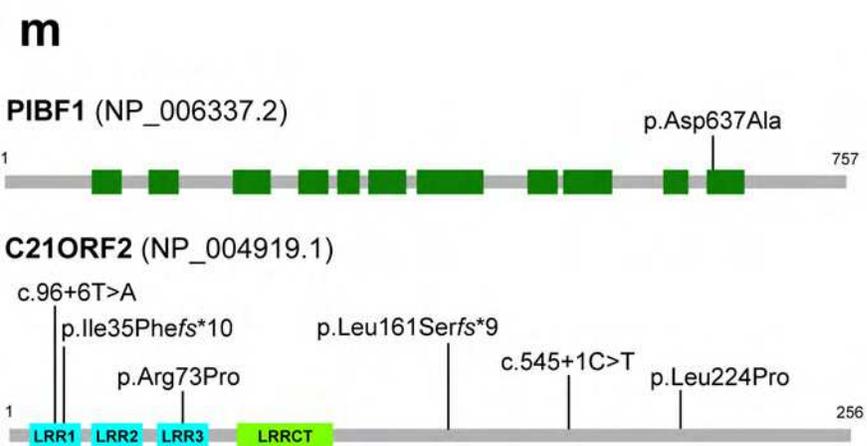
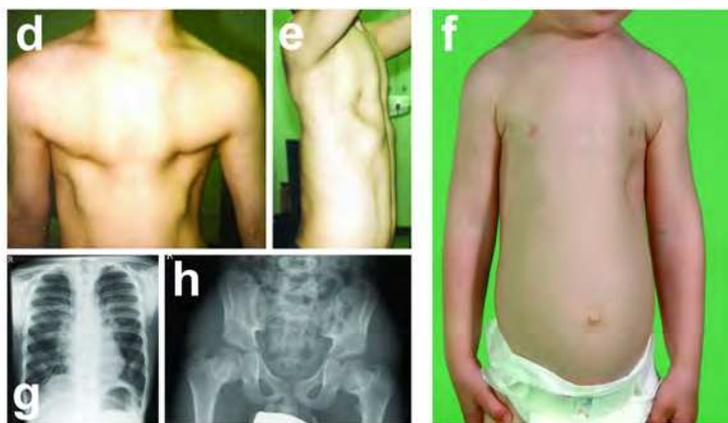
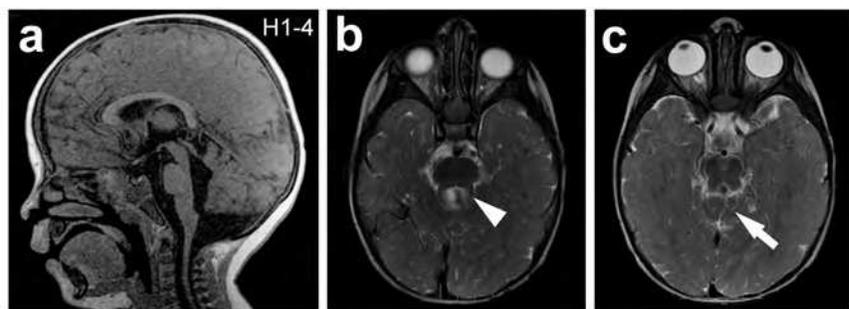






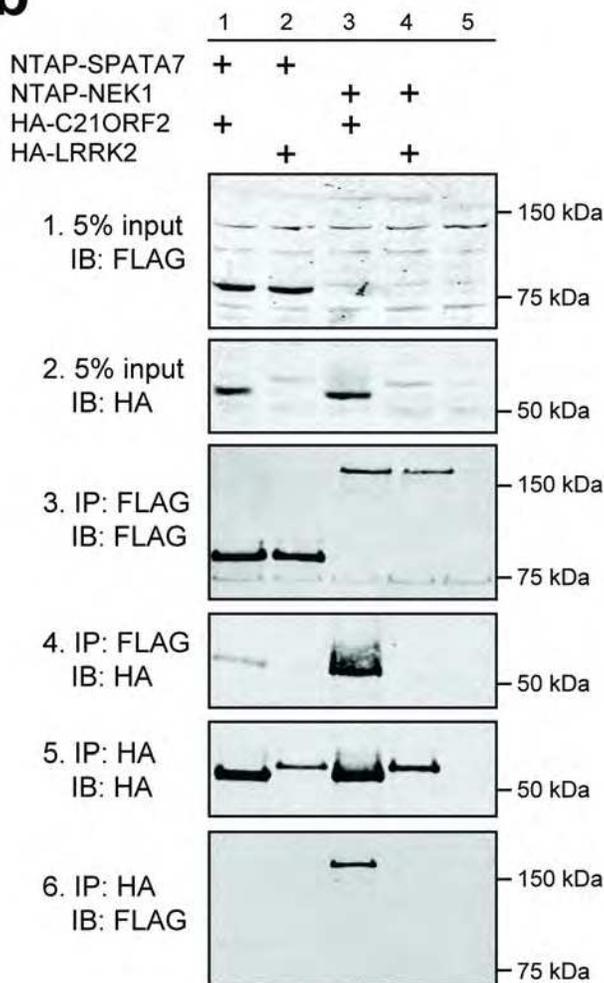
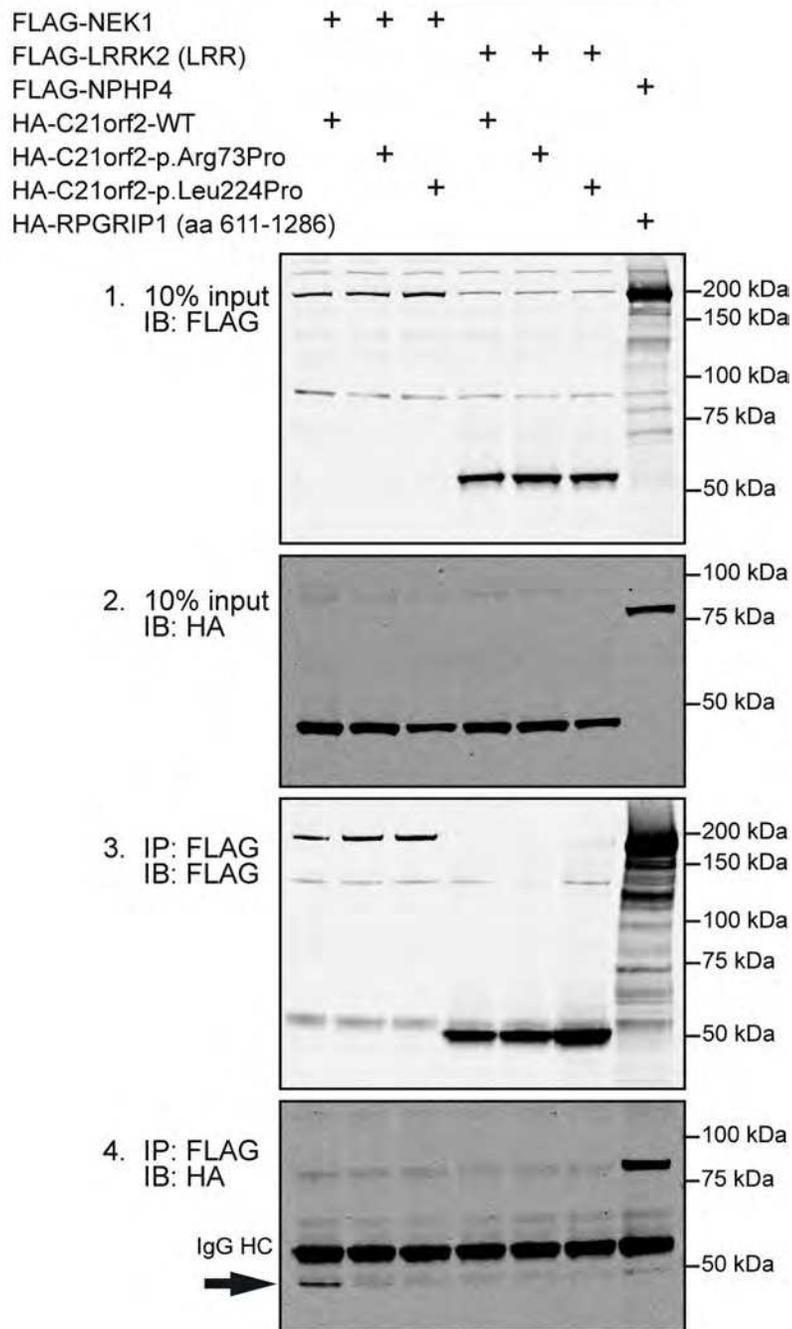
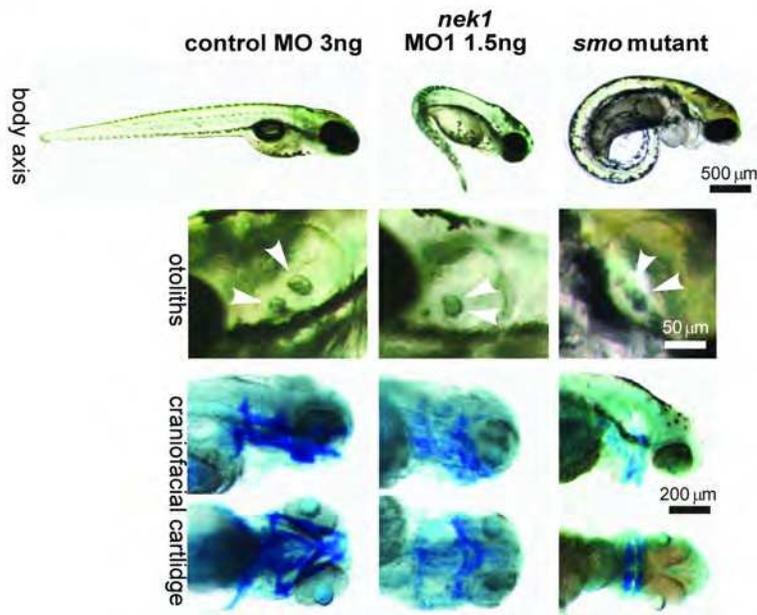
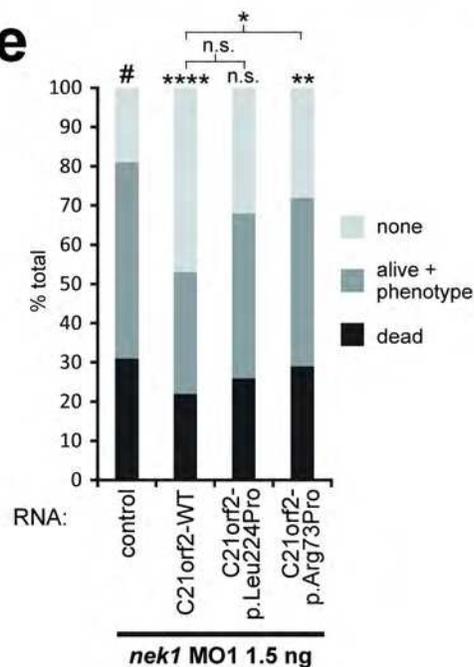






a

		BAIT		
		C21orf2	NEK1	SPATA7
PREY	C21orf2	0.681	0.496	0.335
	NEK1	0.388	0.704	0.392
	SPATA7	0	0	0.786

b**c****d****e**

Mouse Gene ID	Mouse Gene Symbol	Human Gene Symbol	miMCD3 siRNA			hTERT-RPE1 siRNAs		Protein function	OMIM examples (number)
			#1	#2	#3	cilia number	cilia length		
76719	1700081111Bik	KANSI1	-2.13			-5.20	-5.05	histone acetyltransferase	Koolen-De Vries syndrome (610443)
67884	1810043G02Bik	C21orf2	-1.72	-1.67			+4.94	centriolar protein	
69612	2310037124Bik	KANSI2	-2.97	-2.52	-2.28			histone acetyltransferase	
67217	2810055E11Bik	C14orf149	-1.69					isomerase activity	
71617	9130011E15Bik	C10orf26	-4.20					membrane protein	
30806	Adams8	ADAMTS8	-1.78					metalloendopeptidase	
268822	Adck5	ADCK5	-2.28					protein serine/threonine kinase	
11768	Ao1m2	AP1M2	-2.15	-1.62			+2.05	post-Golgi vesicle-mediated transport	
11910	Atf3	ATF3	-2.94	-2.03	-1.90			transcription factor	
242341	At6v0d2	ATP6V0D2	-1.34			+14.57	-7.03	ATP hydrolysis/proton transport	
14595	B4galt1	B4GALT1	-2.94			-3.06	+4.24	post-translational protein modification	Congenital disorder of glycosylation IId (607091)
12419	Cbx5	CBX5	-2.55	-2.53				transcription factor	
77048	Ccdc41	CDC41	-4.41	-2.45		-11.25	-3.37	centriolar protein	
217232	Cdc27	CDC27	-7.38					mitotic cell cycle	
214498	Cdc73	CDC73	-2.35					mitotic cell cycle	Hypoparathyroidism, familial primary (145000)
225523	Cep120	CEP120	-1.95	-1.47				centrosomal protein	Jeune asphyxiating thoracic dystrophy
12922	Chtr2	CFHR2	-4.32					corticotrophin-releasing factor	
227292	Ctdsp1	CTDSP1	-2.46					CTD phosphatase activity	
83603	Elovl4	FLOVL4	-2.16			+2.14		detection of visible light	Macular dystrophy, autosomal dominant (600110)
14428	Galr2	GALR2	-2.57			-2.78		galanin receptor	
14563	Gdf5	GDF5	-3.67					growth factor activity	Brachydactyly, type C (113100)
384009	Glir2	GLIPR2	-1.37					Golgi membrane	
70771	Gpr173	GPR173	-2.27					orphan G-protein coupled receptor	
239530	Gpr20	GPR20	-2.36				+6.45	orphan G-protein coupled receptor	
15107	Hadh	HADH	-2.74	-2.20				fatty acid beta-oxidation	3-hydroxyacyl-CoA dehydrogenase deficiency (231530)
15551	Htr1b	HTR1B	-2.92	-2.70			-2.93	serotonin receptor	
59026	Huwe1	HUWE1	-6.62			-2.23		base-excision repair	Mental retardation, X-linked, Turner type (300706)
15959	Ifit3	IFIT3	-2.78	-1.86			+2.98	type I interferon-mediated signaling	
16535	Kcna1	KCNQ1	-2.39					potassium ion export	Long QT syndrome-1 (192500)
16663	Krt13	KRT13	-3.22					keratin filament	White sponge nevus (193900)
109594	Lmo1	LMO1	-2.70	-1.99			-2.58	metal ion binding	Leukemia, T-cell acute lymphoblastic (613065)
27756	Lsm2	LSM2	-3.14	-2.62	-2.12	-7.95	-2.35	RNA splicing	
66373	Lsm5	LSM5	-2.32	-2.05		-3.36		RNA splicing	
17171	Mas1	MAS1	-2.26	-1.93				angiotensin type II receptor	
51812	Mcrs1	MCRS1	-3.96	-2.38		-8.30	-3.26	histone acetyltransferase	
17349	Mif1	MLF1	-2.23				+3.33	transcription, DNA-dependent	Leukemia, acute myeloid (601626)
17751	Mt3	MT3	-2.13				+3.69	protein import into nucleus	
17754	Mtap1a	MAP1A	-1.88				-5.55	microtubule-associated complex	
217166	Nr1d1	NR1D1	-2.19				+2.15	steroid hormone receptor	
192292	Nrbp1	NRBP1	-2.76				1.65	protein kinase	
77976	Nuak1	NUAK1	-3.45			+4.60	-2.94	response to DNA damage stimulus	
237222	Oid1	OFD1	-3.97	-2.75	-2.63		+2.96	primary cilium protein	Oral-facial-digital syndrome 1 (311200)
404315	Olfir372	OR2Z1	-1.57				+3.97	olfactory receptor	
258961	Olfir631	OR51M1	-3.33	-1.44				olfactory receptor	
14539	Opp1mw	OPN1MW	-1.36			-3.79	-3.00	vision perception	Colorblindness, deutan (303800)
12057	Opp1sw	OPN1SW	-2.35					photoreceptor outer segment	Colorblindness, tritan (190900)
18389	Oprl1	OPRL1	-1.54					nociceptin receptor	
140795	P2rv14	P2RY14	-2.14	-1.98				purinergic nucleotide receptor	
93714	Pcdhaa6	PCDHGA6	-2.98					calcium ion binding	
52023	Pibf1	PIBF1	-3.45	-1.76		-6.81		centrosomal protein	
20873	Plk4	PLK4	-2.10	-1.63		-4.03	+2.77	mitotic cell cycle	Microcephaly and chorioretinopathy (616171)
68988	Prpf31	PRPF31	-10.50	-6.64			+2.35	spliceosome RNA splicing	Retinitis pigmentosa 11 (600138)
230596	Prpf38a	PRPF38A	-8.07			-8.26		spliceosome RNA splicing	
68879	Prpf6	PRPF6	-9.51	-6.54	-5.89	-3.11		spliceosome RNA splicing	Retinitis pigmentosa 60 (613983)
192159	Prpf8	PRPF8	-14.42	-9.26		-8.65		spliceosome RNA splicing	Retinitis pigmentosa 13 (600059)
74760	Rab3il1	RAB3IL1	-2.00				+4.81	Rab GTPase-activating protein	
19387	Rangoap1	RANGAP1	-2.71	-1.53				Ran GTPase activator	
19400	Rapsn	RAPSN	-2.36				+4.84	acetylcholine receptor binding	Myasthenic syndrome, congenital (608931)
56438	Rbx1	RBX1	-6.22					DNA repair	
216227	Slc17a8	SLC17A8	-3.28	-1.41			+2.35	L-glutamate transmembrane	Deafness, autosomal dominant 25 (605583)
20526	Slc2a2	SLC2A2	-1.60					D-glucose transmembrane transporter	Fanconi-Bickel syndrome (227810)
83701	Srrt	SBRT	-3.08	-2.42	-1.77			primary miRNA processing	
66592	Stoml2	STOML2	-3.62					mitochondrial inner membrane protein	
67902	Sumf2	SUMF2	-2.33	-1.93			+3.70	endoplasmic reticulum lumen protein	
78829	Tsc22d4	TSC22D4	-2.05	-1.53				transcription factor	
56085	Ubaln1	UBALN1	-2.01					regulator of protein ubiquitination	
28035	Usp39	USP39	-6.08	-4.89		-3.64	+8.37	RNA splicing	
74716	Wbp2nl	WRP2NL	-2.12				-3.16	meiotic protein	

SUPPLEMENTARY DATA

Supplementary Figure 1. Characterization of the antibody specificities and sub-cellular localizations of hit G protein-coupled receptors.

(a) Localization of the indicated GPCRs (green; with grey-scale image for the green channel on the right) for mlMCD3 cells to the ciliary base (white arrowheads). GPCRs include GPR20 (orphan G-protein coupled receptor 20), GPR173 (orphan G-protein coupled receptor 173), CRHR2 (corticotropin releasing hormone receptor 2), HTR1B (5-hydroxytryptamine/serotonin receptor), MAS1 (MAS1 proto-oncogene receptor), OPRL1 (opioid receptor-like 1), P2RY14 (purinergic receptor P2Y, G-protein coupled, 14) and RAPSIN (receptor-associated protein of the synapse). Knock-down with the cognate siRNA pools for all GPCRs caused loss of primary cilia (visualized by acetylated α -tubulin; red), compared to scrambled (scr.) negative control. Scale bar = 5 μ m.

(b) Localization of GPCRs (green) to the base of primary cilia (stained for polyglutamylated α -tubulin or γ -tubulin as indicated; red) in the different cortical layers of *Tmem67*^{+/+} wild-type embryonic (E18.5) mouse neocortex (left panels). In *Tmem67*^{-/-} knock-out embryos mutated for the transition zone receptor TMEM67, the ciliary localization of GPCRs was lost (right panels). Images are representative of two independent experiments. Scale bar = 20 μ m. Magnified regions are shown in the smaller panels.

Supplementary Figure 2. Mutations in *PIBF1* cause Joubert syndrome.

Homozygosity mapping and pedigrees for Hutterite families with Joubert syndrome. **(a)** Plot of SNP genotypes in family H1, showing two large intervals of absence of heterozygosity (red lines) on chromosomes 10 and 13. **(b)** Pedigrees showing segregation of c.1910A>C in *PIBF1* (NM_006346.2) in Hutterite families H1 to H4 with Joubert syndrome. * indicates individuals with microphthalmia and anterior segment dysgenesis that does not segregate with the homozygous variant. NA, not available. **(c)** Electropherogram of the homozygous *PIBF1* mutation c.1910A>C p.Asp637Ala in affected individual H1-4 compared to a control individual, indicated by red arrowhead. **(d)** Pedigree and electropherograms for the compound heterozygous *PIBF1* variants c.1669delC p.Leu637Phefs*18 and c.1214G>A

p.Arg405Gln in affected individual UW155-3 with Joubert syndrome, indicated by red arrowheads. **(e)** Pedigree of family MTI-121 with compound heterozygous *PIBF1* variants c.1214G>A p.Arg405Gln and a genomic deletion encompassing exons 6 to 9 (c.673-?_1322+?del) of *PIBF1* in affected individuals II.5 and II.7. Electropherogram of the hemizygous c.1910A>C mutation in *PIBF1* in affected individual II:2 (lower electropherogram), indicated by the red arrowhead, as compared to a control individual (upper electropherogram).

Supplementary Figure 3. Range of brain MRI findings in individuals with *PIBF1*-related Joubert syndrome. **(a-c)** Mild vermis hypoplasia, mildly elevated and thickened superior cerebellar peduncles (arrowhead), and superior cerebellar dysplasia (arrow) in H1-3. **(d-i)** Classic molar tooth sign in H2-4 and H3-3. **(j-l)** Classic molar tooth sign with more severe vermis hypoplasia, cervicomedullary heterotopia (white arrow), and foramen magnum cephalocele (white arrowhead) in H4-3. **(a, d, g, j)** are T1-weighted images. **(b, c, e, f, h, i, k, l)** are T2-weighted images.

Supplementary Figure 4. Ciliogenesis rescue experiments with *PIBF1* and *C21orf2* missense mutations. Bar graphs show % ciliated mIMCD3 cells following reverse transfection with either scrambled (scr.) negative control or gene-specific pooled siRNAs, and the indicated FLAG-tagged expression constructs. Ciliogenesis was assayed by automated high content imaging in 96-well plates (see **On-line Methods** and **Supplementary Note** for details), with additional immunostaining with anti-FLAG for normalization of protein expression levels. Values are derived from six fields of view for n=3 independent experiments, expressed as percentages of the % ciliated cells value following scrambled siRNA and empty vector transfection (indicated by #), with the statistical significance of pair-wise comparisons as follows: * p<0.05, ** p<0.01 and *** p<0.001; n.s., not significant (unpaired two-tailed Student's t-test). Other pair-wise comparisons are indicated by braces. Error bars indicate standard error of the mean.

Supplementary Figure 5. Mutations in *C21orf2* cause Jeune syndrome. (a-f) Overview of Jeune family pedigrees that carry *C21orf2* (NM_004928.2) mutations and show segregation consistent with an autosomal recessive inheritance pattern. All affected individuals have cone-rod dystrophy (CRD) with the exception of GC4693.1 shown in **(b)** who has mild retinal degeneration (RD) marked by an asterisk. The key indicates phenotypic variability of the skeletal features ranging from severe (black symbol), mild (mixed symbol) to absent (grey symbol). Consanguineous unions are indicated by double lines. **(e)** Electropherogram showing the effect of the c.545+1C>T mutation by RT-PCR performed on cDNA derived from RNA isolated from peripheral blood cells from individual F2.1 (lower electropherogram) compared to control individual (upper electropherogram). c.545+1C>T results in deletion of five nucleotides at the 3' end of *C21orf2* exon 5, causing a predicted frame-shift of the protein product. **(f)** Predicted effects of the *C21orf2* c.96+6 T>A variant on splicing in patient UCL-78.1. Left panel: visualisation of splicing probabilities for wild-type and mutant *C21orf2* c.96+6 T>A sequence (red text) using Alamut software. Gene Splicer predicts a lack of splicing for the mutant compared to wild-type (wildtype score 1.73, mutant 0; range 0-15); MaxEnt predicts a reduction of splicing probability of 40.1 % (reduction from 9.11 to 5.46; range 0-12); NNS splice predicts a reduction of splicing probability for the mutant of 11.6 % (wild-type score 0.99, mutant 0.87, range 0-1); SSF predicts a splicing probability reduction for the mutant of 6.3% (wild-type score 88.33, mutant 79; range 0-100) and HSF predicts a splicing probability reduction of 2.7% (wild-type score 87.37, mutant 85.04, range 0-100). Right panel: visualisation of predicted splicing enhancer binding sites using Alamut software shows loss of an SRp55 enhancer (highlighted in yellow) in the mutant *C21orf2* c.96+6 T>A sequence (red text) compared to wild-type.

Supplementary Figure 6. Subcellular localisation of *C21orf2*. (a) Indirect immunofluorescence showing co-localization of *C21orf2* (green) with γ -tubulin (red) in ciliary regions in longitudinal sections of adult mouse retinas. Magnified inset indicated by white

frame. Abbreviations: CR, ciliary regions of photoreceptor cells; ONL, outer nuclear layer; RPE, retinal pigment epithelium. Scale bar = 50 μ m. **(b)** Immunohistochemical detection of C21orf2 in bone primordia of mouse embryonic ribs and hip joints at E14.5 and E18.5, respectively (arrowheads). Scale bars = 0.1mm. **(c)** Upper panel: over-expressed eCFP-C21orf2 (cyan) localizes to the cytoplasm with enrichment at the base of the cilium in hTERT-RPE1 cells (arrowhead). Second panel: eCFP-C21orf2 co-localizes with mRFP-NEK1 (red) in the cytoplasm and at the base of the cilium (arrowhead). Third panel: eCFP-C21orf2 is recruited to the microtubule network via its interaction with mRFP-SPATA7 (red). Lower panel: in the presence of mRFP-SPATA7 (red), both eCFP-C21orf2 (cyan) and 3xFLAG-NEK1 (magenta) are localized to the microtubule network (yellow). Acetylated (Ac) α -tubulin antibody is used as a marker for cilia and the microtubule network (red/magenta/yellow), γ -tubulin antibody is used to stain basal bodies and DAPI (blue) is used to stain nuclei. Primary cilia indicated by arrowheads are shown in magnified insets. Scale bars = 10 μ m.

Supplementary Figure 7. Validation of the NEK1 antibody, GST pull-downs of endogenous NEK1 by GST-C21orf2 and GST-SPATA7 fusion proteins from bovine retinal extract, and phenotypes of *nek1* morphant zebrafish embryos. **(a)** Upper panel: western immunoblot with affinity-purified rabbit polyclonal raised against hNEK1H3T1 (1:2000; red frame) for WCEs from control fibroblasts (C), patient fibroblasts (N) with a homozygous *NEK1* nonsense mutation¹ in exon 5 treated with either normal media (5) or serum-starved (0), and HEK293 cells (H). Control fibroblasts and HEK293 cells express full-length NEK1 (predicted size 140kDa) that is absent in patient fibroblasts. Loading control is β -actin (1:3000; green frame). Lower panel: IF labelling of basal ciliary regions in serum-starved control fibroblasts by affinity-purified anti-NEK1 (green; region in magnified inset indicated by arrowhead) with primary cilia labelled with acetylated (Ac) α -tubulin (red). Scale bar = 10 μ m. **(b)** Upper panel: Simply Blue stain of gel showing 10% input GST negative control (lane 2) and GST-fusion proteins GST-C21orf2 and GST-SPATA7 (lanes 3-4). Lower

panel: immunoblot (IB) for endogenous NEK1 on input bovine retinal extract (lane 1) and material pulled down from bovine retinal extract by GST-C21orf2 and GST-SPATA7 (lanes 2-4). **(c)** Verification of morpholino oligonucleotide (MO) effects using RT-PCR. For the *nek1* morpholino oligonucleotide (MO1) targeting the exon6 splice donor site, the wild-type product size is approx. 750 bp whereas MO-injected embryos result in a product of approximately 1250 bp suggesting inclusion of intron 6/7 (intron length 575 bp; see **On-line Methods** for further details). Amplification of *gapdh* shows approximately equal cDNA input. **(d,e-)** H&E staining of eyes of embryos injected with 2ng *nek1* MO1 revealed both a reduced outer nuclear layer and generally smaller eyes compared to controls. Abbreviations: INL= inner nuclear layer, ONL= outer nuclear layer. Scale bar = 50 μ m. **(f,g)** 2ng *nek1* MO1 injections in rhodopsin:GFP transgenic embryos caused loss of photoreceptor cells (arrows) at 4 dpf compared to controls. Scale bar = 25 μ m. **(h,i)** Acetylated (Ac) α -tubulin staining of *nek1* MO1-injected embryos revealed a severe reduction of cilia length in the pronephros compared to control embryos (cilia indicated by arrows). Cilia length appeared normal in the neural tube (data not shown). Scale bar = 10 μ m.

Supplementary Figure 8. Full images of SDS-PAGE gels/western blots. Full unprocessed SDS-PAGE gel images and signals detected upon western blotting, with the regions used in the corresponding main display items indicated by red frames.

Supplementary Table 1. Complete dataset for the genome-wide siRNA primary screen for ciliogenesis. The first spreadsheet ("All results, robust z scores") summarises the results of the whole genome mouse siRNA screen performed in mIMCD3 cells and lists robust z scores for % cells with a single cilium (Z_{cilia}) and whole cell number (Z_{cell}). This spreadsheet details: the mouse gene symbol and Entrez GeneID, the gene symbol of the human orthologue (if any) and Entrez GeneID, sublibrary and plate position of the siRNA pool, and the results for % cells with a single cilium and whole cell number after 72 hours siRNA knockdown. The screen results are expressed as robust z scores, and these are calculated

per 10-plate batch. Each siRNA pool has two results (run 1 and run 2), which represent two independent biological replicates. As an arbitrary measure, $z=-2$ is usually considered an acceptable z score cutoff for significant hits. It is more meaningful, however, to set the cutoff according to the median robust z score of the positive controls. For this reason, the median and median absolute deviation (MAD) robust z scores of the positive controls are given for each batch in columns next to the test results. If a test result is greater in value than the median robust z score for the positive controls ($Z_{\text{cilia}} < Z_{\text{cilia_cutoff}}$) for both replicates, then this siRNA has a greater effect on ciliogenesis than the median of the positive controls and the result is taken as a significant hit. Results are presented in batches in order of their processing in chronological time. Within each batch, the results are ordered by average robust z score for % cells with a single cilium. Cells indicated in the "Fail" column should be interpreted with caution: the results in columns A and H in run 2 of batch mDT51-60 are not reliable because of technical issues with primary antibody application. Plate mG37 in run 1 had to be repeated because of technical issues, and is therefore analysed independently from the rest of batch mG31-40 with different median and MAD values for the controls. The first spreadsheet ("All results, robust z scores") also lists additional information about the mouse gene including Genetic Ontology codes, phenotypes of any mouse models (from MGI) and GeneRIFs (from NCBI). Human orthologues are also listed, with additional summaries of known or inferred function of the encoded protein and any OMIM annotations. The next two tabs list genes that are only partially targeted by siRNA (i.e. not all transcripts were targeted by the siRNA reagent) or have off-target effects (i.e. the siRNA reagent targets transcripts from genes other than the one it was designed for). Both of these lists are derived from BLASTN analyses of all siRNA duplex sequences in the ThermoFisher siGENOME library. Genes are listed by Entrez GeneID. The fourth tab lists genes by NCBI GeneID with abnormal copy number generated by the analysis of array CGH data. '-2' indicates a biallelic deletion, '-1' monoallelic deletion, '1' monoallelic duplication, '2' biallelic duplication. Genes with two or more values indicate that the siRNA duplexes had potential off-target effects. The

fifth tab lists genes by NCBI GeneID with siRNAs that have potential miRNA-like effects, as determined by analysis for seed sequences by GESS.

Supplementary Table 2. Complete dataset for primary screen hits. The spreadsheet lists robust z scores for % cells with a single cilium (Z_{cilium}) and whole cell number (Z_{cell}) for primary screen hits following screening in mIMCD3 cells. Details are as for **Supplementary Table 1**. In total, 2174 siRNA pools significantly decreased the percentage of cells with a single primary cilium ($Z_{\text{cilium}} < Z_{\text{cilium cutoff}}$ in a processed batch for both replicates, where $Z_{\text{cilium cutoff}} =$ median z of all positive controls), with 1956 hits targeting a gene with a human orthologue. Potential non-specific siRNAs comprising those with predicted off-target effects (n=115) or to have microRNA-like effects (n=12) left a total of 1829 mouse genes with a human orthologue.

Supplementary Table 3. Complete datasets for secondary and tertiary screen hits. The spreadsheets list robust z scores for % cells with a single cilium (Z_{cilium}) and whole cell number (Z_{cell}) for secondary and tertiary screen hits following screening in mIMCD3 and hTERT-RPE1 cells. The first spreadsheet (IMCD3 secondary screen 1) lists n=40 out of the 1829 primary screen hits that had human orthologues and were functional candidates and representative components of a significantly enriched pathway that was predicted by DAVID analysis. Results are given for four individual siRNA duplexes for each gene, listing gene ID and mouse gene symbol. The second spreadsheet (IMCD3 secondary screen 2) lists n=154 hits with human orthologues following filtering of the 1829 primary screen hits for those that significantly affect the percentage of cells with a single cilium (Z_{cilium}) but did not have a significant effect on cell number ($Z_{\text{cell}} > Z_{\text{cell cutoff}}$ in a processing batch for both replicates, where $Z_{\text{cell cutoff}} =$ median z-2MAD for all negative controls). Results are given for four individual siRNA duplexes for each gene, listing gene ID and mouse gene symbol. The third spreadsheet lists robust z scores for cilia length ($Z_{\text{cilium length}}$) and % cells with a single cilium (Z_{cilium}) for n=68 secondary screen hits following re-screening with pooled siRNAs in hTERT-RPE1 cells.

Results are given for pooled siRNA duplexes for each gene, listing gene ID and mouse gene symbol and gene symbol of the human orthologue.

Supplementary Table 4. Complete dataset for secondary screen of ubiquitin-proteasome system (UPS) hits. Primary screen hits are listed (n=57), with a total of n=34/57 genes with a significant effect on ciliogenesis (robust z scores ≤ -2 or $\geq +2$; highlighted by red or green text, respectively) and n=29/57 with a significant effect on cilium length (robust z scores ≤ -2 or $\geq +2$) versus in-plate non-targeting (NT) control duplexes, for n ≥ 2 out of 3 individual siRNA duplexes using hTERT-RPE1 cells. A total of n=44/57 genes had a significant effect on cilia number and/or cilia length ciliogenesis. Human Entrez gene symbol, gene ID, siRNA ID and siRNA sequences are listed. Transfection and image analysis methods were the same as for the tertiary screen, listing robust z scores for cilia length ($z_{\text{cilia length}}$) and for % cells with a single cilium (z_{cilia}).

Supplementary Table 5. Variants identified by whole-exome sequencing in patients with Joubert and Jeune syndromes. Genes with causative mutations are underlined. Abbreviations: dam, damaging; del, deleterious; EVS, Exome Variant Server; GS, Grantham Score; MAF, Minor Allele Frequency; neut, neutral; P2, Polyphen2; prob, probably; tol, tolerated. **COL18A1*: mutations cause Knobloch syndrome, (KNO1; OMIM #267750) an autosomal recessive disorder defined by high myopia, vitreoretinal degeneration with retinal detachment, macular abnormalities, and occipital encephalocele.

Supplementary Table 6. MRI features of Joubert patients with *PIBF1* mutations. Abbreviations: d, days; DD, developmental delay; F, female; M, male; mo, months; L, left; MTS, molar tooth sign; na, not applicable; nd, not determined; PMG, polymicrogyria; R, right; SC, superior cerebellum; SCP, superior cerebellar peduncle; yr, years; ?, unknown.

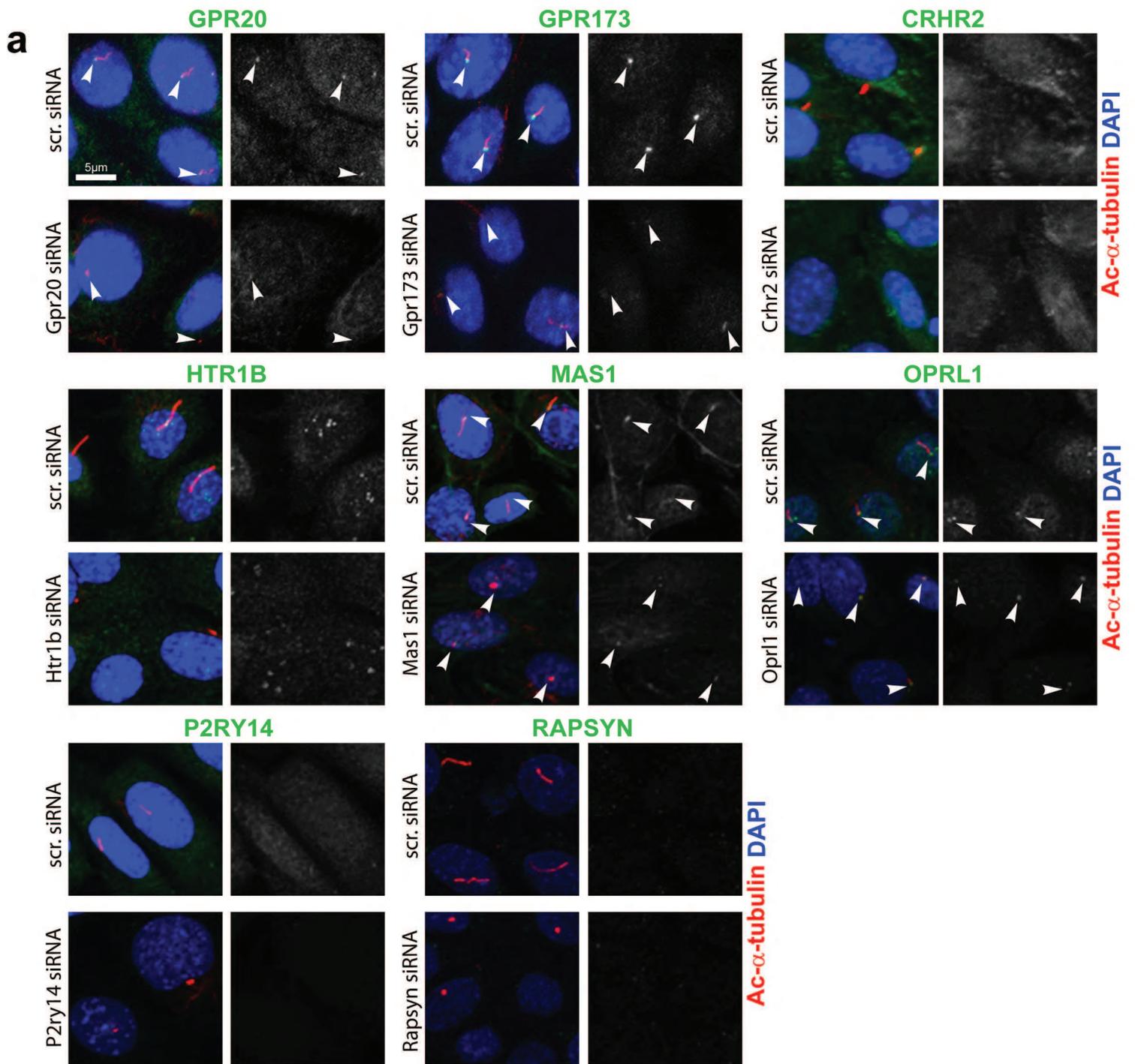
Supplementary Table 7. Clinical features of ciliopathy patients. ^a individuals CR and F2 have been described previously²³; ^b *C21orf2* c.218C>G annotated as variant rs140451304 (ExAC allele calls C=111976/G=37); ^c individual investigated by CT imaging only; ^d tonic seizures in infancy; ^e complex eye defects (microphthalmia, opaque corneas, severe mesodermal dysgenesis); ^f high hyperopic refractive error; ^g mild retinal degeneration; Abbreviations: d, days; DD, developmental delay; F, female; M, male; mo, months; MTS, molar tooth sign; na, not applicable; nd, not determined; OMA, oculomotor apraxia; SD, splice donor; VHD, vermis hypoplasia and cerebellar dysplasia; yr, years; ?, unknown.

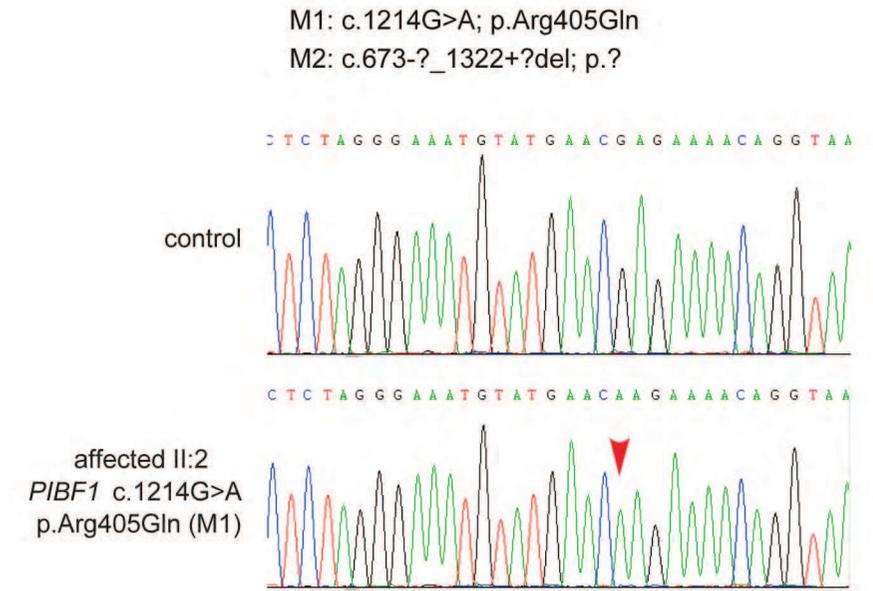
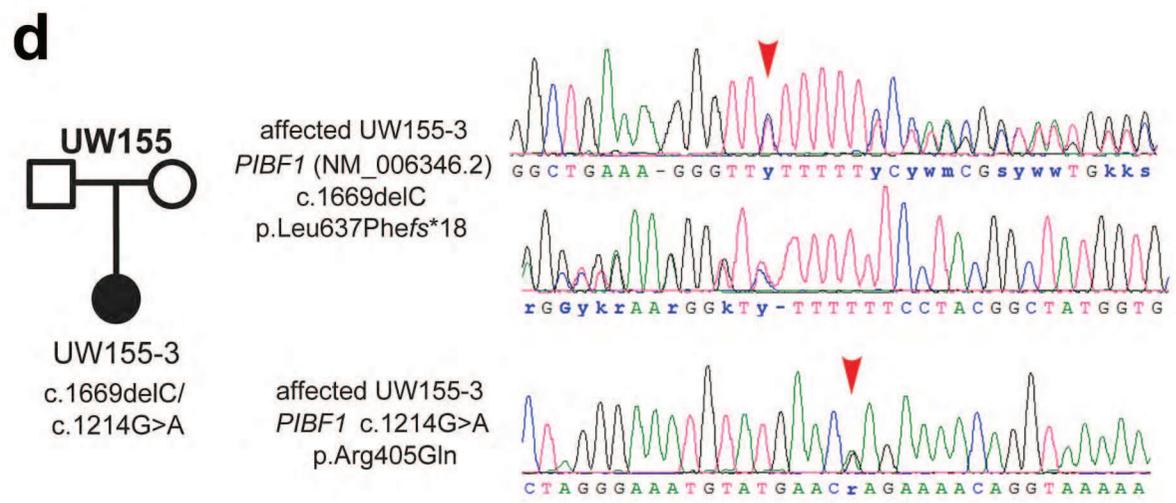
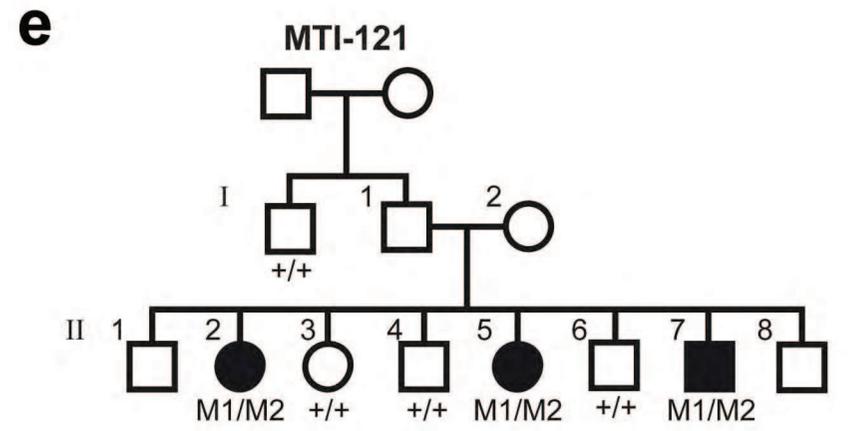
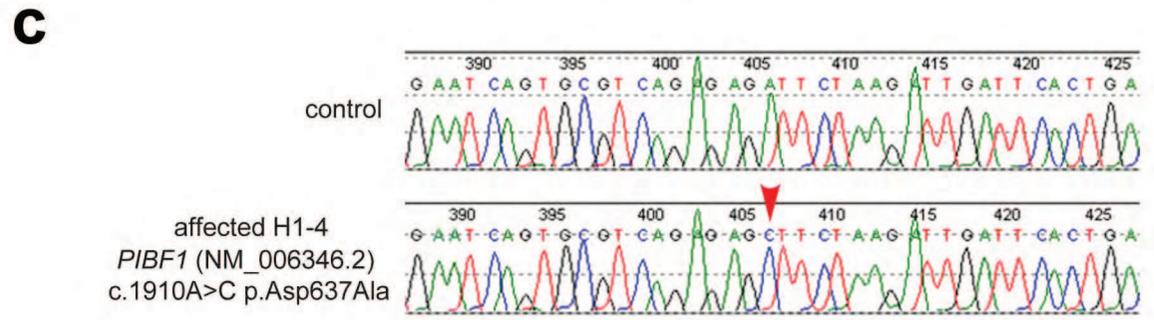
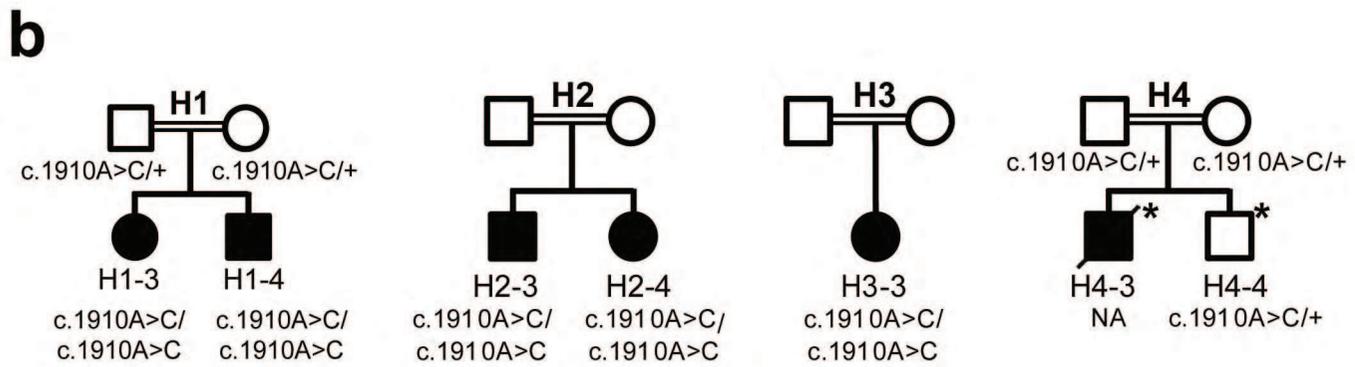
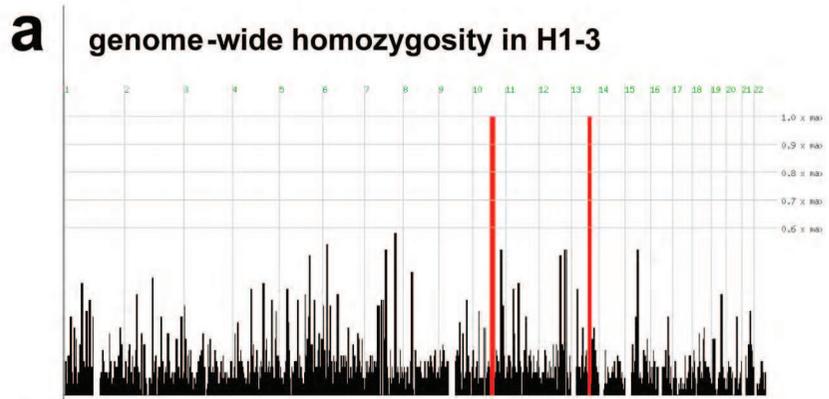
Supplementary Table 8. Strep-tag II/FLAG tandem-affinity purification (SF-TAP) and mass spectrometry analysis of SF-TAP-tagged C21orf2, NEK1 and SPATA7 in HEK293 cells. Three individual SF-TAP experiments N-terminal (n=2) and C-terminal (n=1) SF-TAP-tagged C21orf2 (pale yellow columns), NEK1 (green columns) and SPATA7 (purple columns) were performed. Listed are the SWISS Prot Accession numbers, Entrez gene symbol and name, unique peptides and sequence coverage for the preys identified per TAP experiment and for the majority of TAP experiments per bait protein. Rows for C21orf2, NEK1 and SPATA7 are highlighted in yellow.

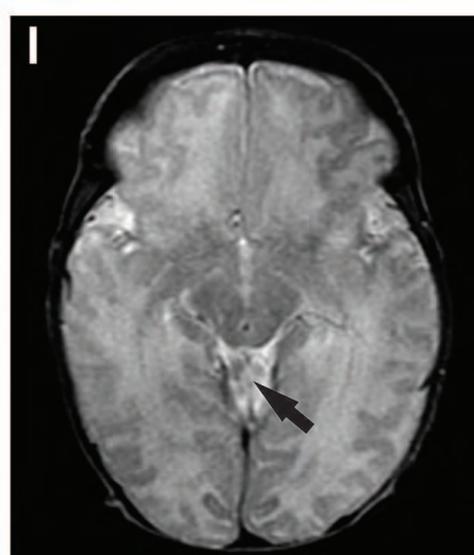
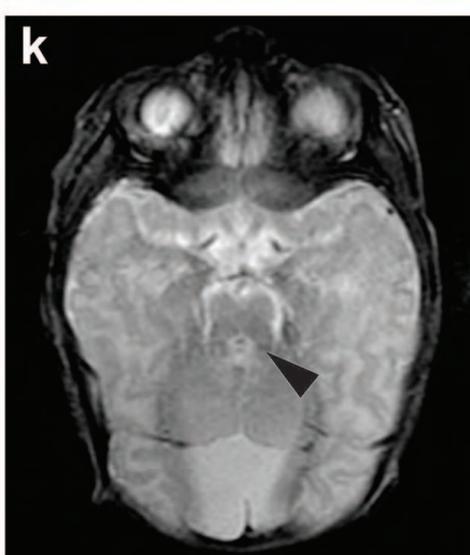
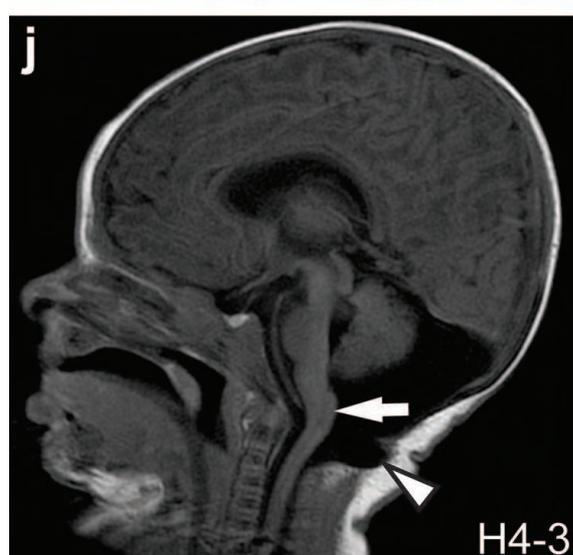
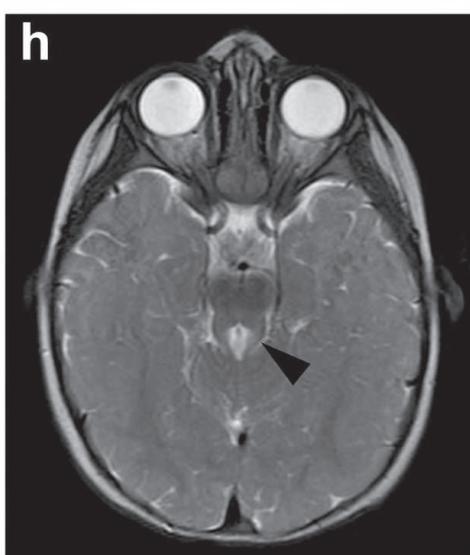
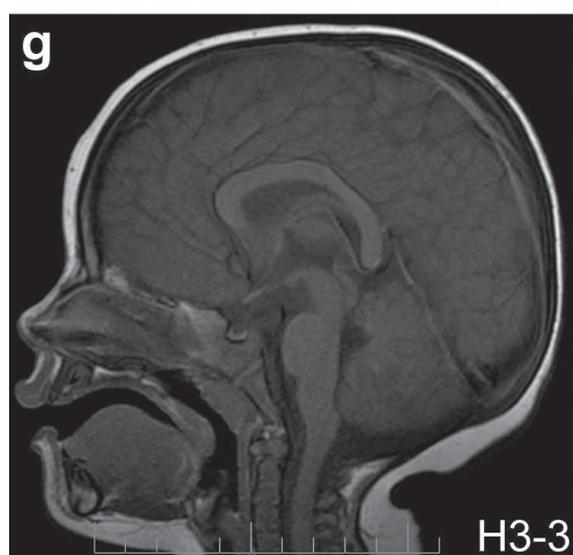
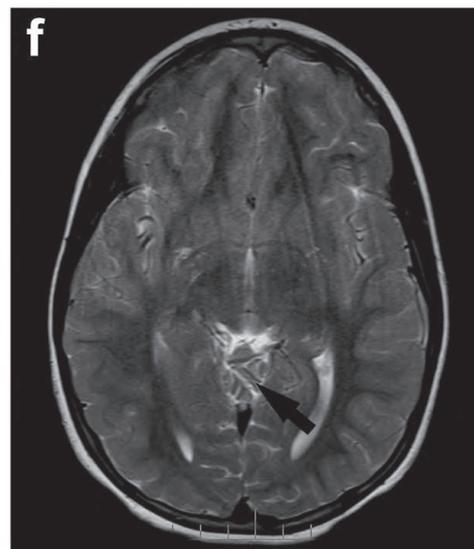
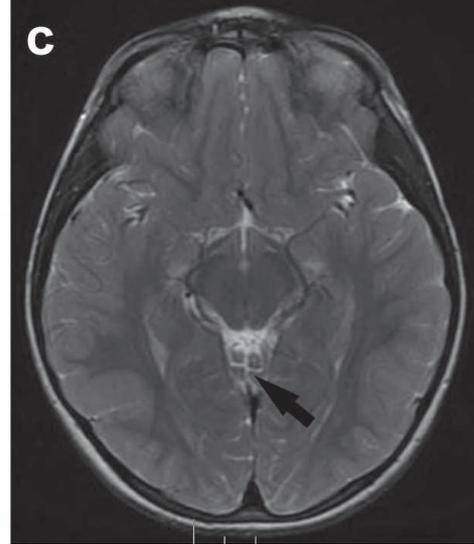
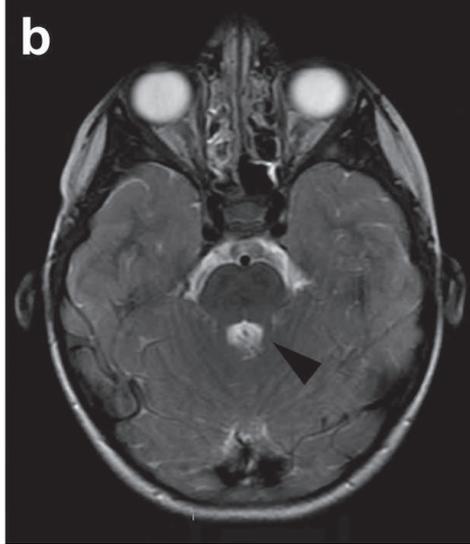
Supplementary Table 9. Mass spectrometry analysis of NEK1 endogenous co-immunoprecipitation and GST pull-down of SPATA7 from bovine retinal lysates. Unique peptide counts for C21orf2, NEK1 and SPATA7 from mass spectroscopy analysis are listed for immunoprecipitation (IP) experiments (n=3 biological replicates) with anti-NEK1 antibody compared to anti-FLAG negative control, and GST pull-down assays with GST-SPATA7 fusion protein compared to GST negative control. Protein symbols, Uniprot accessions (Acc.) and names are listed.

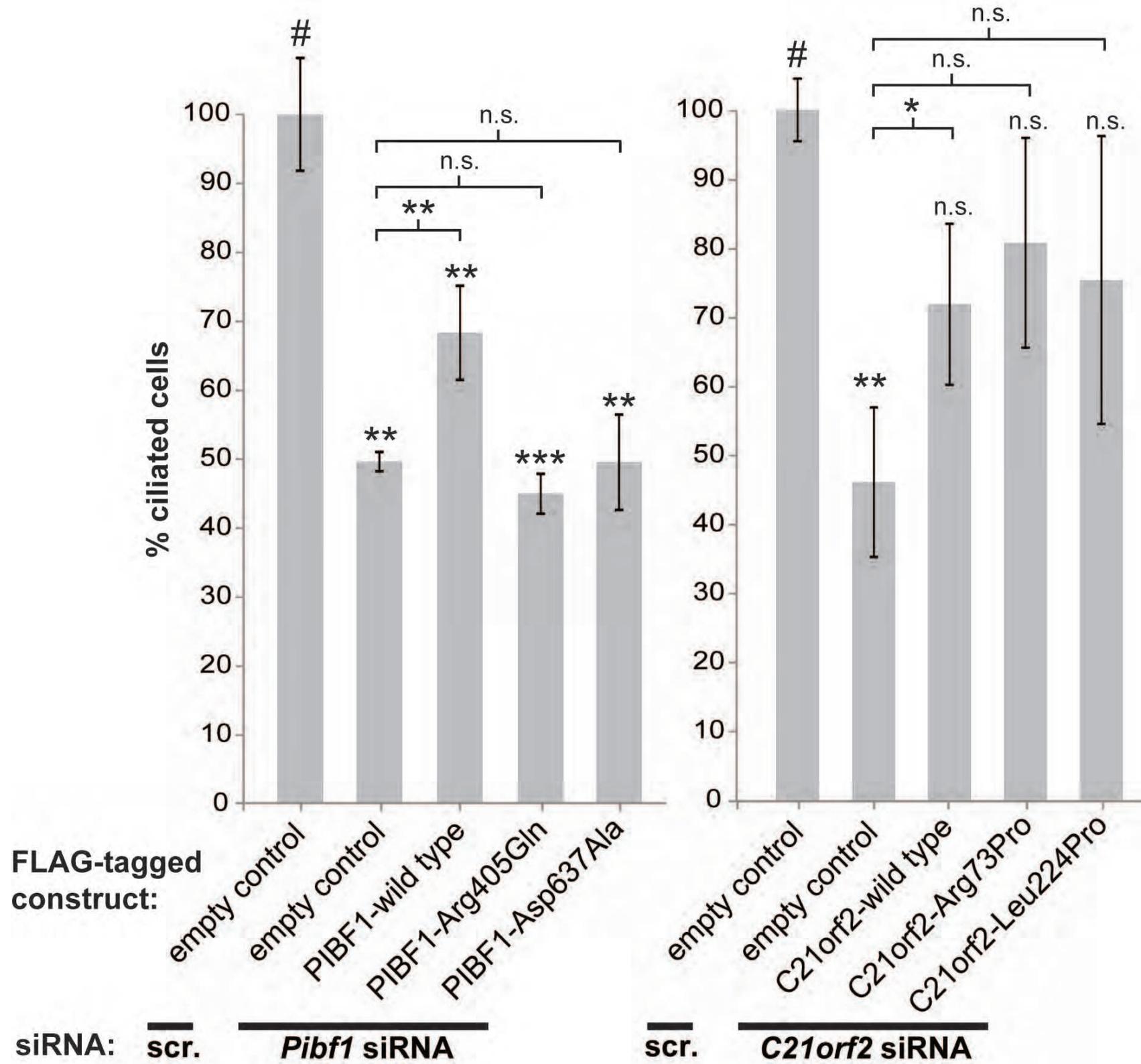
Supplementary Table 10. Sequences of siRNA duplexes used to target positive and negative control genes for ciliogenesis in mlmCD3 cells.

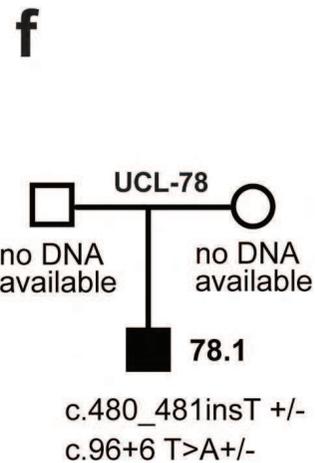
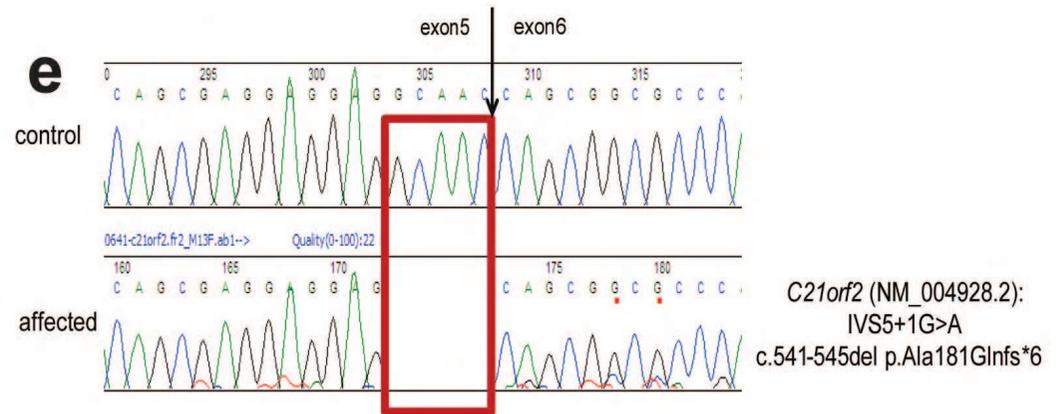
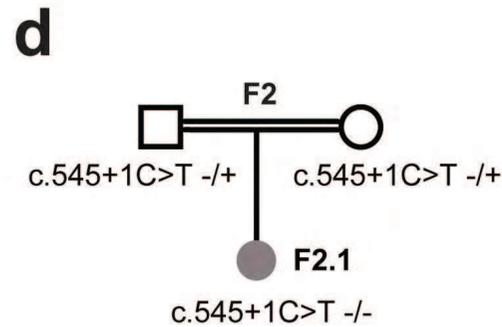
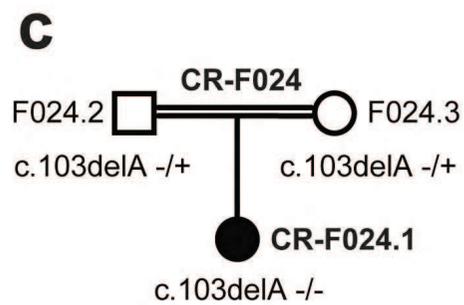
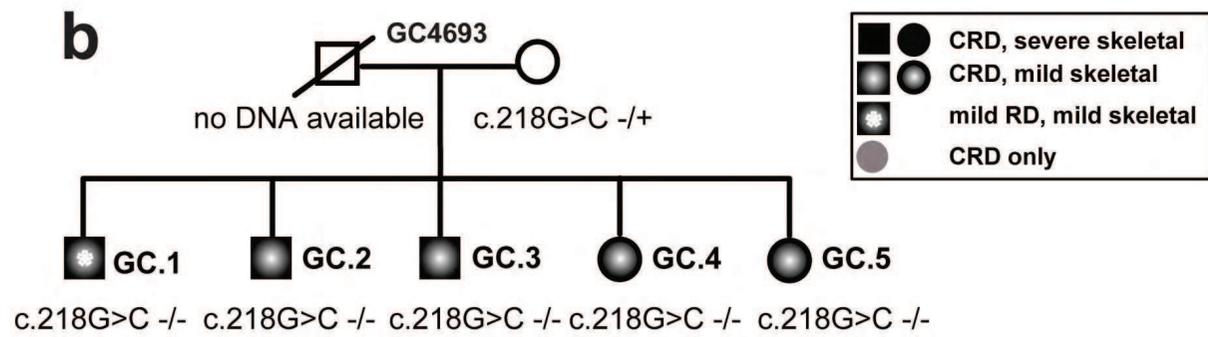
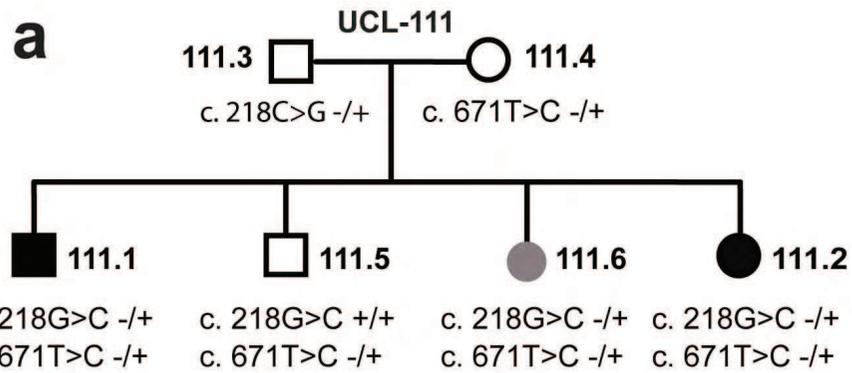
Supplementary Table 11. Complete dataset for additional ciliopathy and ciliary-related hits. Dataset for n=780/1829 of the total hits from the primary screen with an effect on cilia number that is significantly greater ($z > +2$) than any effect on cell number. Details are as for **Supplementary Table 1.**

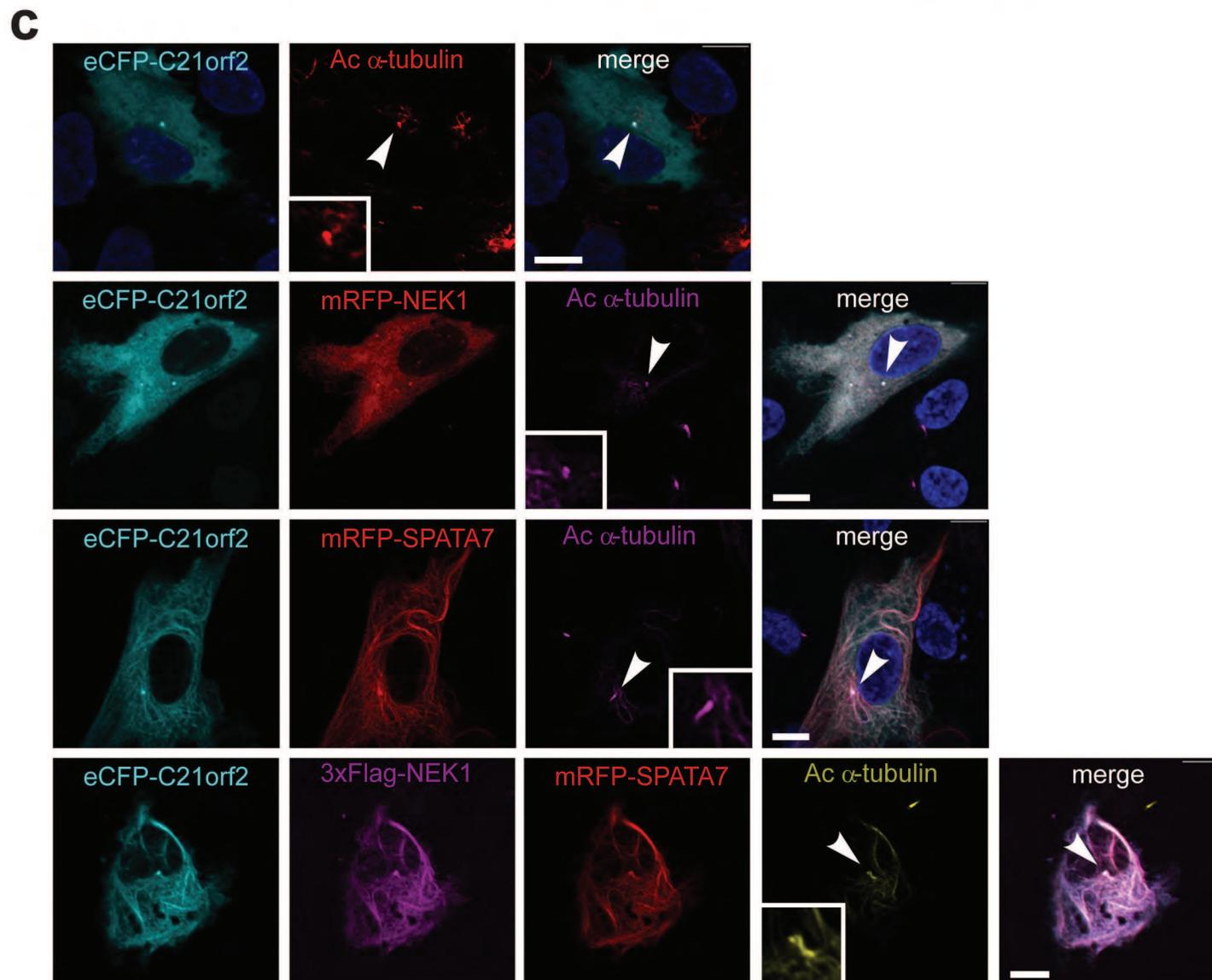
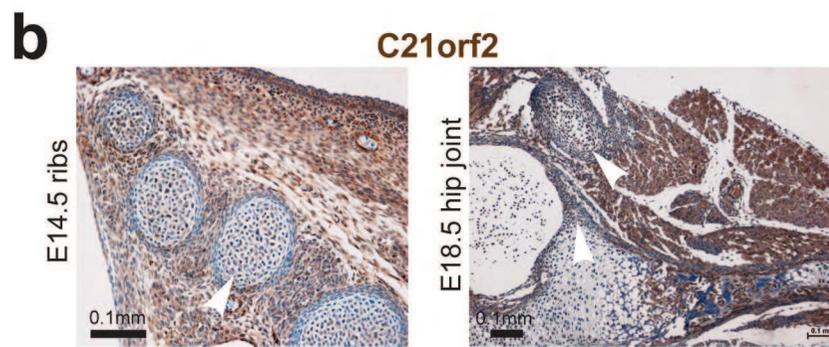
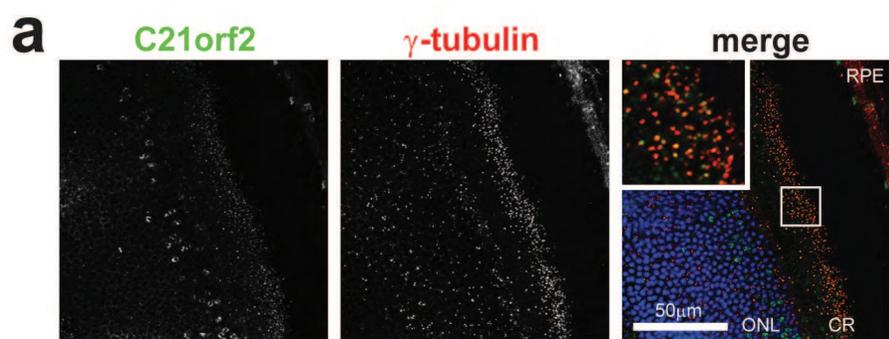












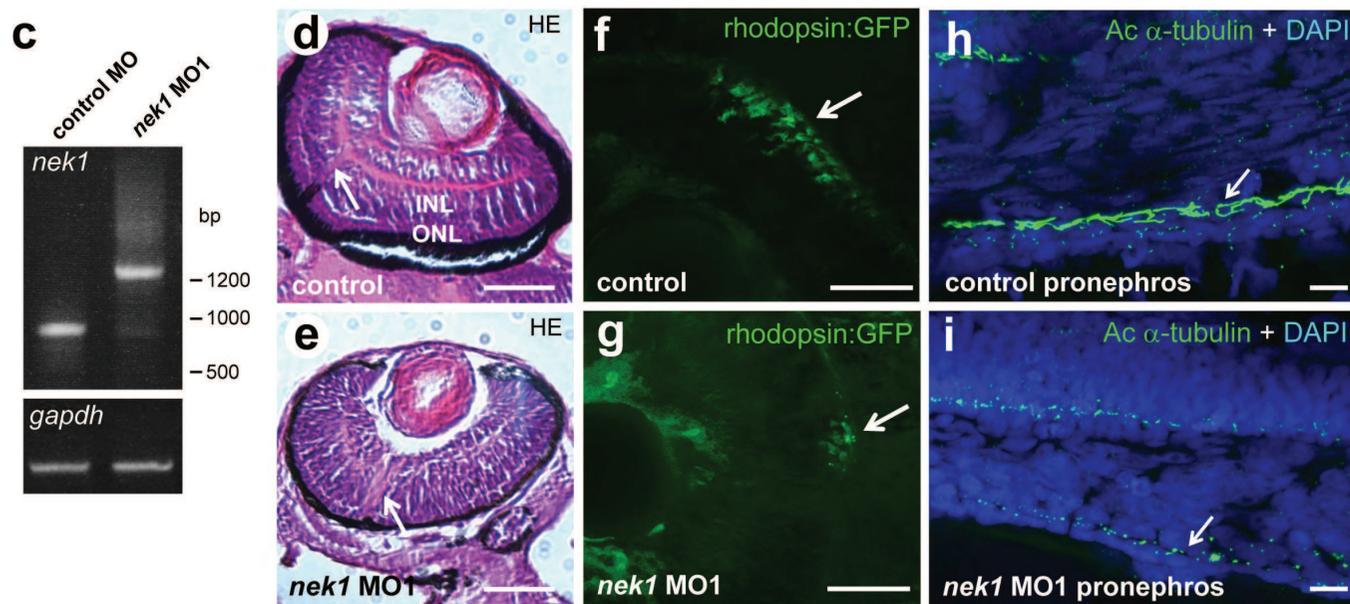
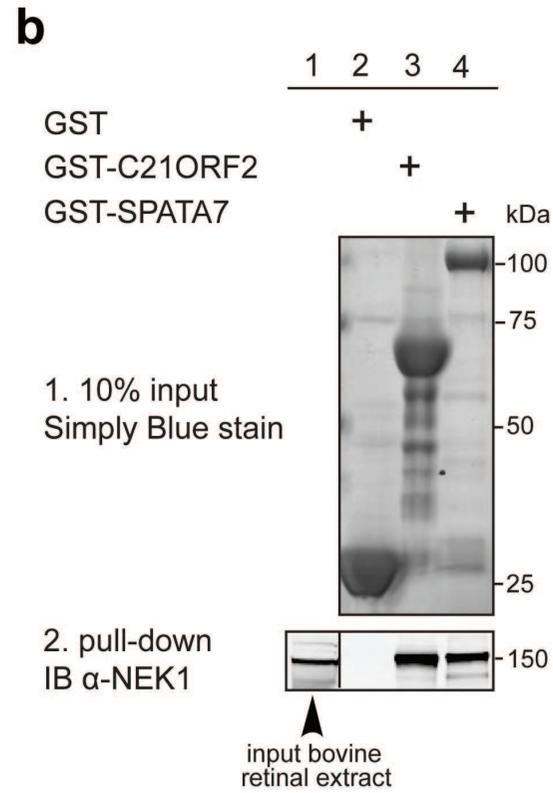
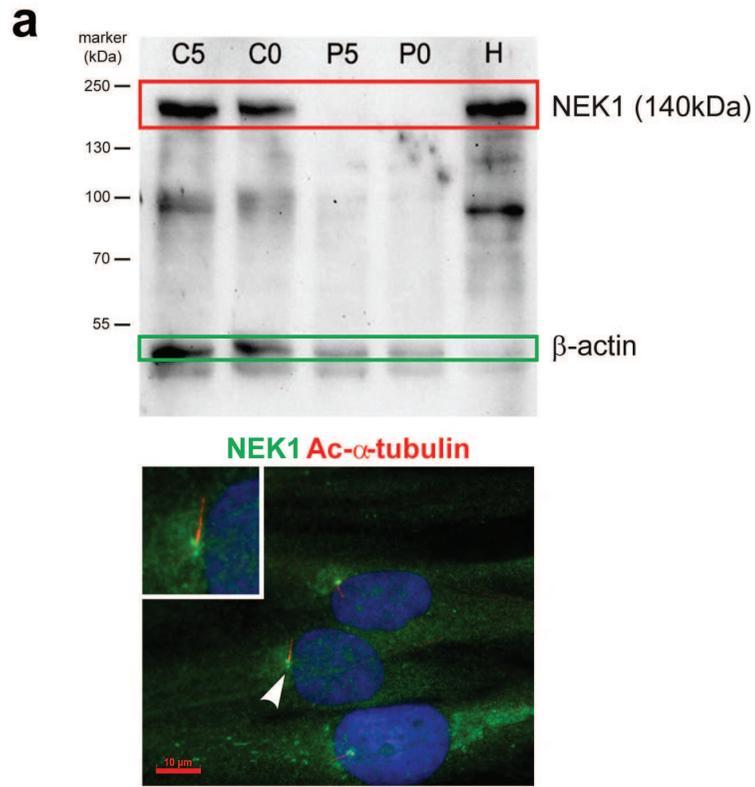
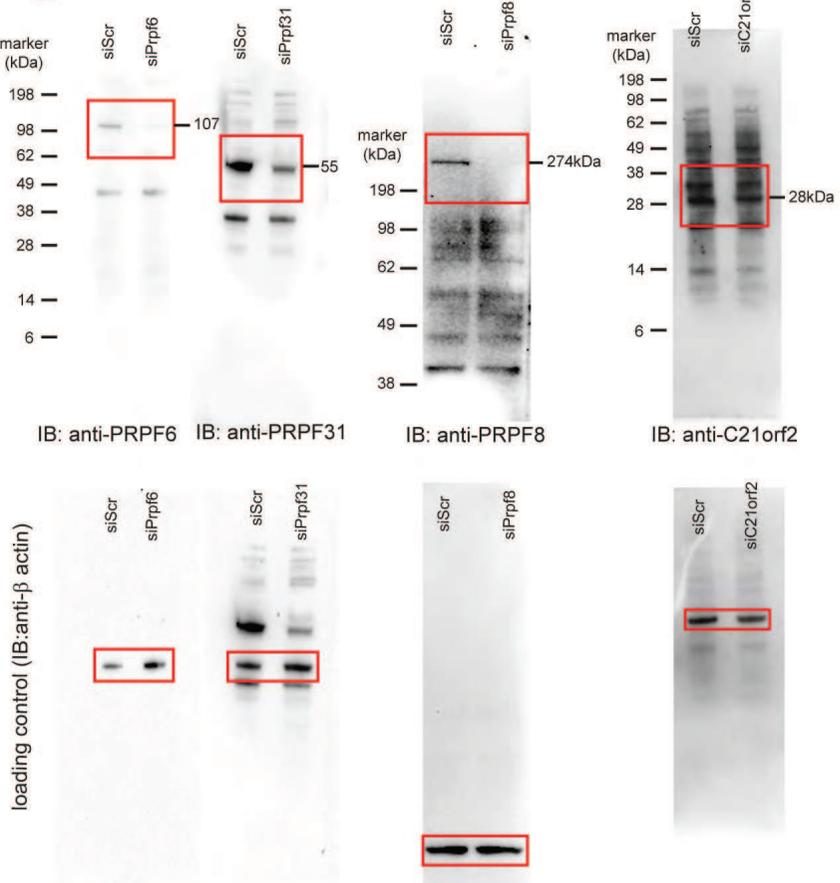
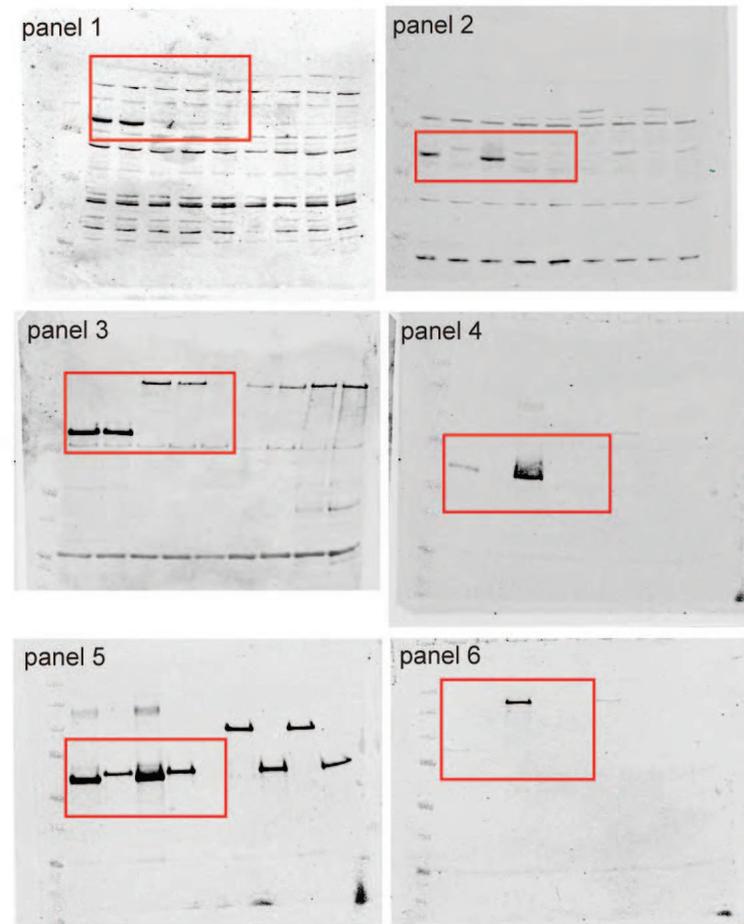
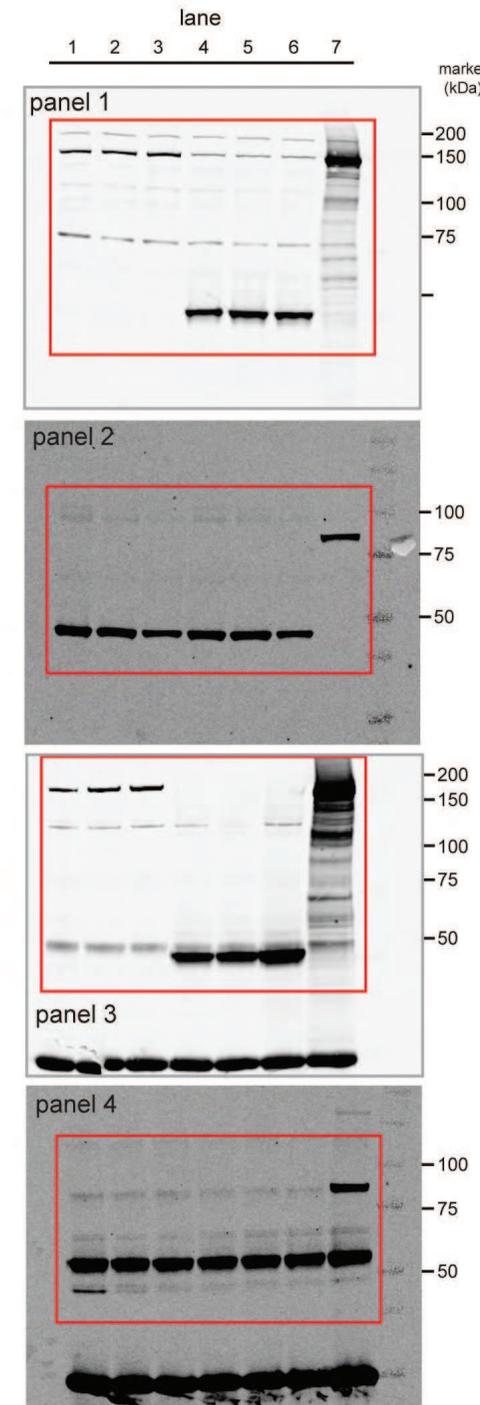
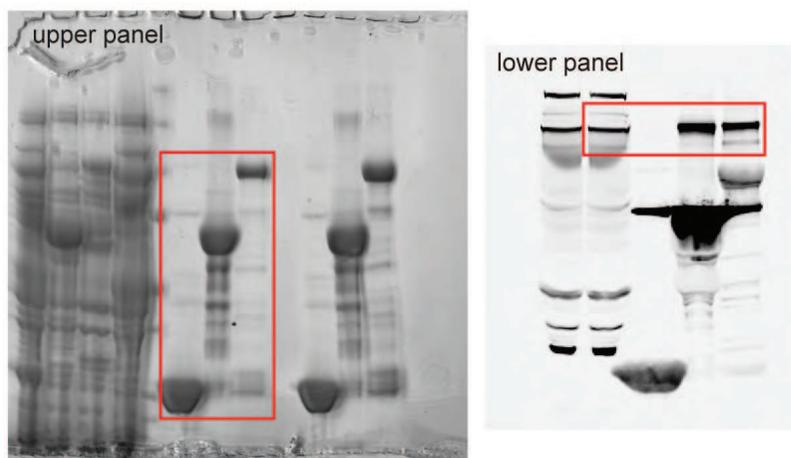


Figure 3c**Figure 7b****Figure 7c****Suppl. Figure 7b**

ON-LINE METHODS AND MATERIALS

Cell culture

Mouse inner medullary collecting duct (mIMCD3), human telomerase reverse transcriptase-transformed retinal pigment epithelium (hTERT-RPE1) and human neuroblastoma (SH-SY5Y) cells were derived from American Type Culture Collection (ATCC) and maintained in DMEM/Ham's F12 medium supplemented with 10% foetal calf serum (FCS), under standard conditions (37°C, 5% CO₂). Cells were passaged at a split ratio of 1:10 twice a week. mIMCD3 cells were obtained from ATCC at passage 13 and were used for screening purposes between passage 17 and 25. SH-SY5Y cells were differentiated over 6 days by retinoic acid (RA) treatment followed by brain derived neurotrophic factor (BDNF) treatment. Cells were plated in standard culture medium for 24 hours, in standard medium containing 10µM RA for 3 days and in serum-free medium containing 50ng/ml BDNF cells for 3 days. Human dermal fibroblasts were derived from skin biopsies and cultured in fibroblast growth medium (Genlantis). hTERT-RPE1 cells were serum starved in normal media with 0.2% FCS for 48 hours to induce ciliogenesis.

Mice

C57Bl/6J wild-type mice and the B6;129P2-*Tmem67*^{tm1Dgen/H} line were kept on a 12 h light-dark schedule with unlimited access to food and water. All procedures were in accordance with local laws on animal protection, the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research, and under the guidance issued by the Medical Research Council in *Responsibility in the Use of Animals for Medical Research* (July 1993) in accordance with UK Home Office regulations under the Project Licence no. PPL40/3349.

Plasmids

Full-length Gateway Entry clones for human C21orf2 (IOH23207), SPATA7 (IOH62585), PRPF6 (IOH3880), PRPF31 (IOH41469) and PRPF38A (IOH46080) were kindly provided by Nico Katsanis (Duke University, Durham, USA). Human PIBF1 (kindly provided by Kym Boycott) and

NEK1 (NM_012224.2, kindly provided by Andreas Gießl) were cloned into the Gateway system (LifeTechnologies). Full-length NPHP4, amino acids 611-1286 of RPGRIP1 and the LRR domain of LRRK2 constructs were generated as previously described³¹ and cloned into the Gateway system. Entry clones were sequence-verified and further cloned into HA-, GST-, FLAG-, mRFP, eCFP and strep-II tag/FLAG-tandem affinity purification (SF-TAP)-tagged destination vectors using the Gateway system. Missense mutations for PIBF1 and C21orf2 were introduced with the Q5 Site-Directed Mutagenesis Kit (New England Biolabs), using conditions recommended by the manufacturer.

Generation of hNEK1 specific antibody

A 444 bp hNEK1 (NM_012224) cDNA (hNEK1 H3T1; 1759-2202 bp, encoding amino acids 587-734)³⁵ was cloned into the pGEX-4T3 expression vector (GE Healthcare Life Sciences) using *Bam*HI and *Xho*I restriction sites. Expression and purification of the GST fusion protein was performed according to the manufacturer's instructions (GE Healthcare Life Sciences). After cleavage of the fusion protein with thrombin on the column, hNEK1H3 was eluted with 1 x PBS followed by peptide immunization of rabbits by a commercial supplier (Pineda Antikörper-Service, Berlin, Germany). Polyclonal antisera against recombinantly-expressed hNEK1H3T1 were affinity-purified on high trap N-hydroxysuccinimide columns (HiTrap NHS-Activated HP; product code:17-0716-01; GE Healthcare Life Sciences). Validation experiments for antisera are shown in **Suppl. Figure 7a**.

Antibodies and staining reagents for immunofluorescence

The following antibodies were used with the indicated titres for immunofluorescent staining: rabbit anti-PRP6 (Santa Cruz sc-48786) 1:100, mouse anti-PRPF8 (Santa Cruz sc-55534) 1:100, rabbit anti-PRPF8 (Santa Cruz sc-030207) 1:100, goat anti-PRPF31 (Abnova PAB7154) 1:100, PIBF1/CEP90 (Novus NBP2-19823) 1:100, rabbit anti-PLK4 (Cell Signalling Technologies 3258) 1:100, rabbit anti-GPR20 polyclonal (Thermo Scientific, PA3-068) 1:100, rabbit anti-SREB3 (Novus Biologicals NBP1-71064) 1:100, rabbit anti-MAS1 (Novus Biologicals

NBP1-78444) 1:100, rabbit anti-HTR1B (Novus Biologicals NBP1-01013) 1:100, rabbit anti-P2RY14 (Novus NBP1-70970) 1:100, rabbit anti-CRHR2 (Novus NBP1-00767) 1:100, rabbit anti-OPRL1 (Abcam ab66219) 1:50, rabbit anti-C21orf2 (GeneTex [N1C2], GTX119046) 1:100, mouse anti-acetylated alpha tubulin (Sigma Aldrich clone 6-11 B1 T6793) 1:1000, rat anti-ZO1 (1:400, Santa Cruz), rabbit anti- β -catenin (1:400, BD Bioscience), rabbit anti-gamma tubulin (AbCam ab11317) 1:1000, rabbit anti-ARL13B (Proteintech 17711-1-AP) 1:1000, mouse anti-GT335 monoclonal (Enzo Life Sciences ALX-804-885-C100) 1:1000 and mouse anti-FLAG (Sigma F3165) 1:1000. The mouse monoclonal antibody against centrin-3 has been described previously³⁶. Secondary antibodies were Alexa-Fluor-conjugated goat anti-mouse IgG and goat anti-rabbit IgG (Life Technologies Ltd.)

siRNA rescue experiments in mIMCD3

Rescue experiments followed our standard siRNA transfection and image analysis protocols for pooled siRNA mouse siGENOME siRNAs (Thermo Scientific) in mIMCD3 cells for either *Pibf1* or *C21orf2*. Reverse transfections in 96-well assay plates (PerkinElmer View Plates) included 100ng of wild-type or missense mutated FLAG-tagged human PIBF1 or C21orf2 constructs. Our standard high-content image analysis protocol was adapted to include immunofluorescent staining with anti-FLAG to enable quantitation of transfection efficiency.

Immunofluorescence and confocal microscopy of hTERT-RPE1 and mIMCD3 cells

hTERT-RPE1 cells were cultured as described above. To induce ciliogenesis, cells were seeded on coverslips and serum starved (culture medium with 0.2% FCS) for 24 hrs prior to transfection. Subsequently, Lipofectamine2000 (LifeTechnologies) was used to transfect cells with the appropriate vector according to the manufacturer's protocol. 24 hrs post-transfection, ciliated cells were fixed in 2% PFA for 20 minutes, treated with 1% Triton X-100 in PBS for 5 minutes, and blocked in 2% bovine serum albumin (BSA) in PBS for 30 minutes. Fixed cells were stained for 1 hour with the corresponding primary antibodies: rabbit anti-FLAG (Sigma-Aldrich) 1:1000 and mouse anti-acetylated alpha tubulin (Sigma-Aldrich) 1:1000. Coverslips

were then washed in PBS and stained for 45 minutes with AlexaFluor-405 or AlexaFluor-647-conjugated secondary antibodies (1:500, LifeTechnologies). Coverslips were washed again with PBS and briefly with milliQ water before mounting in Vectashield with or without DAPI (Vector Laboratories). mIMCD3 cells were seeded at 2.5×10^5 cells/well on sterile glass coverslips in six-well plates and fixed in ice-cold methanol. Cells were blocked in 1% non-fat milk for 5 min each. Immunofluorescence staining was performed as described previously³⁷.

Confocal images were obtained using a Zeiss Axio Imager ZI fluorescence microscope (Zeiss), equipped with a 63x objective oil lens, or a Nikon A1R confocal microscope with x100 oil objective lens controlled by NIS-Elements AR 4.20.01 (Nikon) software. Optical sections were generated through structured processed using Axiovision 4.3 (Zeiss) or NIS-Elements AR 4.20.01 (Nikon) software. Images were analysed using Adobe Photoshop CS and FIJI software. Images were assembled with Adobe Illustrator CS.

mIMCD3 spheroid assay

1×10^4 mIMCD3 cells per well were seeded in a 48 well-plate and transfected using Lipofectamine RNAiMax (LifeTechnologies) according to the manufacturer's protocol. Cells were trypsinized 24 h post-transfection and resuspended cells were then mixed 1:1 with growth factor-depleted Matrigel (BD Bioscience) in Labtech chamber slides. After polymerization for 20 min at 37°C, warm medium was dripped over the matrix until just covered. The mIMCD3 cells formed spheroids with cleared lumens 3 days later. Medium was removed by pipetting and the gels were washed three times for 10 min with warm PBS supplemented with Ca^{2+} and Mg^{2+} . The gels were then fixed in fresh 4% PFA for 30 min at room temperature. After washing three times in PBS, the cells were permeabilized for 15 min in gelatin dissolved in warm PBS (350 mg/50 ml) with 0.5% Triton X-100. Primary antibody (mouse anti-acetylated tubulin, Sigma, 1:20,000), rat anti-ZO1 (1:400, Santa Cruz) and rabbit anti- β -catenin (1:400, BD Bioscience) was diluted in permeabilization buffer and incubated overnight at 4°C. After washing the spheroids three times for 30 min in permeabilization buffer, goat anti-mouse-Cy5, donkey anti-rabbit-TRITC, and goat anti-rat-FITC secondary antibodies (1:500, LifeTechnologies) were each diluted in

permeabilization buffer and incubated with the spheroids for 4 hr at room temperature. Spheroids were washed three times in permeabilization buffer for 10 min per wash and then incubated for 1 hr with DAPI, before being washed an additional three times in PBS, and finally mounted in Fluoromount-G (Cell Lab, Beckman Coulter). Images and z-stacks were taken with a Zeiss LSM700 confocal microscope. Cilia were counted manually from maximum projections of z-stacks, with axonemes $>1\mu\text{m}$ in length scored as positive, as described previously³⁷.

Confocal immunofluorescence microscopy and high resolution immunofluorescence analysis of mouse photoreceptor cells

Images were obtained using a Nikon Eclipse A1 scanning confocal microscope, controlled and processed using Nikon Elements software. For high magnification and high resolution immunofluorescence analysis eyes of C57Bl/6J mice were cryofixed, sectioned, and immunostained as previously described^{38,39} for two independent experiments. Cryosections were double stained for PRPF6, PRPF8 (A5463), GPR20, MAS1, SREB3, HTR1B (1:200) and centrin-3 (1:100) as a molecular marker for the connecting cilium, the basal body, and the adjacent centriole of photoreceptor cells⁴⁰. Subsequently, cryosections were incubated with the according secondary antibodies conjugated to Alexa 488 or Alexa 555 (LifeTechnologies), counterstained with DAPI (Sigma-Aldrich), and mounted in Mowiol 4-88 (Hoechst). Specimens were analysed and deconvoluted in a Leica LEITZ DM6000B microscope (Leica). Images were processed with Adobe Photoshop CS using different tools including contrast and colour correction as well as pixel extrapolation.

Immunoelectron microscopy of mouse photoreceptor cells

For immunoelectron microscopy of mouse photoreceptor cells, the pre-embedding labelling protocol was used, as described previously^{41,42}. Ultrathin sections were analysed in a transmission electron microscope (TEM) (Tecnai 12 BioTwin, FEI). Images were obtained with a charge-coupled device camera (SIS MegaView3, Olympus) and processed with Adobe Photoshop CS.

Transmission electron microscopy on *C. elegans* amphid channel cilia

TEM was performed on the amphid channel cilia of strain MR247: *prp-8(rr40);cdc-25.1(rr31)unc-13(e450);rrIs01*. This strain was obtained in a forward genetic screen for suppressors of the cell cycle phosphatase *cdc-25.1(rr31)*, which is a dominant maternal effect gain-of-function missense allele (G47D) sensitizing the intestinal lineage to an extra cell division⁴³. The *rr40* suppressor is a homozygous donor splice site mutation (hypomorphic) in intron 8 of *prp-8*. MR247 also contains an undefined (likely missense) mutation (*e450*) in the neurotransmitter vesicle release regulator *unc-13* (used to map the *prp-8(rr40)* mutation), and an integrated intestinal GFP marker (*rrIs01*)⁴⁴. Although the *prp-8* splice-site mutation is the likely cause of the ciliogenesis defect in this strain, we cannot formally exclude the possibility that a dominant missense mutation in *cdc-25.1* or an undefined (likely missense) mutation in *unc-13* could be causative for these defects. Young adult worms were processed and imaged essentially as described previously⁴⁵, except that MR247 worms were fixed overnight at 4°C in 2.5% gluteraldehyde and 1% paraformaldehyde in Sørensen's buffer.

Confirmation of siRNA knockdown by qRT-PCR

QRT-PCR was conducted as previously described⁴⁶. The following primers were used:

Plk4: 5'-CGTTGTCACTCAGTAGAAATGCT, 5'-GGTGTCTGGTCTGCAAATGG; *Prpf6*: 5'-AAGCTGACTCCTGTTCTGA, 5'-TCATTCCACCTGGGTATGGAG; *Prpf8*: 5'-ACTGGAAGCCATTCAGCTAGAG, 5'-AATGTGAACTGCCAGCGTTG; *Prpf31*: 5'-TATCAGCAAGCAAGCCAACG, 5'-CTATCTCCACAGTCAGGTTGTTG; *Prpf38a*: 5'-AAGAGCCAGAACCGAAATGG, 5'-GACTCTCTCACTGTGCAGGA; *Pibf1*: 5'-CCCAGAAAAACATGATGATCGAC, 5'-CTAATCGCAAAGTTAGCAGCTG; *C21orf2*: 5'-ACCTGCTGAGCTACACAGAG, 5'-TTCTTGTGACTCCTCATCGTGT; *β-actin*: 5'-GCTTCTTTGCAGCTCCTTCGT, 5'-AGCGCAGCGATATCGTCAT.

Western blotting

Whole cell extract and western blots were conducted as previously described⁴⁶. Antibodies were used at the following concentrations: rabbit anti-PRPF6 (Santa Cruz sc-48786) 1:500, mouse anti-PRPF8 (Santa Cruz sc-55534) 1:200, rabbit anti-PRPF8 (Santa Cruz sc-030207) 1:200, goat anti-PRPF31 (Abnova PAB7154) 1:500, rabbit anti-PLK4 (Cell Signalling Technologies 3258) 1:500, rabbit anti-C21orf2 1:500 and rabbit anti-NEK1 (H3T1; validated in **Suppl. Figure 8a**) 1:2000 incubated overnight at 4°C; and mouse anti-FLAG (Sigma Aldrich clone M2 F9291) 1:2000, mouse anti-HA (Sigma Aldrich H9658) 1:1000 and mouse anti-β actin (Abcam ab6276) 1:5000 incubated for 1 hour at room temperature. Membranes were incubated with HRP-conjugated secondary antibodies (Dako Cytomation P0447 and P0448) for 1 hour at room temperature and immunopositive bands visualized with Supersignal West Femto chemiluminescent substrate (Pierce) on a BioRad ChemiDoc MP system. Full unprocessed western blotting scans are shown in **Suppl. Figure 9**, with the regions used in main display items indicated by red frames.

Zebrafish experiments

Zebrafish husbandry was undertaken to UK Home Office Rules under standard conditions. Antisense morpholino oligonucleotide (MO; Gene Tools Inc.) for *nek1* (ENSDART00000035239 zv9) were designed against the splice donor sites of exon 6 and exon 8 (exon 6 sequence MO1: AGCATTAAAACACATCTGACCTATT, exon 8 sequence MO2: GATAAGAAAGACAAGCGCACCTTCC). MO injection into embryos was performed at the one- to two-cell stage and embryos were incubated at 28.5°C. Effective doses were defined as the MO concentration that resulted in the least dead fish and the most fish with the phenotypes of interest. *nek1* knockdown resulted in a marked embryonic lethality of about 25% even at low doses due to developmental defects during gastrulation. MO specificity was confirmed by RT-PCR (**Suppl. Figure 7c**). RNA was extracted at 48 hpf using TRIzol (LifeTechnologies). First-strand complementary DNA was synthesized using random nonamers (Sigma-Aldrich) and Omniscript transcriptase (Qiagen). Standard PCR was carried out using primers in exons 4 and 10 for the *nek1* exon 6 splice MO (expected wild-type product size approx. 750 bp, intron 6/7

length 575 bp), with 60 bp. RT-PCR result suggesting the inclusion of the downstream intron (Suppl. Figure 7c). Amplification of *gapdh* was used as a quantification control. Primer sequences are available on request. For rescue experiments, 150 pg of human *C21orf2* RNA were co-injected with *nek1* MO and embryos assessed for phenotypic features 3 dpf by light microscopy. RNA was generated using the mMessage Machine kit (Ambion) according to the manufacturer's protocol.

For whole-mount immunofluorescence analysis, embryos were dechorionated and fixed in 100% methanol overnight at -20°C . After rehydration, embryos were washed several times in PBST (0.5% Triton X-100/PBS) then blocked in 5% BSA for 1 hour and incubated with anti-acetylated tubulin antibody (Sigma) overnight at 4°C followed by incubation with anti-mouse IgG 488 secondary antibody (Sigma) and staining with DAPI. Embryos were flat mounted on glass slides in Citifluor (Citifluor Ltd) and immunofluorescence analysis was performed using a Zeiss confocal microscope. Image analysis was performed using ImageJ software. Images represent z-stack maximum intensity projections. Alcian Blue staining for cartilage visualization was performed after over-night incubation of dechorionated embryos in 100% methanol at -20°C using a 1% Alcian Blue 70% acidic ethanol solution. Histology of the eye was performed using paraffin embedded embryos and a standard H&E staining protocol. Transgenic zebrafish lines included rhodopsin:GFP (in house heterozygous out-cross of *cu3Tg* to TULP long fin wild-type) and *smo* mutant (in-house heterozygote x heterozygote cross of *smo*^{b641/b641}).

Tandem affinity-purification and mass spectrometry

HEK293T (human embryonic kidney) cells were transfected with constructs encoding either *C21orf2*, or *NEK1*, or *SPATA7* fused to a strep-II tag/FLAG tag for tandem affinity purification (SF-TAP-tag) using polyethylenimine (PEI, PolySciences) as a transfection reagent. 48 hours after transfection the cells were lysed in lysis buffer containing 30mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5% Nonidet-P40 (NP40), freshly supplemented with protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany), phosphatase inhibitor cocktail 2 (Sigma-Aldrich) and PhosphataseArrest III (GBiosciences) for 20 minutes at 4°C . The streptavidin- and FLAG-based

tandem affinity purification steps were performed as previously described⁴⁷. 5% of the final elute was resolved by SDS-PAGE and analysed by silver staining according to standard protocols while the rest was subjected to protein precipitation with chloroform and methanol. Protein precipitates were subsequently subjected to mass spectrometry analysis and peptide identification as previously described⁴⁷.

Exogenous co-immunoprecipitation

Plasmids expressing N-terminal SF-TAP-tagged SPATA7 and N-terminal SF-TAP-tagged NEK1 were co-transfected with plasmids expressing HA-tagged wild-type C21orf2 or mutant constructs C21orf2-p.Arg73Pro or C21orf2-p.Leu224Pro in HEK293T cells. A plasmid expressing functionally unrelated HA-tagged LRR (leucine-rich repeat) domain of LRRK2 was used as a negative control. FLAG-NPHP4 and HA-RPGRIP1 (aa611-1286) were included as positive control as described previously⁴⁷. 24 hrs post-transfection, cells were lysed on ice in lysis buffer (50 mM Tris-HCl [pH7.5], 150 mM NaCl, 1% NP-40) supplemented with complete protease inhibitor cocktail (Roche Diagnostics). Lysates were incubated with either α -FLAG M2-agarose from mouse (Sigma-Aldrich) or anti-HA affinity matrix (Roche) for 2-3 hrs at 4°C. After incubation, beads with bound protein complexes were washed in ice-cold lysis buffer. Subsequently, 4x NuPAGE sample buffer was added to the beads and heated for 10 minutes at 70°C. Beads were precipitated by centrifugation, and supernatant was run on NuPAGE Novex 4%-12% Bis-Tris SDS-PAGE gels. After blotting overnight at 4°C, blots were stained with either rabbit anti-FLAG (Sigma, 1:1000), mouse anti-FLAG (Sigma, 1:1000), rabbit anti-HA (Sigma, 1:1000) or mouse anti-HA (Sigma, 1:1000). Goat anti-rabbit IRDye800 (Li-COR, 1:20,000) or goat anti-mouse IRDye800 (Li-COR, 1:20,000) were used as secondary antibodies. Fluorescence was analysed on a Li-COR Odyssey 2.1 infrared Scanner. Full unprocessed western blotting scans are shown in **Suppl. Figure 8** with the regions used in main display items indicated by red frames.

Endogenous co-immunoprecipitation from retinal extracts

Protein A/G PLUS-agarose beads (Santa Cruz Biotechnology) were washed three times in extraction buffer (10mM HEPES [pH7.9], 10mM NaCl, 3mM MgCl₂, freshly added 1mM DTT, 1mM Na₃VO₄) supplemented with complete protease inhibitor cocktail (Roche) and then coupled to polyclonal anti-NEK1 (H3T1) or polyclonal anti-FLAG (Sigma, F7425) overnight at 4°C. Coupled beads were washed three times in extraction buffer and incubated with fresh bovine retinal lysates overnight at 4°C. Retinas were homogenized by sonication on ice two times for 30 seconds in extraction buffer (10mM HEPES [pH7.9], 10mM NaCl, 3mM MgCl₂, freshly added 1mM DTT, 1mM Na₃VO₄) supplemented with complete protease inhibitor cocktail (Roche). Beads were washed five times with extraction buffer, transferred to microspin columns (GE Healthcare) and washed three times in washing buffer (1xTBS, 0.12% NP-40, supplemented with complete protease inhibitors (Roche)). Then, neutralization buffer (1M Tris/HCl, [pH8.0]) was added. Eluates were retrieved by adding 3x elution buffer (200 mM glycine [pH2.5]) for 20 min. at 4°C and subjected to protein precipitation with chloroform and methanol. Protein precipitates were subsequently subjected to mass spectrometry analysis and peptide identification as previously described⁴⁸.

GST-pulldown from retinal extracts

In order to produce GST fusion proteins, BL21-DE3 bacteria were transformed with either pDEST15-SPATA7, pDEST15-C21orf2 or pDEST15 alone. Protein expression was induced at 28°C for 3 hours with 0.5 mM IPTG (Sigma-Aldrich) and harvested by centrifugation. Subsequently, cells were lysed with STE buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA; and 150 mM NaCl) supplemented with 10 mg/ml lysozyme, 0.5% Sarkosyl, 1% Triton X-100, and Complete protease inhibitor cocktail (Roche). Lysates were then incubated for 2 hours at 4°C with glutathione-sepharose 4B beads (Amersham Biosciences). After incubation, beads were washed with STE and TBSTD (TBS with 1% Triton X-100 and 2 mM DTT). The amount of GST fusion proteins bound to the beads was verified on a NuPAGE Novex 4%–12% Bis-Tris SDS-PAGE gel by staining with SimplyBlue SafeStain (LifeTechnologies). Beads with bound GST-fusion proteins were then incubated over-night with bovine retinal extracts at 4°C. After

overnight incubation, samples were washed three times with extraction buffer. Protein were eluted from beads in 4X NuPAGE sample buffer and heated for 10 minutes at 70°C. Beads were precipitated afterward by centrifugation. The supernatants were run on NuPAGE Novex 4%-12% Bis-Tris SDS-PAGE gels, followed by immunoblotting overnight at 4°C and staining with rabbit polyclonal α -NEK1 as primary antibody (H3T1; 1:1000) and goat anti-rabbit IRDye800 (LiCOR, 1:20,000) as secondary antibody. Fluorescence was analysed on a Li-COR Odyssey 2.1 infrared Scanner. Full unprocessed western blotting scans are shown in **Suppl. Figure 8** with the regions used in main display items indicated by red frames.

For mass spectrometry analysis of the GST pull down, beads were transferred to microspin columns (GE Healthcare), and washed 2 times with filtered TBS(1X). Eluates were retrieved by adding 300 μ l of 100 mM reduced Glutathione (GSH) (Sigma-Aldrich) in 50mM Tris-HCl [pH=8.0] and incubating overnight at 4°C on an Intelli mixer. Eluates were then subjected to protein precipitation with chloroform and methanol. Protein precipitates were subsequently subjected to mass spectrometry analysis and peptide identification as previously described⁴⁸.

Please refer to **Supplementary Note** for methodological details of the primary, secondary and tertiary screens.

Please refer to **Supplementary Note** for methodological details of patient ascertainment and sample preparation, and genetic analyses including SNP genotyping arrays, linkage analysis, whole exome sequencing and Sanger sequencing.

REFERENCES

35. Thiel, C. *et al.* NEK1 mutations cause short-rib polydactyly syndrome type Majewski. *Am J Hum Genet* **88**, 106-14 (2011).
36. Giessl, A., Pulvermüller, A., Trojan, P., Park, J.H., Choe, H.W., Ernst, O.P., Hofmann, K.P., Wolfrum, U. Differential expression and interaction with the visual G-protein transducin of centrin isoforms in mammalian photoreceptor cells. *J Biol Chem.* **279**, 51472-81 (2004).
37. Dawe, H.R. *et al.* Nesprin-2 interacts with meckelin and mediates ciliogenesis via remodelling of the actin cytoskeleton. *J Cell Sci* **122**, 2716-2726 (2009)..
38. Arts, H.H. *et al.* Mutations in the gene encoding the basal body protein RPGRIP1L, a nephrocystin-4 interactor, cause Joubert syndrome. *Nat Genet* **39**, 882-8 (2007).
39. Overlack, N. *et al.* Direct interaction of the Usher syndrome 1G protein SANS and myomegalin in the retina. *Biochim Biophys Acta* **1813**, 1883-92 (2011).
40. Trojan, P. *et al.* Centrins in retinal photoreceptor cells: regulators in the connecting cilium. *Prog Retin Eye Res* **27**, 237-59 (2008).
41. Maerker, T. *et al.* A novel Usher protein network at the periciliary reloading point between molecular transport machineries in vertebrate photoreceptor cells. *Hum Mol Genet* **17**, 71-86 (2008).
42. Sedmak, T., Sehn, E. & Wolfrum, U. Immunoelectron microscopy of vesicle transport to the primary cilium of photoreceptor cells. *Methods Cell Biol* **94**, 259-72 (2009).
43. Hebeisen, M., Drysdale, J. & Roy, R. Suppressors of the cdc-25.1(gf)-associated intestinal hyperplasia reveal important maternal roles for prp-8 and a subset of splicing factors in *C. elegans*. *RNA* **14**, 2618-33 (2008).
44. Kostic, I. & Roy, R. Organ-specific cell division abnormalities caused by mutation in a general cell cycle regulator in *C. elegans*. *Development* **129**, 2155-65 (2002).
45. Cevik, S. *et al.* Joubert syndrome Arl13b functions at ciliary membranes and stabilizes protein transport in *Caenorhabditis elegans*. *J Cell Biol* **188**, 953-69 (2010).

46. Huang, L. *et al.* TMEM237 is mutated in individuals with a Joubert Syndrome related disorder and expands the role of the TMEM family at the ciliary transition zone. *Am J Hum Genet* **89**, 713-30 (2011).
47. Roepman, R. *et al.* Interaction of nephrocystin-4 and RPGRIP1 is disrupted by nephronophthisis or Leber congenital amaurosis-associated mutations. *Proc Natl Acad Sci U S A* **102**, 18520-5 (2008).
48. Coene, K.L. *et al.* The ciliopathy-associated protein homologs RPGRIP1 and RPGRIP1L are linked to cilium integrity through interaction with Nek4 serine/threonine kinase. *Hum Mol Genet* **20**, 3592-605 (2011).

SUPPLEMENTARY NOTE

Whole genome siRNA-based reverse genetics screen for ciliogenesis

Whole genome siRNA reagents: A Dharmacon mouse siGENOME siRNA library (Thermo Scientific) was used, plated in a standard 96-plate well format, consisting of 76269 individual siRNA duplexes. Each gene was targeted by a pool of four duplexes per well, across 40 sub-library plates, 89 drug target sub-library plates, and 114 plates targeting the remaining genes in the mouse genome (the genome sub-library). After the removal of 38 duplicated reagents, the entire library targeted a total of 19059 unique Entrez RefSeq (B37.1) genes across 243 plates, of which 18960 also targeted a unique Ensembl (r69) mouse gene. Library plates contained 80 targets per plate across columns 2-11. Lyophilised siRNA was re-constituted with 1X RNA buffer (Thermo Scientific) to a concentration of 2 μ M using a Fluid-X XRD-384 dispenser, followed by 90 minutes of agitation on a rotary shaker. Reconstituted siRNA plates were stored at -80°C. Thawed siRNAs were directly aliquoted into 96-well assay plates (PerkinElmer View Plates) using an Agilent Bravo liquid handling platform. Eight different control siRNAs (siGENOME, Thermo Scientific) at 2 μ M were added to plates using a Star-pet-E 8-channel electronic pipette (in duplicate to columns 1 and 12) for a 50nM final concentration of pooled siRNAs. Each 96-well plate contained six positive assay controls (siRNA pools for *Mks1*, *Rpgrip11* and *Ift88* each in duplicate; **Suppl. Table 10**) that had significant and reproducible inhibitory effects on ciliogenesis, and one positive transfection control (duplicate siRNA pool for *Plk1*, a gene essential for cell growth and proliferation) that significantly affected cell number (**Figure 1c**). We also included eight negative control siRNAs (two non-targeting or “scrambled” siRNAs, an siRNA targeted against human motilin receptor (*MLNR*) that has no sequence orthologue in the mouse genome (**Suppl. Table 10**), and a mock-transfection control. These controls had no significant effect on the percentage of cells with a single cilium (**Figure 1c,d**), as previously reported¹. The strictly standardised mean difference (SSMD)^{2,3} of the percentage of cells with a single cilium in *Mks1*, *Ift88* and *Rpgrip11* knockdowns, compared to all negative controls was consistently >1.5, reflecting the

suitability of these siRNAs as positive controls for effects on ciliogenesis (data not shown). Sequences of control siRNA duplexes are given in **Suppl. Table 10**. For the validation screen unpooled duplexes of Dharmacon ON-TARGET-PLUS siRNAs were used at a final concentration of 12.5nM.

Whole genome siRNA transfections: Reverse siRNA transfections were set up in batches of 20 plates. 2.5µl siRNA were added per well to optical-bottomed 96-well Perkin Elmer View Plates, followed by 0.2µl Lipofectamine RNAiMAX transfection reagent (LifeTechnologies) suspended in 17.5µl Opti-MEM serum free media (Invitrogen) using a FluidX XRD-384 dispenser at high speed (300rpm). Plates were gently mixed on a rotary shaker at room temperature for a minimum of 20 minutes. 80µl mIMCD3 cells at a density of 10^5 cells per ml in Opti-MEM, in a uniform suspension maintained by constant magnetic stirring, were added per well to plates using an XRD-384 FluidX dispenser at high speed. To minimize edge effects, cells and transfection complexes were left to rest in a laminar flow hood for 1 hour at room temperature before returning to a 37°C 5% CO₂ incubator without further changes of media in wells. Cells were assayed after 72 hours, at which time they formed a uniform monolayer of 95-100% confluency. This confluency was optimal for the detection of cilia at the same single focal plane in every well, since over-confluent cells grew in multiple layers. The proportion of ciliated cells was consistently above 60% which is comparable with previous reports¹. Hits confirmed in the secondary screen were validated in scaled-up format following a previously described protocol¹.

High-throughput liquid handling: Plates were processed for immunofluorescent staining in batches of 20. 72 hours after transfection, media was removed from plates by inverting the plate and blotting on clean paper towels to remove excess liquid. Cells were washed three times with 100µl PBS, using two XRD-384 FluidX dispensers working in parallel on slow speed (100rpm), dispensing to the left side of the well to minimise disturbance of cells. After fixation, all cell washes were performed at medium speed (150rpm). After incubation with primary antibody, cells were washed once using 100µl PBS on a XRD-384 FluidX dispenser and three times on a Biotek ELx405 Select CW plate-washer with a Biostack

automated plate stacker that dispensed against the left side of the well. After incubation with secondary antibody and stains, cells were washed once using 100µl PBS on a XRD-384 FluidX dispenser and six times on a Biotek ELx405 Select CW plate-washer with a Biostack automated plate stacker. Plates were stored with 100µl PBS per well at 4°C for up to 1 week before imaging.

Fixation: Optimisation experiments proved that ice-cold methanol fixation gave much better resolution of cilia staining and lower background compared to other fixation methods in mIMCD3 cells. The success of methanol fixation when optimized for high-throughput depended on maintaining the methanol temperature on dry ice, pre-chilling the dispenser tubing prior to dispense, and placing plates into a -20°C freezer for exactly 5 minutes. Cells were fixed for exactly 5 min. using ice-cold methanol. Methanol was removed from a -20°C freezer and kept on dry ice whilst passing through the dispenser tubing (which was placed between two frozen ice blocks) for 1 minute to ensure the temperature of methanol being dispensed into plates was maintained. 50µl ice cold methanol was dispensed to the left side of each well using FluidX XRD-384 dispenser on slow speed (100rpm), taking ca. 30s to dispense the methanol per plate. Plates were then placed in a -20°C freezer on top of frozen ice blocks for exactly 5 min. After fixation, each plate was removed from the freezer, inverted and blotted to remove methanol. A second pre-primed dispenser was used to dispense 50µl PBS into each well of the plate. The same speeds and volumes were used on the two dispensers to ensure that each plate was fixed in methanol for the same length of time.

Antibodies and staining reagents for immunofluorescent staining: DAPI was used with the far-red RNA stain TOTO3 (LifeTechnologies) to determine cell, cytoplasm and nuclear boundaries. Cilia were detected with a mouse monoclonal antibody against acetylated α -tubulin, a cilia marker with specificity for the entire length of the primary ciliary axoneme and low background staining (clone 6-11 B-1, catalogue number T6793, lot no 070M4755, Sigma-Aldrich) and an Alexa Fluor 488-conjugated secondary antibody (LifeTechnologies). Optimization experiments in pilot screens showed that separate primary and secondary

antibody binding steps were essential for high-quality cilia staining. Plates were blocked with 100µl 1% non-fat milk powder/PBS (w/v) previously clarified of particulates by centrifugation at 3000 x g for 5 min. All antibodies and stains were diluted in blocking solution and clarified by centrifugation at 12,000 x g for 5 min. at room temperature. Primary antibody (anti-acetylated alpha tubulin) was diluted x1000 and cells were incubated with 50µl antibody solution per well for one hour at room temperature. Secondary antibody (x2000 dilution), and DAPI and TOTO-3 (both x5000 dilutions) were incubated for one hour at room temperature in the dark.

High-throughput high content imaging: Plates were processed using the PerkinElmer Operetta high-content wide-field fluorescence imaging system, coupled to Harmony software. Plates were automatically loaded onto the Operetta using a PerkinElmer Plate Handler II robotic arm, operated through Plateworks software. Plates were bar-coded with specific plate information, and barcodes were read by a SICK barcode reader before loading onto the Operetta. Wells were imaged using a x20 objective lens, detecting three colours in three separate focal planes to ensure the best resolution was obtained for each colour. The bottom of each well was detected automatically by the Operetta infra-red focusing LED, and focal planes of detection for each colour were calculated relative to this value. DAPI emission was detected for 60ms, 6µm above the calculated bottom of the well. TOTO3 emission was detected for 500ms, 1µm above the DAPI plane, and Alexa Fluor 488nm emission was detected for 700ms 7µm above the DAPI plane (**Figure 1a**). Six fields of view (each 510x675µm) were imaged per well (with an identical pattern of fields in every well away from the dispense area), with an approximate total of 4000 cells detected and analysed per well.

Image analysis: Modified PerkinElmer image analysis algorithms were used throughout. Cilia in assay plates remained vertical and therefore appeared as discrete dots 1-5µm in diameter (**Figure 1b**). This enabled us to develop highly efficient cilia detection protocols based on pre-designed PerkinElmer 'find spots' algorithms. Cilia were consistently identified at a focal plane above the apical cell surface, 7µm above the focal plane of the nucleus.

Cells were stained with DAPI and TOTO-3, allowing accurate segmentation of cells (**Figure 1a**). This methodology highly improves specificity for high-throughput cilia detection when compared to previously reported HTS cilia assays⁴. Nuclei were detected as blue (DAPI) fluorescent regions $>30\mu\text{m}^2$, with contrast >0.10 . Cytoplasm was detected as far-red (TOTO3) fluorescent regions around nuclei. Border objects were excluded so that only whole cells were analysed. Cilia on whole cells were detected using a modified 'find spots' algorithm, identifying green (Alexa Fluor 488) fluorescent spots with radius <3.8 pixels, contrast >0.1 , spot-to-region intensity >1.3 with a distance from all neighbouring spots >5.6 pixels. Any cells with more than one cilium in the cell area were excluded from analysis. Key output parameters were number of whole cells, and percentage of whole cells with a single cilium. Additional output parameters included number of whole cells with two or more cilia and mean cilium intensity, which could be used as an indicator of cilium length.

Screen quality control and statistical analyses: The strictly standardised mean difference (SSMD) of the percentage of cells with a single cilium in *Mks1*, *Ift88* and *Rpgrip11* knockdowns, compared to all negative controls was consistently >1.5 reflecting the suitability of these siRNAs as positive controls for effects on ciliation, and the consistency and quality of the screen. For cilia length measurements, normal distribution of data was confirmed using the Kolmogorov-Smirnov test (GraphPad Software). Pair-wise comparisons were analysed with Student's two-tailed t-test or one-way analysis of variance (ANOVA) using InStat (GraphPad Software). Significance of pair-wise comparisons indicated as: ns not significant, * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$. Results reported are from at least three independent biological replicates.

GESS analysis: To further control for false positive effects in the screen, we performed GESS analysis⁵ to identify seed sequences which were enriched in our hit list and thus likely to be due to non-specific microRNA-like effects. We identified a total of 79 genes in the whole genome library (**Suppl. Table 1**) and 14 hits from the primary screen (**Suppl. Table 2**) that had enriched seed sequences. These were excluded from further analysis.

Cell line characterisation: To control for false positive effects in the screen, we performed extensive characterisation of mIMCD3 cells, to study whole genome copy number and whole transcriptome expression level.

Array CGH: Array CGH analysis identified 5274 genes with abnormal copy number in mIMCD3 cells, including three biallelic losses. The three genes with biallelic losses were excluded from analysis. None of the 1940 siRNA hits in the ciliogenesis screen had biallelic losses, but 711 were found to have abnormal copy number. No results were filtered or weighted on the basis of this data, but it may be useful to do so in future analyses. Other possible sources of false positive and negative hits are sequence-independent off-target effects due to siRNA competition with/saturation of endogenous RNAi, especially if the cell line used is under-expressing rate-limiting components of RNAi: exportin-5, TRBP, Dicer, Ago1-4. To control for these effects we confirmed normal copy number of rate-limiting components of RNAi in mIMCD3 cells by array CGH. *Xpo5* had normal copy number, *Tarbp2* +1 gain, *Dicer1* +1 gain, *Eif2c1* no info but very close to *Eif2c3* and *Eif2c4* on mouse chromosome 4, (these both have normal copy number therefore *Eif2c1* likely to have normal copy number), *Eif2c2* +1 gain, *Eif2c3,4* normal. Importantly, there were no genomic losses in these crucial parts of RNAi machinery.

BLASTN alignment: We undertook BLASTN alignment analysis of the entire Dharmacon mouse siGENOME siRNA library to identify siRNAs with sequence-specific off-target effects and pools of siRNAs that did not target all transcripts of a gene. Of the 19059 unique Entrez RefSeq genes in the entire library, 832 were found to have off-target mappings. These genes were excluded from our analysis, including n=133/2174 hits from the primary screen (**Suppl. Table 2**). Potential false negative results may arise from incomplete targeting of mRNAs by siRNAs in the Dharmacon siGENOME siRNA pools. BLASTN alignments of siRNA sequences identified 2031 siRNAs that only partially target their target gene, with at least one non-targeted transcript per gene (**Suppl. Table 1**). These siRNAs were not excluded from our screen analysis, but weighting could be applied to effects seen in these knockdowns to account for the only partial targeting of genes. BLASTN analysis confirmed

that the *Plk1*, *Mks1*, *Rpgrip1l* and *Ift88* siRNA pools targeted every Ensembl annotated and protein coding alternative transcript of the target genes, with the exception of one incomplete *Rpgrip1l* transcript (ENSMUST00000132757) encoding a predicted 78 amino acid isoform of the protein.

DAVID analysis: To minimize false negative results from the primary screen list, we performed functional annotation clustering and enrichment analysis for either BioCarta and KEGG pathways or gene ontology (GO) terms using DAVID^{6,7}. We selected a total of 373 hits that had human orthologues and were part of a pathway that was significantly enriched based on a Benjamini-Hochberg adjusted p value <0.05 . Based on functional categorization and relevant literature information, we prioritized enriched components of the spliceosome, the proteasome and the ubiquitin-proteasome system (UPS). We excluded hits that could be due to non-specific secondary effects (*e.g.* associated with the ribosome, RNA polymerase, purine and pyrimidine biosynthesis), leaving a final number of 40 hits (secondary screen 1, **Suppl. Table 3**). The $n=154$ hits that passed the $Z_{\text{cell}} > Z_{\text{cell cutoff}}$ filter (secondary screen 2, **Suppl. Table 3**) were also analysed in DAVID, and these identified enrichment of GPCRs, the non-motile primary cilium and the photoreceptor outer segment.

Additional unvalidated data-sets: We also imaged other cellular phenotypes during data acquisition for the primary screen. These comprised modified PerkinElmer image analysis algorithms (available on request) to measure cilia intensity and pyknotic nuclear morphology to enable the assessment of cilia length and apoptosis. In brief, cilia intensity measurements used the modified 'find spots' algorithm described above. Apoptotic nuclei were detected as regions of DAPI staining with area $>15\mu\text{m}^2$ and $<50\mu\text{m}^2$, mean intensity >2000 , and roundness <0.5 . We include these numerical features expressed as robust z scores (**Suppl. Table 1**) to enable interrogation and further annotation of our screen dataset in future validation screens. We also note that $n=780/1829$ of the total hits from the primary screen also have an effect on ciliogenesis (Z_{cilia}) that is significantly greater ($z > -2$) than any effect on cell number (Z_{cell}) (**Suppl. Table 11**). Comparison of this "long list" of 780 genes against known ciliary components (the SYSCILIA Gold Standard⁸), revealed a

significant 2.3-fold enrichment of known ciliary components ($p = 1.0 \times 10^{-5}$ hypergeometric test, observed 29, expected 12.6). The “long list” therefore also comprises an additional resource for future validation of genes mediating ciliary processes or ciliopathy candidate genes.

Secondary siRNA screens in mIMCD3 and hTERT-RPE1 cells

Secondary validation screens were performed in mIMCD3 and hTERT-RPE1 cells to validate the hits from the primary siRNA screen. Screening in mIMCD3 cells consisted of duplicate siRNA knockdowns, using four unpooled siRNA duplexes with a different chemistry (Dharmacon ON-TARGET PLUS) from the primary screen (Dharmacon siGENOME) for a total of 194 genes (**Suppl. Table 3**), and assaying for effects on ciliogenesis defects using the same platform and analysis as the primary screen. The complete dataset for this screen is shown in **Suppl. Table 3** and consists of 68 genes.

Screening in hTERT-RPE1 cells consisted of single siRNA knockdowns for all 57 UPS hits identified in the primary siRNA screen using Silencer Select siRNA duplexes (LifeTechnologies). For transfection, siRNAs (15 nM final concentration) were plated in black MW96 plates (BD Falcon), Lipofectamine RNAiMax (LifeTechnologies) and Opti-MEM (LifeTechnologies) were added to the duplexes and incubated for 10-20 minutes according to manufacturer’s protocol to allow the formation of transfection complexes. hTERT-RPE1 cells (2.5×10^4 cells per well) were then plated in the MW96 plates. Per plate, three non-targeting duplexes and three *PLK1* duplexes were included as negative controls or transfection controls, respectively. After 24 hours of transfection, cells were serum-starved (DMEM/Ham’s F12 with 0.2% FCS) to induce ciliogenesis. After 72 hours of transfection, cells were washed with PBS, and fixed with 2% paraformaldehyde for 20 min. After fixation, cells were washed briefly in PBS, permeabilized in 1% Triton-X/PBS for 5 minutes and blocked in 2% BSA/PBS for 30 minutes. After blocking, cells were stained with anti-acetylated tubulin (Sigma, clone 6-11B-1; 1:1000) and anti-RPGRIP1L⁹ (SNC040; 1:500) for one hour. Then, cells were washed in PBS (3x5 min) and stained with goat anti-mouse

Alexa Fluor 488 (LifeTechnologies; 1:500), goat anti-guinea pig Alexa Fluor 568 (LifeTechnologies; 1:500) and DAPI (LifeTechnologies; 1:8000) for 45 minutes. Finally, cells were washed in PBS (3x5 min) and stored in PBS at 4°C prior to microscopy analysis. Microscopy analysis of this screen was performed using a Leica DMI-6000B high-content microscope. Multiple focal plane images were captured from four regions per well for each channel. Data was analyzed automatically to define the number of nuclei, number of cilia and cilia length using FIJI-based plug-ins (FindFocus and CiliaMeasurement plug-ins) which were developed specifically for these experimental conditions. First, for each image the focal plane for analysis was determined using the FindFocus plug-in. Second, the CiliaMeasurement plug-in was applied to determine the number of nuclei and cilia, and length of cilia. Edged objects of each image were excluded from the analysis. Data for each image were then exported to Excel and analyzed for significant effect on ciliated cell per well (sum of four images), using robust z scores for cilia length ($z_{\text{cilia length}}$) and for % cells with a single cilium (z_{cilia}). The results from this screen are shown in **Suppl. Table 4**.

Tertiary siRNA screen in hTERT-RPE1 cells

In the tertiary screen, we re-screened the n=68 gene hits validated in the secondary screen using pooled Dharmacon ON-TARGET PLUS siRNAs in hTERT-RPE1. Transfection and image analysis methods were the same as for the validation siRNA screen in hTERT-RPE1, listing robust z scores listed for cilia length ($z_{\text{cilia length}}$) and for % cells with a single cilium (z_{cilia}). The complete datasets for this screen are shown in **Suppl. Table 3**.

Patient ascertainment and sample preparation

All samples used in this study were obtained with informed consent according to the protocols approved by the ethical committees of the Institute of Child Health/Great Ormond Street Hospital (REC ref. no. 08/H0713/82), South Yorkshire Research Ethics Committee (11/H1310/1), King Faisal Hospital, Saudi Arabia (RAC# 2070 023), and the University of Washington, Seattle WA (Protocol #28853). Inclusion criteria for Joubert syndrome were:

(1) molar tooth sign or cerebellar vermis hypoplasia and superior cerebellar dysplasia on brain imaging, and (2) clinical findings consistent with JBTS (cognitive impairment, hypotonia and ataxia). Imaging findings for *PIBF1*-mutated individuals ranged from the classic molar tooth sign to moderate vermis hypoplasia with mildly thick superior cerebellar peduncles and characteristic superior cerebellar dysplasia (**Figure 6a-c, Suppl. Figure 3, Suppl. Table 6**). One individual had peri-Sylvian polymicrogyria, and several had prominent perivascular spaces. All individuals for whom information was available had ataxia and developmental delay, ranging from mild to moderate (**Suppl. Table 7**). Three had oculomotor apraxia, and abnormal breathing pattern was noted in only one. None of the affected individuals had other features commonly associated with JBTS including polydactyly, coloboma, retinal dystrophy, nephronophthisis and liver fibrosis.

Inclusion criteria into the UCL Whole Exome Study was clinical diagnosis of Jeune syndrome based on typical radiological features such as short ribs, narrow thorax, handlebar clavicles, trident acetabulum, cone shaped epiphyses and other typical clinical symptoms such as renal, liver and retinal involvement as previously described in the literature. Inclusion criteria for primary ophthalmological patients was clinical diagnosis of retinal degeneration with assignment to a specific clinical subtype (rod-cone dystrophy) based on clinical and, in selected cases, electrophysiological assessment. Genomic DNA was isolated from peripheral venous blood or saliva by standard salt extraction. Total RNA was extracted from muscle tissue with the TRIzol method (LifeTechnologies).

All JBTS patients were tested for linkage to known JBTS disease genes, or had mutations excluded by direct Sanger sequencing, MIPS or WES. Founder pathogenic variants in the two known Hutterite JBTS genes *TMEM237*¹⁰ and *CSPP1*^{11,12} were excluded from Hutterites.

SNP genotyping arrays and linkage analysis

WES and homozygosity mapping were used in combination to identify the rare homozygous missense variant in JBTS individual H1-3, which was surrounded by a 15.6 MB region of

absence of heterozygosity on chromosome 13. For family GC4693, all five affected siblings and the unaffected mother were genotyped using a 50k Xba SNP Array (Affymetrix Inc.) according to the manufacturer's protocol. As there was no skeletal nor retinal disease in parents or other relatives, data were analyzed assuming an autosomal recessive model in which all five siblings would share identical genotypes around the causative gene. Linkage was excluded for regions surrounding known candidate retinal dystrophy genes. Seven regions over 1Mb were consistent with linkage. There were no significant regions of shared autozygosity.

Genetic analyses including whole exome sequencing

DNA isolation and exome sequencing for samples CR-F024.1 and F2.1 have been previously published¹³. For all other individuals, DNA was isolated from human EDTA blood samples using Genra Puregene Blood kit (Qiagen) according to the manufacturer's protocol. 5µg of DNA were used for whole exome sequencing and processed using a standard protocol as previously described¹⁴. Samples UCL111.1 and UCL111.2 were processed as part of the UK10K project (www.uk10k.org) using the Agilent V2 Sure Select 50 MB whole exome kit. Sample UCL78.1 was sequenced (Axeq Technologies) using Nimblegen V3 whole exome kit. DNA samples with a 'UCL' prefix were analysed for CNVs using the ExomeDepth Program¹⁵. In brief, 50 bp reads were aligned to hg19 reference genome and followed by variant calling (Lifescape V.2.1 software). Variants from all samples were annotated and prioritised to identify pathogenic mutations as previously described^{14,16}. Inclusion criteria for variants consisted of being covered by at least 10 sequence reads wherein the variant must be present in at least 20% of the reads (variants detected in 20%–80% of the sequence reads were considered heterozygous). Variants annotated in dbSNP132, the 1000 Genomes project, the Seattle exome database (<http://www.evs.gs.washington.edu>) or our UCL in-house database with a minor allele frequency (MAF) >0.5%, were removed as were all variants outside of known coding genetic sequencing and up to 20 bp intronic around exons.

For the siblings UCL111.1 and UCL111.2, we further filtered for variants present in both individuals using Exome Variants Analysis and Reporting (EVAR) software (<http://www.exome.info>). An autosomal recessive inheritance model was applied for gene identification, with known JATD and SRPS genes manually visually analysed (Integrative Genomics Viewer; <http://www.broadinstitute.org/igv/>).

DNA from subject GC4693.1 was analyzed by WES, using the Illumina TruSeq Exome Enrichment Kit and run on an Illumina HiSeq2000 sequencer. Reads were aligned to the human reference sequence (UCSC Genome Browser hg19) with Novoalign (Novocraft) version 2.05. The ANNOVAR tool (OpenBioinformatics) was used for annotating SNPs and small indels. Candidate pathogenic variants were validated and assessed for familial segregation by Sanger sequencing.

The Hutterite variant p.Asp637Ala in *PIBF1* is predicted to be deleterious by Polyphen2 (0.999/0.996) but tolerated by SIFT, whilst p.Arg405Gln is predicted deleterious by Polyphen2 (1.0/0.996) and damaging by SIFT¹⁷. The overall minor allele frequency for p.Arg405Gln in all ethnic groups for ExAC datasets is 1.416% (rs17089782; 1720/122044 alleles, 79 homozygous genotypes reported). The p.Asp637Ala variant was absent in all ESP or ExAC datasets, however in 550 North American Hutterite individuals from all three Leut groups it was identified as heterozygous in 23 and homozygous in one 57 year old woman for whom no other medical information was available. In whole genome sequence data for a cohort of 98 Schmiedeleut Hutterite individuals, the variant was found to be heterozygous in 7, and homozygous in one 8 year old boy who scored as 10-25 percentile on the Raven's Progressive Matrices¹⁸ and had night-time incontinence but was not noted to have other medical issues (no further information available). Using SNP genotyping data to impute the missense variant genotype in a total cohort of 1401 Schmiedeleut Hutterite individuals identified no additional individuals homozygous for the variant, despite an estimated carrier frequency of 7%¹⁹.

Targeted molecular inversion probe sequencing (MIPS) of *PIBF1* in a cohort of affected individuals from 353 additional families identified single heterozygous truncating

PIBF1 variants in five additional affected individuals. WES also identified two predicted deleterious heterozygous *PIBF1* variants (c.1669delC, p.Leu557Phefs*18 and c.673-?_1322+?del, for affected individual UW155-3 and for those in family MTI-121; **Suppl. Table 5**) that segregated with disease (**Suppl. Figure 2d-e**). Both deleterious variants were in combination with c.1214G>A p.Arg405Gln. No other heterozygous variants in *PIBF1* (including SNVs and CNVs) were identified in these individuals, suggesting that the combination of a unique truncating allele and a more common variant (c.1214G>A p.Arg405Gln) may be pathogenic²⁰. Although exogenous expression of human wild-type *PIBF1* rescued ciliogenesis in mIMCD3 cells following siRNA knockdown of endogenous *Pibf1*, expression of *PIBF1* containing either p.Asp637Ala or p.Arg405Gln was unable to rescue normal ciliogenesis suggesting that they are pathogenic missense mutations (**Suppl. Figure 4**).

The *C21orf2* p.Pro73Arg minor allele frequency in all ethnic groups for ExAC datasets is 0.03304% (rs140451304; 37/111976 alleles, no homozygous genotype reported). Although this variant is predicted benign by Polyphen2 and Mutationtaster, amino acid residue Pro73 is highly conserved in a leucine-rich repeat (LRR) domain (**Figure 6m**) and the substitution is non-conservative (Grantham score is >100). The p.Leu224Pro mutation is predicted damaging by Polyphen2 and Mutationtaster. No material was available to confirm the effect of the c.96+6T>A variant on splicing in individual UCL78.1 (**Suppl. Figure 5f**). However, Alamut software predicted that the c.96+6T>A variant caused a significant decrease in splicing probability and loss of a predicted SRp55 splicing enhancer binding site (**Suppl. Figure 5f**).

An affected sibling of UCL-111.1 and UCL111.2 carried the same mutations in *C21orf2* and had cone-rod dystrophy (**Figure 6k**) but with no signs of skeletal features of Jeune syndrome. This is consistent with a previous report that *C21orf2* is a candidate gene for non-syndromic retinal degeneration (cone-rod dystrophy), with homozygous variants found in two consanguineous families of Arabic origin¹³. The two individuals reported in Abu-Safieh et al. (2013)¹³ were therefore re-assessed as adults for any skeletal

involvement. This revealed that individual CR-F024.1 (c.103delA, p.Ile35Phefs*10; **Suppl. Figure 5c**) in addition to cone-rod dystrophy (**Figure 6l**) has a narrow thorax and very short stature but no cone shaped epiphyses (**Figure 6l,j, Suppl. Table 7**). The second individual F2.1 (c.545+1C>T, p.Ala181Glnfs*6 **Suppl. Figure 6e-f**), has no skeletal abnormalities (**Suppl. Table 7**). However, since individual F2.1 was examined at 4 years of age and early but mild skeletal Jeune syndrome phenotypes may regress in later life^{14,16,21,22}, it is possible that additional mild signs of Jeune syndrome were present neonatally. Further screening of *C21orf2* identified, in a non-consanguineous individual of northern European origin (GC4693.1), a homozygous missense change shared with the UCL-111 family (p.Arg73Pro). GC4693.1 has cone-rod dystrophy, severe scoliosis and hip dysplasia (**Suppl. Table 7**). The same homozygous mutation was found in his four younger siblings who are all affected with a similar phenotype but more severe retinal degeneration, and Sanger sequencing confirmed that this variant segregates with the disease (**Suppl. Figure 5b**). These findings indicate that mutations in *C21orf2* (**Figure 6m**) cause both syndromic and non-syndromic forms of retinal dystrophy, with a variable spectrum of skeletal involvement.

Sanger sequencing

Mutation verification was undertaken by performing PCR reactions under standard conditions and Sanger sequencing was performed at Source Bioscience, Cambridge, UK (PCR protocol and primer sequences are available on request). Sanger sequencing confirmed segregation of variant with phenotype in all cases in Schmiedeleut and Leherleut Hutterite JBTS families H2, H3 and H4 and in individual UW155-3.

SUPPLEMENTARY REFERENCES

1. Dawe, H.R. *et al.* Nesprin-2 interacts with meckelin and mediates ciliogenesis via remodelling of the actin cytoskeleton. *J Cell Sci* **122**, 2716-2726 (2009)..

2. Birmingham, A. *et al.* Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods* **6**, 569-75 (2009).
3. Zhang, X.D. A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. *Genomics* **89**, 552-61 (2007).
4. Lai, C.K. *et al.* Functional characterization of putative cilia genes by high-content analysis. *Mol Biol Cell* **22**, 1104-19 (2011).
5. Yilmazel, B., Hu, Y., Sigoillot, F., Smith, J.A., Shamu, C.E., Perrimon, N., Mohr, S.E. Online GESS: prediction of miRNA-like off-target effects in large-scale RNAi screen data by seed region analysis. *BMC Bioinformatics* **15**, 192 (2014).
6. Huang da, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).
7. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
8. van Dam, T.J. *et al.* The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. *Cilia* **2**, 7 (2013).
9. Arts, H.H. *et al.* Mutations in the gene encoding the basal body protein RPGRIP1L, a nephrocystin-4 interactor, cause Joubert syndrome. *Nat Genet* **39**, 882-8 (2007).
10. Huang, L. *et al.* TMEM237 is mutated in individuals with a Joubert Syndrome related disorder and expands the role of the TMEM family at the ciliary transition zone. *Am J Hum Genet* **89**, 713-30 (2011).
11. Shaheen, R. *et al.* Mutations in CSPP1, encoding a core centrosomal protein, cause a range of ciliopathy phenotypes in humans. *Am J Hum Genet* **94**, 73-9 (2014).
12. Tuz, K., *et al.* Mutations in CSPP1 cause primary cilia abnormalities and Joubert syndrome with or without Jeune asphyxiating thoracic dystrophy. *Am J Hum Genet* **94**, 62-72 (2014).

13. Abu-Safieh, L. *et al.* Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. *Genome Res* **23**, 236-47 (2013).
14. Schmidts, M. *et al.* Exome sequencing identifies DYNC2H1 mutations as a common cause of asphyxiating thoracic dystrophy (Jeune syndrome) without major polydactyly, renal or retinal involvement. *J Med Genet* **50**, 309-23 (2013).
15. Plagnol, V. *et al.* A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* **28**, 2747-54 (2012).
16. Schmidts, M. *et al.* Combined NGS approaches identify mutations in the intraflagellar transport gene IFT140 in skeletal ciliopathies with early progressive kidney Disease. *Hum Mutat* **34**, 714-24 (2013).
17. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**,1073-81 (2009).
18. Raven, J. The Raven's progressive matrices: change and stability over culture and time. *Cogn Psychol* **41**, 1-48 (2000).
19. Uricchio, L.H., Chong, J.X., Ross, K.D., Ober, C. & Nicolae, D.L. Accurate imputation of rare and common variants in a founder population from a small number of sequenced individuals. *Genet Epidemiol* **36**, 312-9 (2012).
20. Tory, K., *et al.* Mutation-dependent recessive inheritance of NPHS2-associated steroid-resistant nephrotic syndrome. *Nat Genet* **46**, 299-304 (2014).
21. Baujat, G. *et al.* Asphyxiating thoracic dysplasia: clinical and molecular review of 39 families. *J Med Genet* **50**, 91-8 (2013).
22. de Vries, J. *et al.* Jeune syndrome: description of 13 cases and a proposal for follow-up protocol. *Eur J Pediatr* **169**, 77-88 (2010).