Supplementary Methods.

Details of sequencing analysis approaches used for genome sequencing and exome sequencing

Proband 1: Library construction and genome sequencing of the proband's DNA was conducted at The Centre for Applied Genomics (TCAG) at the Hospital for Sick Children according to methods previously published[1]. CNVs were called according to methods previously published[2]. Sequence reads were aligned to the GRCh37 human reference genome assembly. Genome sequencing was filtered according to a standard lab filtering pipeline (**Figure S1**). Briefly, single nucleotide variants (SNVs), indels, structural variants (SVs), and transposable elements (TEs) were kept 1) if rare (<1% MAF in GnomAD[3], Bravo[4]), 2) if predicted to be damaging by various predictive pathogenicity scores (Sorting Intolerant from Tolerant (SIFT)[5], Polyphen-2[6], Mutation Taster[7], Mutation Assessor[8], Combined Annotation – Dependent Depletion (CADD PHRED)[9], PhyloPMam[10], PhyloPVert[10], **Pro**tein **V**ariation **E**ffect **An**alyzer (PROVEAN)[11]), 3) if predicted to cause splicing defects by three or more of the tools Splice Site Finder (SSF)[12], MaxEntScan (MaxEnt)[13], Splice Site Prediction by Neural Network (NNSPLICE)[14], and Human Splicing Finder (HSF)[15] accessed through Alamut Visual 2.11 (Interactive Biosoftware, Rouen, France), and 4) if segregating with the disease phenotype in the family. Variants in inherited retinal disease genes (RetNet; https://sph.uth.edu/retnet/) and genes with retinal expression[16] were prioritized. The R script used for filtering proband 1 genome sequencing is available upon request. Candidate disease-causing variants underwent segregation analysis using PCR and Sanger sequencing of DNA extracted from blood (primer sequences are listed in **Table S1**). Following these steps candidate biallelic disease-causing variants were identified in the genes *DYNC2H1* and *KIF7* (**Table S2**). *KIF7* variants were eliminated as candidate disease-causing variants for several

reasons. First, there is lack of evidence of a deleterious effect on protein structure as predicted *in silico.* Both *KIF7* variants occur at the disordered tail end of the protein which makes them unlikely to induce a structural perturbation[17]. Secondly, the patient cellular phenotype does not match what is reported for *KIF7* mutants in the literature. Specifically, cilia in a patient-derived fibroblast cell lines did not show abnormal ciliary architecture which is characteristic of mutant *KIF7* cell lines[18,19]. Additionally, patient-derived fibroblast cell lines show defects in IFT which is not characteristic of *KIF7* variants [18] . Finally, the *KIF7* clinical disease phenotype does not fit the patient clinical phenotype well since there is an absence of RD reported in *KIF7* patients[18,19].

Proband 2: Proband 2 underwent exome sequencing as part of the UK Inherited Retinal Disease Consortium, an ongoing study of inherited retinal disease in families without a molecular diagnosis. In brief, 200 ng of genomic DNA was sheared, processed, and captured using the Human All Exon XT Library V6 enrichment kit (Agilent, Santa Clara, CA, USA) Pooled libraries were sequenced on an Illumina HiSeq3000, generating 150bp paired end reads. Following QC and adapter trimming, the resulting fastq files were aligned to the GRCh37 version of the human genome using BWA and processed for duplicates and in/del realignment using Samtools, Picard and the Genome Analysis Toolkit (GATK), according to best practice guidelines. Variant calling was also conducted using the Genome Analysis Tool Kit (GATK, version 3.3-0) and variants were annotated with Variant Effect Predictor, with filtering to remove common variants (>0.1% in dbSNP150) with a CADD (v1.3) score < 15 using VCFhacks (https://github.com/david-a-parry/vcfhacks). As the parents of proband 2 are related and both unaffected, a recessive mode of inheritance was assumed, with the likely pathogenic variant

predicted to be in a region identical-by-descent. The remaining list was therefore filtered to include homozygous variants only. Seven variants remained after filtering (**Table S3**). Missense variants in *SLC1A1* and *TJP2* were excluded based on the proband not having the phenotypes associated with recessive mutations in either gene. Four variants were identified in genes not currently associated with a disease phenotype (*PACSIN1*, *DDR1*, *EYA2* and *AGER*), but were predicted to be benign (Polyphen2) or tolerated (SIFT), therefore considered less likely to be the cause of disease in this case. This left the null variant in *DYNC2H1*, which scored the highest CADD PHRED score (51) in the list, and is in a gene previously associated with a retinal phenotype. Candidate disease-causing variants were segregated by PCR and Sanger sequencing.

Proband 3: Genome sequencing was filtered according to Genomics England Ltd (GEL) clinical filtering pipeline. Briefly, SNVs and indels were kept if rare (<1% MAF), protein altering in at least one transcript, the allelic state matched the known mode of inheritance for the gene and the disorder, and familial segregation was observed (where applicable). A virtual gene panel analysis was applied to tier variants according to likelihood of causing disease. Panels of genes confidently linked to specific groups of disorders were constructed in PanelApp by a group of expert curators. Panels were selected automatically by the custom Panel Assigner tool, according to the Human Phenotype Ontology (HPO) assigned to the proband at recruitment. Protein truncating variants or *de novo* variants in genes on selected panels were tiered 1, protein altering variants (missense, splice region variants) in genes on selected panels were tiered 2. All other variants were tiered 3. This proband had no tier 1 or 2 variants *i.e*. no variants in genes previously linked to rod-cone dystrophy. 440 tier 3 variants were identified. In an independent study attempting to identify novel ciliopathies and ciliopathy patients, in which a virtual gene

panel was applied to tier 3 variants, based on Syscilia Gold standard v1[20] and CiliaCarta[21] to identify predicted pathogenic in cilia genes, this patient was identified with a homozygous frameshift *DYNC2H1* as an excellent candidate for further investigation. Homozygous missense variant in *C2CD3* (Arg413Gln, SIFT deleterious, Polyphen probably damaging) was identified as a candidate disease variant but it carried a high allele frequency in gnomAD. Independently, following the negative clinical pipeline analysis, rare (MAF<0.01) homozygous protein altering variants in regions of autozygosity affecting genes outside of the posterior segment abnormalities gene panel were interrogated, this revealed just two candidate variants: chr11:103241543C>A, *DYNC2H1* c.9836C>A, p.Ser3279* and chr11:82733311G>A, *FAM181B* c.419C>T, p.Pro140Lys. In light of the damaging nature of the *DYNC2H1* variant, its position in a seemingly retinal specific exon, the importance of the protein in cilia function and prior association of the gene with ciliopathy, this variant was prioritised as the most likely candidate.

Proband 4: Genome Sequencing and variant filtering was performed as previously described[22]. Following the exclusion of variants in known IRD genes, regions of homozygosity were interrogated for rare (MAF <0.01) predicted to be loss of function or damaging coding variants. Two variants survived filtering: chr11:103112272C>G, *DYNC2H1* c.9836C>A, p.Ser3279Ter and chr11:82444354G>A, *FAM181B* c.419C>T, p.Pro140Leu. Of these, the *DYNC2H1* variant was deemed to be the higher priority candidate due to the damaging nature of the variant.

Proband 5: Exome sequencing for proband 5 was performed according to a previously described protocol[23]. After quality filtering, variants in known IRD genes were investigated.

Subsequently, rare variants (MAF <0.05) were sorted into three categories: nonsense and frameshift, canonical and non-canonical splice sites, in-frame deletions and insertions and missense alterations. For missense variants, MetaSVM prediction was set to damaging; tolerated could be included only if CADD score was above 20[9,24]. Silent mutations were included only with CADD >20. ClinPred was applied to estimate clinical relevance of the variants (threshold 0.8)[25]. Only one gene with biallelic variants survived the filtering: *DYNC2H1* with two missense changes: chr11:103199375 A>C, c.7987A>C, p.Thr2663Pro and chr11:103468635 T>G, c.12716T>G, p.Leu4239Arg.

Cell lines for proband 1

The proband 1-derived fibroblast cell line was established from a sample of proband 1 skin at the Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada according to their standard protocols and procedures. The cell line tested negative for mycoplasma contamination and was authenticated by Sanger sequencing for V1 and V2. Proband 1-derived fibroblast cells were cultured in AMEM medium (Wisent, Saint-Jean-Baptiste, Quebec) supplemented with 10% fetal bovine serum (FBS) (Gibco, Gaithersburg, MD). When cells reached 90-100% confluence, they were washed once with phosphate buffered saline (PBS) and starved for 72 hours in 1XOptiMEM (Gibco, Gaithersburg, MD). Before fixation, cells were placed in ice for 30 mins to destabilize cytoplasmic microtubules. Starved cells were washed once with room temperature (RT) PBS and fixed with ice cold methanol for 20 mins at -30ºC. Fixed cells were washed 3x with RT PBS and blocked with blocking buffer (5% FBS in 0.05% PBST) for 1 hour at RT. Following blocking, coverslips were incubated with primary antibodies (diluted in blocking buffer) overnight in a humid chamber at 4ºC. The next day, coverslips were washed 3 times with 0.05% PBST. Coverslips were then incubated with

secondary antibodies (diluted in blocking buffer) for 1 hour at RT protected from light followed by Hoescht for 10 minutes at RT also protected from light. Coverslips were then washed 3 times with 0.05% PBST and left to dry for 45 minutes at RT, protected from light. Coverslips were mounted onto 12mm glass slides using mounting media (1M Tris pH 8, 0.5% propyl gallate, 90% glycerol) and sealed. Immunofluorescence experiments were visualized using a laser scanning confocal microscope (Zeiss, Oberkochen, Germany). 100 0.2uM z-stacks were acquired at 63x using the Zen software for each image in order to ensure capture of cilia which protrude from the cell membrane. Once coverslips were brought into focus, and the slide was moved horizontally and vertically at random to a new position where the image was acquired. For the cilia formation and function assays, approximately 100 cells were counted manually in each patient and control ciliated fibroblast cell lines using the FIJI software (NIH, USA)[26]. A sample size of 100 cilia were counted to ensure adequate power to detect changes in IFT88 fluorescence intensity. This is similar to the number of cilia that were counted in a previous publication also quantifying IFT88 fluorescence intensity in cilia[27]. The individual doing the quantification was blinded to which sample was from proband 1 vs WT. Fluorescence intensities of cilia were quantified manually using the sum intensity images in FIJI. Rectangle areas were drawn around each cilium and the fluorescence intensities and rectangle areas were measured. Background fluorescence intensity measurements (the average of mean intensities from 10 rectangles drawn on the background of the image) were multiplied by the area of each cilia rectangle and subtracted from the fluorescence intensity of each cilia rectangle.

Antibodies for Immunofluorescence

The following antibodies were used: 1:500 rabbit anti-IFT88 (Proteintech 13967-1-AP), 1:2,000 mouse anti-gamma tubulin (Sigma-Aldrich T6557), and 1:500 mouse anti-acetylated tubulin (Sigma-Aldrich T7451). Secondary antibodies used were 1:500 goat-anti-mouse-Alexa555 (Invitrogen A28180) and 1:500 goat-anti-rabbit-Alexa488 (Invitrogen A27034). Nuclei were stained with 10 mg/mL Hoescht (Life technologies, Carlsbad, CA, USA).

RNA extraction, cDNA synthesis, and DYNC2H1 transcript validation

RNA was extracted from a proband 1-derived fibroblast cell line and a control-derived fibroblast cell lines using the RNeasy Plus mini-kit (Qiagen, Hilden, Germany). Approximately 6 μg of resulting RNA was reverse transcribed into cDNA using the SuperScript IV First Strand kit (Thermo Fisher Scientific, Waltham, MA, USA). For the derived cDNA, at least one primer was designed across *DYNC2H1* exon-exon boundaries to minimize the chances of amplifying genomic DNA (gDNA). To favor detection of the low-abundance normally spliced [(NS) transcript 2b], an Amplification Refractory Mutation System (ARMS) approach[28] was used, followed by Sanger sequencing.

For validation of *DYNC2H1* microexon expression in probands 3, 4 and 5, RNA was extracted from cell lines and retinal organoids using the RNeasy Micro Kit (Qiagen, Hilden, Germany) following the manufacturer instructions. 50 ng of RNA were converted to cDNA by the Tetro cDNA Synthesis Kit (Bioline, London, UK) and a mixture of oligo-dT and random hexamer primers. 2.5 ng of cDNA were then used to detect the *DYNC2H1* gene isoforms. For control BJ-derived retinal organoid cDNA, PCR amplicons for the *DYNC2H1* retinal isoform only differed

by 21 bp. This caused formation of nonspecific cDNA heteroduplex. For this reason, samples were subjected to digestion by 3 units/sample of T7 Endonuclease I (New England BioLabs, Ipswich, MA, USA) 37°C, 60 min prior to loading a 2.5% agarose gel for visualisation. Bands were excised and extracted from the gel using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) and sent for Sanger sequencing.

Protein expression and purification

Frozen cell pellets from 225–250 ml Sf9 cultures were resuspended in 20 ml purification buffer (30 mM HEPES [pH 7.4], 300 mM KCl, 50 mM K-acetate, 2 mM Mg-acetate, 1 mM EGTA, 10% [v/v] glycerol, 1 mM DTT, 0.2 mM Mg-ATP, 1 mM PMSF) supplemented with a cOmplete™ EDTA-free Protease Inhibitor Cocktail (Roche, Basel, Switzerland). Cells were lysed using a Dounce homogenizer and 10 strokes with a small clearance pestle. Lysates were clarified by ultracentrifugation in a Type 70 Ti rotor at 183,960 *g* for 30 min. The supernatant was incubated for 1 hour on a roller with 0.5 ml IgG Sepharose 6 resin (GE Healthcare, Chicago, IL, USA) pre-washed in purification buffer. Resin and bound-proteins were collected by gentle centrifugation at 670 *g* for 5 min, transferred into a 10 ml column, and washed with 2x 10 ml volumes of purification buffer and 1x 10 ml volume of TEV buffer (50 mM Tris [pH 7.5], 150 mM K-acetate, 2 mM Mg-acetate, 1 mM EGTA, 10% [v/v] glycerol, 1 mM DTT, 0.2 mM Mg-ATP).

Microtubule gliding assay

Wild type, p.Glu2211Val (V2) and p.Asp4203Trpfs*7 (V1) dynein-2 motor protein samples were biotinylated for microtubule gliding assays via their N-terminal $SNAP_f$ tag as described[29] (**Figure 3**). Flow chambers were made between glass slides, PEG-biotin coverslips, and double-sided tape. Fluorescently-labelled microtubules (prepared as described[29]) were visualized on an Eclipse Ti-E inverted microscope with a CFI Apo TIRF 1.49 N.A. oil objective, Perfect Focus System, H-TIRF module, LU-N4 laser unit (Nikon, Minato, Japan) and a quad band filter set (Chroma, Guishan, Taiwan). Images were recorded with 100 ms exposures on an iXon DU888 Ultra EMCCD camera (Andor, Belfast, UK), controlled with NIS-Elements AR Software (Nikon). Different areas of the coverslip were randomly selected for imaging. Temperature of the flow chamber was maintained at 25ºC by an objective heater (Okolab, Ottaviano, Italy). Chambers were sequentially incubated with 1) blocking solution (0.75% Pluronic F-127, 5 mg/ml casein) for >10 min, followed by two washes with B80-TK (80 mM PIPES [pH 6.9], 2 mM $MgCl_2$, 1 mM EGTA, 1 mM DTT, 20 μM taxol, 50 mM KCl); 2) 0.5 mg/ml neutravidin for 2 min, followed by two washes with B80-TK; 3) biotinylated motor protein (1.6 nM) for 2 min, followed by two washes with B80-TK supplemented with 1 mg/ml casein; 4) 0.1 μM Alexa-488 microtubules in assay solution (B80-TK supplemented with 1 mg/ml casein, 1 mM Mg-ATP, 71 mM β-mercaptoethanol, 20 mM glucose, 300 μg/ml glucose oxidase, 60 μg/ml catalase). Microtubule gliding velocities were calculated from kymographs generated in FIJI[26]. Graphing was performed in Prism5 (GraphPad).

References for Supplementary Methods

1. Tavares E, Tang CY, Vig A, et al. Retrotransposon insertion as a novel mutational event in Bardet-Biedl syndrome. *Mol Genet Genomic Med.* 2019;7(2):e00521.
2. C Yuen RK, Merico D, Bookman M, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nature Neuroscience.* 2017;20(4):602-611.
3. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-443.
4. The NHLBI Trans-Omics for Precision Medicine (TOPMed) Whole Genome Sequencing Program. University of Michigan and NHLBI. https://bravo.sph.umich.edu/freeze5/hg38/. Published 2018. Accessed.
5. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols.* 2009;4:1073.
6. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One.* 2012;7(10):e46688-e46688.
7. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods.* 2010;7:575.
8. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research.* 2011;39(17):e118-e118.
9. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 2019;47(D1):D886-D894.
10. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 2010;20(1):110-121.
11. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics.* 2015;31(16):2745-2747.
12. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Research.* 1987;15(17):7155-7174.
13. Yeo G, Burge CB. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *Journal of Computational Biology.* 2004;11(2-3):377-394.
14. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved Splice Site Detection in Genie. *Journal of Computational Biology.* 1997;4(3):311-323.
15. Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research.* 2009;37(9):e67-e67.
16. Farkas MH, Grant GR, White JA, Sousa ME, Consugar MB, Pierce EA. Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC Genomics.* 2013;14(1):486.
17. Klejnot M, Kozielski F. Structural insights into human Kif7, a kinesin involved in Hedgehog signalling. *Acta Crystallographica Section D Biological Crystallography.* 2012;68(2):154-159.
18. He M, Subramanian R, Bangs F, et al. The kinesin-4 protein Kif7 regulates mammalian Hedgehog signalling by organizing the cilium tip compartment. *Nature Cell Biology.* 2014;16:663.
19. Dafinger C, Liebau MC, Elsayed SM, et al. Mutations in KIF7 link Joubert syndrome with Sonic Hedgehog signaling and microtubule dynamics. *J Clin Invest.* 2011;121(7):2662-2667.

20. van Dam T, Wheway G, Slaats G. The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. Cilia 2: 7. doi: 10.1186. In: PMID; 2013.
21. van Dam TJP, Kennedy J, van der Lee R, et al. CiliaCarta: An integrated and validated compendium of ciliary genes. *PLoS One.* 2019;14(5):e0216705.
22. Taylor RL, Arno G, Poulter JA, et al. Association of Steroid 5α-Reductase Type 3 Congenital Disorder of Glycosylation With Early-Onset Retinal Dystrophy. *JAMA Ophthalmol.* 2017;135(4):339-347.
23. Lipinski P, Stawinski P, Rydzanicz M, et al. Mild Zellweger syndrome due to functionally confirmed novel PEX1 variants. *J Appl Genet.* 2019.
24. Kim S, Jhong JH, Lee J, Koo JY. Meta-analytic support vector machine for integrating multiple omics data. *BioData Min.* 2017;10:2.
25. Alirezaie N, Kernohan KD, Hartley T, Majewski J, Hocking TD. ClinPred: Prediction Tool to Identify Disease-Relevant Nonsynonymous Single-Nucleotide Variants. *American journal of human genetics.* 2018;103(4):474-483.
26. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 2012;9(7):676-682.
27. Ocbina PJR, Eggenschwiler JT, Moskowitz I, Anderson KV. Complex interactions between genes controlling trafficking in primary cilia. *Nature Genetics.* 2011;43(6):547-553.
28. Little S. Amplification-Refractory Mutation System (ARMS) Analysis of Point Mutations. *Current Protocols in Human Genetics.* 1995;7(1):9.8.1-9.8.12.
29. Toropova K, Mladenov M, Roberts AJ. Intraflagellar transport dynein is autoinhibited by trapping of its mechanical and track-binding elements. *Nature Structural &Amp; Molecular Biology.* 2017;24:461.

Supplementary Results.

Details of Clinical Evaluations (Summary in Table 1)

Proband 1 (p.Asp4203Trpfs*7/p.Glu2211Val): This female proband, last seen at 19 years old was part of a sibship of three children, born to a non-consanguineous marriage of mixed Barbados and Greek origin (**Figure 1**). Other than childhood complaints from decreased vision and nyctalopia, the proband was healthy and the only affected member of the family. Her central vision loss in each eye was moderate (0.5, 0.6 LogMAR at 19 years) but her field of vision was very constricted (<10 degrees at 19 years) and her ERG non-recordable. Systemically, the patient was thoroughly examined and had a normal chest X-ray (**Figure 1**), skeletal system, kidney, liver, cognition, hearing, balance, height, and weight. She displayed signs and symptoms of autosomal recessive non syndromic retinitis pigmentosa (ARRP) with normal cognitive skills. Panel–based clinical genetic testing did not identify any potentially pathogenic variants.

Proband 2 (p.Ser3279*/ p.Ser3279*): Proband 2, a female last seen at 36 years old, first noticed symptoms of nyctalopia at age 25 years which progressed only very slowly. At age 31 years her visual acuity was 0.4 LogMAR in both eyes. She was the only affected child of consanguineous parents. Proband 2 had normal hand and chest X-rays and no signs of renal failure. Hence she was diagnosed only with non syndromic ARRP. Though on OCT she showed preserved central islands of photoreceptors, she was registered severely sight impaired at age 31years due to extensive peripheral field loss (data not available).

Proband 3 (p.Ser3279*/ p.Ser3279*): Proband 3 was a 47 year old male born to a consanguineous union and was the youngest of four siblings. He was the only family member known to be affected with non-syndromic ARRP. Symptoms of visual field loss were only noted at age 43. He was recruited at the Moorfields Eye Hospital Inherited Eye Disease clinic for whole genome sequencing as part of the Genomics England 100,000 genomes project.

Proband 4 (p.Ser3279*/ p.Ser3279*): Proband 4 was a 32 year old male diagnosed with non-syndromic ARRP. First symptoms of peripheral vision loss were at age 18 years.

Proband 5 (p.Thr2663Pro/ p.Leu4239Arg): This young female with non-syndromic RP was diagnosed in her early twenties.

Table S1. Primers used.

Primer	Sequence 5'→ 3'
Primers used for Proband 1 *DYNC2H1* transcript amplification	
DYNC2H1_39-40_F	TCTTTCTTAGTGATGAAGAGAC
DYNC2H1_42-43_R	CAGCATCCCTTTGCCACATC
DYNC2H1_g.103058126_R	AGAAGTGGTTTGTGCACTTCAG
DYNC2H1_86-87_F	AATACAGAACTGGGTAGATAAAGC
DYNC2H1_88-89_R	ACAAGCCACTGATCTTAATTTGTAG
Primers used for Agena massARRAY genotyping	
Primers used for initial DYNC2H1 amplification	
DYNC2H1_g.103057048_F	ACGTTGGATGCAGTCAAGTCTTCTGGCTTC
DYNC2H1_g.103057048_R	ACGTTGGATGCCTCCAGACTTTCACAAACC
DYNC2H1_g.103055779_F	ACGTTGGATGAAGTCTGGAGGAGATTCTCG
DYNC2H1_g.103055779_R	ACGTTGGATGGGGACTTGGTGGAAATCTG
DYNC2H1_g.103055625_F	ACGTTGGATGGGTTTGTGAAAGTCTGGAGG
DYNC2H1_g.103055625_R	ACGTTGGATGTGAAAAGGCTTTACAATGGG
Extension primers	
DYNC2H1_g.103057048_E1	ACTCTACTAGGGGTCG
DYNC2H1_g.103055779_E2	ACGTTTGGAATTTACCAAAG
DYNC2H1_g.103055625_E3	CAATGGGTTCTAAAGCAG
Primers used for BJ fibroblast cell line and retinal organoid *DYNC2H1* transcript amplification	
DYNC2H1_63-65_F	CTGCTGCTCCTGAATCTCTG
DYNC2H1_63-65_R	TGTAGCTTGGGAAGAAGGATC
PAX6_F	AACGATAACATACCAAGCGTGTC
PAX6_R	GTCTGCCCGTTCAACATCCT
VSX2_F	GTGGCTACTGGGGATGCAC
VSX2_R	TCCTGCTCCATCTTGTCGAG
CRX_F	TTTGCCAAGACCCAGTACC
CRX_R	GTTCTTGAACCAAACCTGAACC
NRL_F	CACTGACCACATCCTCTCGG

NRL_R	GAGGGTTCCCGCTTTACCTC
NR2E3_F	TCTTCAAGCCAGAGACGCG
NR2E3_R	CTCAAAGACGGGAGGAGCAG
GAPDH_F	GCCCAATACGACCAAATCC
GAPDH_R	GGTACTTTATTGATGGTACATGACAAG

Table S2. Variants remaining in proband 1 after GS filtering.

Position (hg19)	Gene (Transcript)	SNP ID	DNA Variant	Protein Variant	CADD phred (v1.3)	PolyPhen2 (HumVar)	SIFT	Disease Associations (MIM#)
11:103055779 11:103327020	*DYNC2H1* (NM_001080463.1)	rs929322688 NA	c.6632A>T c.12605_12606dup	p.Glu2211Val p.Asp4203Trpfs*7	18.06 NA	0.04 (Benign) NA	0.004 (Deleterious) NA	short-rib thoracic dysplasia-3 (613091)
15:90171684 15:90171738	*KIF7* (NM_198525)	rs145726393 rs150248985	c.3998C>A c.3944C>T	p.Pro1333Gln p.Pro1315Leu	14.53 N/A	0.014 (Benign) 0.074 (Benign)	0.094 0.473	Al-Gazali-Bakalinova syndrome (607131) Hydrolethalus syndrome 2 (614120) Acrocallosal syndrome/ Joubert syndrome 12 (200990)

Table S3. Homozygous variants remaining in proband 2 after ES filtering.

Position (hg19)	Gene (Transcript)	SNP ID	DNA Variant	Protein Variant	S.Asian GMAF	CADD PHRED (v1.3)	PolyPhen2 (HumVar)	SIFT	Disease Associations (MIM#)
11:103112272	DYNC2H1 (NM_001080463.1)	rs762578912	c.9836C>G	p.Ser3279*	0.00003	51	N/A	N/A	short-rib thoracic dysplasia-3 (613091)
9:4573977	SLC1A1 (NM_004170)	rs564231110	c.838C>T	p.Arg280Cys	0.00007	35	0.999 (Probably Damaging)	0.02 (Deleterious)	Dicarboxylic aminoaciduria (222730)
9:71854860	TJP2 (NM_001170416.1)	N/A	c.2456C>T	p.His819Leu	N/A	31	0.685 (Possibly Damaging)	0.0 (Deleterious)	Familial hypercholanemia (607748) Familial intrahepatic cholestasis-4 (615878)
6:34498084	PACSIN1 (NM_020804.4)	N/A	c.853A>G	p.Arg285Gly	N/A	23.1	0.614 (Possibly Damaging)	0.0 (Deleterious)	N/A
6:30863247	DDR1 (NM_001297654.1)	rs201876615	c.1580C>T	p.Pro527Leu	N/A	20.3	0.000 (Benign)	0.74 (Tolerated)	N/A
20:45702903	EYA2 (NM_005244.4)	rs535313513	c.590A>C	p.Leu197Pro	0.00026	17.81	0.861 (Possibly Damaging)	0.32 (Tolerated)	N/A
6:32150675	AGER (NM_001206934.1)	rs1469908411	c.634C>T	p.Pro212Ser	0.00006	17.39	0.683 (Possibly Damaging)	0.27 (Tolerated)	N/A

Variants are ranked on CADD score from highest to lowest (CADD > 15, MAF< 0.001)

Table S4. Predictive pathogenicity scores and population frequency for *DYNC2H1* variants 1-5. Variant genomic positions are reported according to GRCh37 *DYNC2H1*, chromosome 11.

	Variant	V1	V2	V3	V4	V5
Predictive Pathogenicity Scores		g.103327020_103327021dup NM_001377.2: c.12584_12585dup p.Asp4196Trpfs*7 NM_001080463.1: c.12605_12606dup p.Asp4203Trpfs*7	g.103055779 A>T NM_001377.2: c.6632A>T p.Glu2211Val NM_001080463.1: c.6632A>T p.Glu2211Val	g.103112272 C>G NM_001080463.1: c.9836C>G p.Ser3279*	g.103070104 A>C NM_001377.2: c.7987A>C p.Thr2663Pro NM_001080463.1: c.7987A>C p.Thr2663Pro	g.103339363 T>G NM_001377.2: c.12695T>G p.Leu4232Arg‡ NM_001080463.1: c.12716T>G p.Leu4239Arg‡
	SIFT[1]	NA	0.04	NA	0.001	0.01
	PolyPhen-2[2]	NA	0.004	NA	0.603	0.905
	MA[3]	NA	2.77	NA	3.085	3.215
	MT[4]	0.877	NA	0.81	0.81	0.81
	phyloMam[5]	2.44	1.95	0.852	0.964	1.199
	PhyloVert[6]	6.59	8.93	3.334	7.064	7.359
	CaddPhred[7]	NA	18.06	51	27.3	25.2
	PROVEAN[8]	NA	-4.52	NA	-4.7	-3.79
Population Frequency	**GnomAD**[9]	Not found	Not found	0.0004%	Not found	Not found
	Bravo[10]	Not found	Not found	Not found	Not found	Not found
ACMG Classification	**Classification**	Pathogenic	Likely Pathogenic	Pathogenic	Uncertain Significance	Uncertain Significance
	Pathogenicity Criteria[a]	PVS1, PS3, PM2, PM3, PP3	PS3, PM2, PM3, PP3	PVS1, PS4, PM2, PP3	PM2, PP3	PM2, PP3

Legend: [1]Sorting Intolerant From Tolerant (SIFT)[1] values ≤ 0.05 are predicted pathogenic; [2]Polymorphism Phenotyping v2 HumVar (PolyPhen-2)[2] values ≥ 0.09 are predicted pathogenic; [3]Mutation Assessor (MA)[3] values ≥ 1.9 are predicted pathogenic; [4]Mutation Taster (MT)[4] values ≥ 0.5 are predicted pathogenic; [5]Phylogenetic P value mammals (PhyloPMam_avg)[5] values > 2.3 are predicted pathogenic; [6]Phylogenetic P value vertebrates (PhyloVert_avg)[5] values > 4 are predicted pathogenic; [7]Combined Annotation – Dependent Depletion (CADD phred)[6] values ≥15 are predicted pathogenic; [8]**Pro**tein **V**ariation **E**ffect **An**alyzer (PROVEAN)[7] values ≤ -2.5 are predicted pathogenic; [9]Genome Aggregation Database (GnomAD)[8]; [10]Bravo[9]. ‡ Variant has been previously reported in Schmidts et al. (2013)[10].

[a]ACMG classification of variants:
PVS1: null variant (nonsense, frameshift, canonical +- 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease.
PS3: Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product.
PS4: The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls.
PM2: Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium.
PM3: For recessive disorders, detected in trans with a pathogenic variant.
PP3: Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.).
Variants are listed for both *DYNC2H1* transcripts except for V3 which is specific to NM_001080463.1.
ClinVar submission number: SUB7370183.

References

1. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols.* 2009;4:1073.
2. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248-249.
3. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research.* 2011;39(17):e118-e118.
4. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods.* 2010;7:575.
5. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 2010;20(1):110-121.
6. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One.* 2012;7(10):e46688-e46688.
7. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics.* 2015;31(16):2745-2747.
8. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-443.
9. The NHLBI Trans-Omics for Precision Medicine (TOPMed) Whole Genome Sequencing Program. University of Michigan and NHLBI. https://bravo.sph.umich.edu/freeze5/hg38/. Published 2018. Accessed.
10. Schmidts M, Arts HH, Bongers EMHF, et al. Exome sequencing identifies DYNC2H1 mutations as a common cause of asphyxiating thoracic dystrophy (Jeune syndrome) without major polydactyly, renal or retinal involvement. *Journal of Medical Genetics.* 2013;50(5):309.

Table S5. Splice donor site prediction scores for the V2 allele compared to the wild-type (WT).

Splice Prediction Tools used	WT ccaaagaggtaat	V2 ccaaagtggtaat
[1]SSF (≥70)	79.53	71.02
[2]MaxEnt (≥ 0)	8.73	4.46
[3]NNSPLICE (≥ 0.4)	0.90	0.63
[4]HSF (≥65)	83.99	79.24

Legend: [1]Splice Site Finder (SSF) values ≥ 70 predict that the underlined site acts as a splice donor, [2]MaxEntScan (MaxEnt) values ≥ 0 predict that the underlined site acts as a splice donor, [3]Splice Site Prediction by Neural Network (NNSPLICE) values ≥ 0.4 predict that the underlined site acts as a splice donor, and [4]Human Splicing Finder (HSF) values ≥ 65 predict that the underlined site acts as a splice donor. All splicing prediction tools were accessed through Alamut Visual 2.11 (Interactive Biosoftware).

These tools were used to assess potential splicing defects resulting from V2. When the invariant AG in the wild-type donor site is disrupted by **V2** (bold), scores from all prediction tools are lower than WT scores which predicts a loss of the WT donor site.

References

1. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Research.* 1987;15(17):7155-7174.
2. Yeo G, Burge CB. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *Journal of Computational Biology.* 2004;11(2-3):377-394.
3. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved Splice Site Detection in Genie. *Journal of Computational Biology.* 1997;4(3):311-323.
4. Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research.* 2009;37(9):e67-e67.

Table S6. Haplotype analysis around *DYNC2H1* of 3 UK individuals homozygous for V3.

Chr.	Position	rs number	Ref	Alt	LDS	MAN	LON	Alt. Allele Freq.	
11	78372536	rs2277277	G	A	G/A			0.54438	
11	78498022	rs11237621	C	T	C/T			0.22590	
11	85906362	rs3844144	T	C	C/C	C/C	C/C	0.77915	
11	89936093	rs3758757	T	C	C/C	C/C	C/C	0.48805	
11	89939448	rs1943381	G	A	A/A	A/A	A/A	0.68383	
11	100641200	rs7944598	A	T	T/T	T/T	T/T	0.26475	
11	100830579	rs6590829	G	A	A/A	A/A	A/A	0.50214	
11	102822733	rs3819089	C	T	T/T	T/T	T/T	0.17042	
11	**103112272**	**rs762578912**	**C**	**G**	**G/G**	**G/G**	**G/G**	**0.000001**	***DYNC2H1* (V3)**
11	107197640	rs3758911	T	C	C/C	C/C	C/C	0.29804	
11	110023628	rs665013	C	T	T/T	T/T	T/T	0.74512	
11	111156836	rs3087967	T	C	C/C	C/C	C/C	0.73303	
11	111635655	rs7106104	T	C	T/C			0.26370	
11	111724133	rs10502151	C	T	C/T			0.35000	

Legend: Variants within 50Mb of *DYNC2H1* were investigated in proband 2 (LDS) to identify if the variant was within a homozygous region. A large homozygous region, flanked by rs11237621 and rs7106104, was identified and investigated in probands 3 (MAN) and 4 (LON). All variants were found to be homozygous for the same alleles in all 3 cases, confirming a shared haplotype (green). The frequencies of the alternative alleles were obtained from dbSNP151.

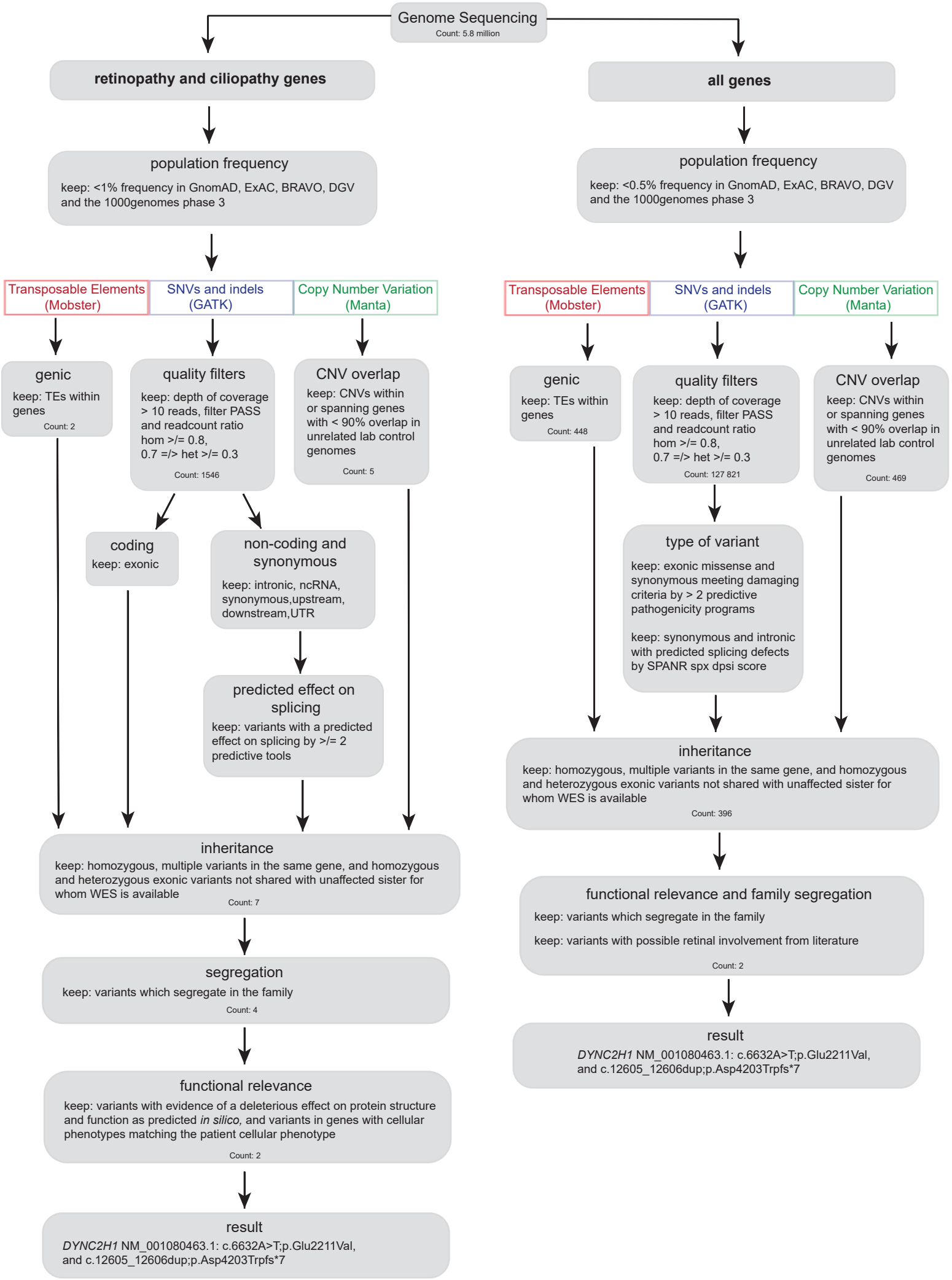
Genome Sequencing
Count: 5.8 million
retinopathy and ciliopathy genes
all genes
population frequency
keep: <1% frequency in GnomAD, ExAC, BRAVO, DGV and the 1000genomes phase 3
population frequency
keep: <0.5% frequency in GnomAD, ExAC, BRAVO, DGV and the 1000genomes phase 3
Transposable Elements (Mobster)
SNVs and indels (GATK)
Copy Number Variation (Manta)
Transposable Elements (Mobster)
SNVs and indels (GATK)
Copy Number Variation (Manta)
genic
keep: TEs within genes
Count: 2
quality filters
keep: depth of coverage > 10 reads, filter PASS and readcount ratio hom >/= 0.8, 0.7 =/> het >/= 0.3
Count: 1546
CNV overlap
keep: CNVs within or spanning genes with < 90% overlap in unrelated lab control genomes
Count: 5
genic
keep: TEs within genes
Count: 448
quality filters
keep: depth of coverage > 10 reads, filter PASS and readcount ratio hom >/= 0.8, 0.7 =/> het >/= 0.3
Count: 127 821
CNV overlap
keep: CNVs within or spanning genes with < 90% overlap in unrelated lab control genomes
Count: 469
coding
keep: exonic
non-coding and synonymous
keep: intronic, ncRNA, synonymous,upstream, downstream,UTR
type of variant
keep: exonic missense and synonymous meeting damaging criteria by > 2 predictive pathogenicity programs
keep: synonymous and intronic with predicted splicing defects by SPANR spx dpsi score
predicted effect on splicing
keep: variants with a predicted effect on splicing by >/= 2 predictive tools
inheritance
keep: homozygous, multiple variants in the same gene, and homozygous and heterozygous exonic variants not shared with unaffected sister for whom WES is available
Count: 7
inheritance
keep: homozygous, multiple variants in the same gene, and homozygous and heterozygous exonic variants not shared with unaffected sister for whom WES is available
Count: 396
segregation
keep: variants which segregate in the family
Count: 4
functional relevance and family segregation
keep: variants which segregate in the family
keep: variants with possible retinal involvement from literature
Count: 2
functional relevance
keep: variants with evidence of a deleterious effect on protein structure and function as predicted in silico, and variants in genes with cellular phenotypes matching the patient cellular phenotype
Count: 2
result
DYNC2H1 NM_001080463.1: c.6632A>T;p.Glu2211Val, and c.12605_12606dup;p.Asp4203Trpfs*7
result
DYNC2H1 NM_001080463.1: c.6632A>T;p.Glu2211Val, and c.12605_12606dup;p.Asp4203Trpfs*7

Figure S1. Genome Sequencing variant filtering strategy for proband 1.

Schematic of the approach used to filter the genome sequence of proband 1. The other approaches are similar.

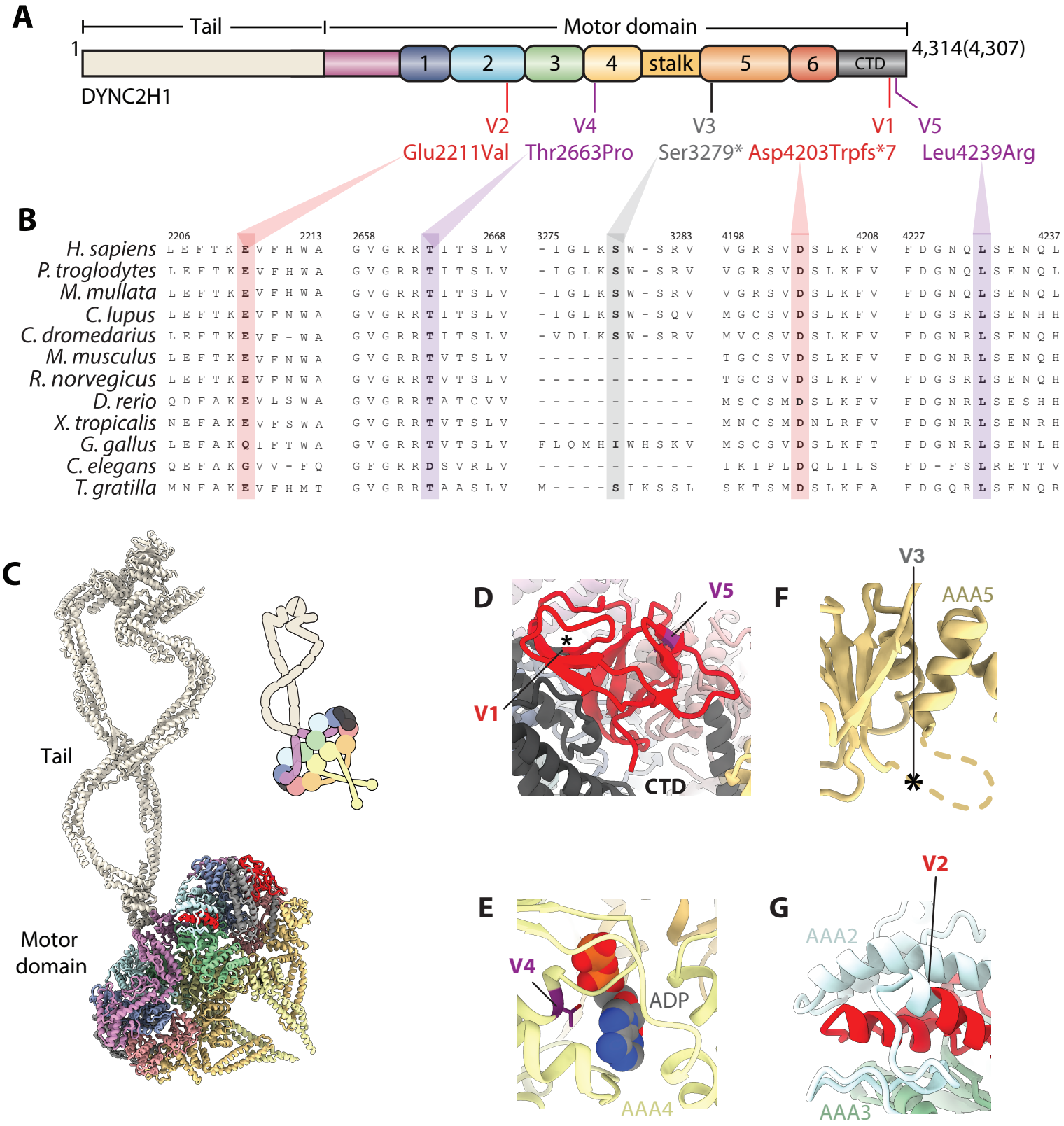
A
Tail
Motor domain
1
4,314(4,307)
1
2
3
4
stalk
5
6
CTD
DYNC2H1
V2
Glu2211Val
V4
Thr2663Pro
V3
Ser3279*
V1
Asp4203Trpfs*7
V5
Leu4239Arg
B
H. sapiens
P. troglodytes
M. mullata
C. lupus
C. dromedarius
M. musculus
R. norvegicus
D. rerio
X. tropicalis
G. gallus
C. elegans
T. gratilla
C
Tail
Motor domain
D
V5
V1
CTD
F
V3
AAA5
E
V4
ADP
AAA4
G
V2
AAA2
AAA3

Figure S2. Schematic mapping of proband 1–5 variants on the dynein-2 heavy chain. Variants are coloured according to proband; proband 1 (V1/V2): red, probands 2-4 (V3): grey, and proband 5 (V4/V5): purple. A) Domain organisation of the dynein-2 heavy chain (*DYNC2H1*). AAA+ modules 1–6 are labelled. CTD; C-terminal domain. The canonical isoform of *DYNC2H1* is 4,307 amino acids long, and a non-canonical isoform including an extra 7 amino acids in AAA5 is 4,314 amino acids long. Variants are numbered relative to the 4,314 amino acid isoform (NM_ 001080463.1). B) Amino acid alignment of dynein-2 orthologues across animals shows that V1-V5 fall on well-conserved amino acid residues. The amino acids are color shaded according to the probands in A. C) Structure of the dynein-2 heavy chain homodimer [36] shown in cartoon representation and colored by domain as in (A) with schematic inset. The tail (beige) forms an elongated structure that binds associated subunits but is not needed for activity of the motor domains. D)-G) Close up views of the proband variants. Red coloring denotes regions that would be missing within the CTD in V1 (D) and within AAA2 in V2 (G). Purple coloring denotes substitutions within the CTD for V5 (D) and within AAA4 for V4 (E). The latter contributes to the ADP nucleotide binding pocket in AAA4 (E). Dashed gold line approximates the 7 amino acid insert in the non-canonical isoform, with an Asterix marking the approximate V3 position (F).