TMEM107 recruits ciliopathy proteins to subdomains of the ciliary transition zone and causes Joubert syndrome

Nils J. Lambacher[1]*, Ange-Line Bruel[2]*, Teunis J. P. van Dam[3]*, Katarzyna Szymańska[4], Gisela G. Slaats[5], Stefanie Kuhns[1], Gavin J. McManus[6], Julie E. Kennedy[1], Karl Gaff[1], Ka Man Wu[5], Robin van der Lee[3], Lydie Burglen[7], Diane Doummar[7], Jean-Baptiste Rivière[2,8], Laurence Faivre[2,8], Tania Attié-Bitach[9,10,11], Sophie Saunier[9,10], Alistair Curd[12], Michelle Peckham[12], Rachel H. Giles[5], Colin A. Johnson[4], Martijn A. Huynen[3#], Christel Thauvin-Robinet[2,8#], Oliver E. Blacque[1#]

[1]School of Biomolecular and Biomedical Science, UCD Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland.

[2]EA4271 GAD, Genetics of Development Abnormalities, Burgundy University, Dijon, France

[3]Centre for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud university medical center, Geert Grooteplein 26-28, 6525 GA Nijmegen, Netherlands

[4]Section of Ophthalmology and Neurosciences, Leeds Institute of Biomolecular & Clinical Sciences, University of Leeds, Leeds, LS9 7TF, UK

[5]Department of Nephrology and Hypertension, University Medical Center Utrecht, Utrecht, The Netherlands

[6]School of Biochemistry and Immunology, Microscopy facility, Trinity Biomedical Sciences Institute, Trinity College Dublin, 152-160 Pearse Street, Dublin 2, Ireland

[7]Centre de référence des malformations et maladies congénitales du cervelet et Service de Génétique, APHP, Hôpital Trousseau, Paris, France; INSERM U1141, Paris, France

[8]FHU TRANSLAD, CHU Dijon, France

[9]INSERM UMR1163, Hôpital Necker-Enfants Malades, Paris, France

[10]Université Paris Descartes, Sorbonne Paris Cité, France, Institut IMAGINE, Paris, France

[11]Département de Génétique, Hôpital Necker-Enfants Malades, AP-HP, Paris, France

[12]School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, UK

*equal contribution

[#]Co-corresponding authors

Correspondence to: oliver.blacque@ucd.ie, martijn.huijnen@radboudumc.nl and christel.thauvin@chu-dijon.fr

INTRODUCTORY PARAGRAPH

The transition zone (TZ) ciliary subcompartment is thought to control cilium composition and signaling by facilitating a protein diffusion barrier at the ciliary base, and TZ defects cause ciliopathies such as Meckel-Gruber syndrome (MKS), nephronophthisis (NPHP) and Joubert syndrome (JBTS) [1]. However, the molecular composition and mechanisms underpinning TZ organisation and barrier regulation are poorly understood. To uncover candidate TZ genes, we employed bioinformatics (co-expression and co-evolution) and identified TMEM107 as a TZ protein mutated in oral-facial-digital syndrome (OFD) and JBTS patients. Mechanistic studies in *Caenorhabditis elegans* showed TMEM107 controls ciliary composition and functions redundantly with NPHP4 to regulate cilium integrity, TZ docking and assembly of membrane to microtubule Y-link connectors. Furthermore, nematode TMEM107 occupies an intermediate layer of the TZ-localised MKS module by organising recruitment of ciliopathy proteins MKS1, TMEM231 (JBTS20) and TMEM237 (JBTS14). Finally, MKS module membrane proteins are immobile and super-resolution microscopy (STED, dSTORM) in worms and mammalian cells reveals periodic localisations within the TZ. This work expands the MKS module of ciliopathy-causing TZ proteins associated with diffusion barrier formation and provides insight into TZ subdomain architecture.

MAIN TEXT

Cilia are microtubule-based structures serving motility, sensory, and signaling functions, and ciliary defects cause broad spectrum symptoms including cystic kidneys, blindness and nervous system defects[2]. Cilia possess subcompartments with distinct molecular compositions, enriched for receptors, channels, and signaling components[3]. Protein targeting to cilia depends on intraflagellar transport (IFT) and membrane trafficking pathways that sort, deliver and recycle ciliary components[4]. A recently established paradigm implicates 'gated' diffusion barriers at the ciliary base that regulate protein entry in a size-dependent manner and restrict lateral transport between ciliary and non-ciliary membranes[1, 5-10].

The ciliary gate is associated with the proximal-most 0.2-0.8 μm of the axoneme, termed TZ, characterised by Y-shaped structures that connect the microtubules and membrane. Y-links are also thought to organise the ciliary necklace, a membrane specialisation consisting of intramembranous particles that may contribute to diffusion barrier properties[11, 12]. Multiple MKS, JBTS and NPHP proteins are TZ-localised and regulate ciliary composition and signaling [5-7, 13]. In *C. elegans* (Ce) sensory neurons, 13 ciliopathy protein orthologues localise at the TZ, where they functionally associate as MKS (MKS-1, MKS-2, MKS-3, MKS-5, MKS-6, MKSR-1, MKSR-2, JBTS-14, TMEM-17, TMEM-231), NPHP (NPHP-1, NPHP-4) or CEP-290-associated modules to regulate Y-link formation and TZ docking during cilium assembly[10, 13-19].

To predict new TZ genes, we employed a weighted gene co-expression screening strategy [20], and a training set of 20 known TZ genes (**Supplementary Table 1**)[1], to interrogate >1600 mouse and

92 human expression microarray datasets in GEO (**Supplementary Table 2;** top 500 co-expressed
93 genes shown in **Supplementary Table 3).** In the integrated datasets, TZ genes display higher co-
94 expression with each other than genomic background (Mann-Whitney *U* tests: $p=1.57e^{-14}$
95 (human); $p=1.42e^{-9}$ (mouse); **Fig. 1a** and **Supplementary Fig. 1a**). Ciliary genes[21] are also
96 enriched (Mann-Whitney *U* tests: $p=7.35e^{-55}$ (human); $p=1.43e^{-44}$ (mouse)), but less than TZ
97 genes (Mann-Whitney *U* tests: $p=1.36e^{-10}$ (human); $p=1.01e^{-6}$ (mouse); (**Fig. 1a** and
98 **Supplementary Fig. 1a**). Cross validation (leave-one-out analysis) shows that our method
99 retrieves TZ genes versus ciliary genes, although this is not pronounced among top scoring genes
100 (**Fig. 1b**).
101
102 We compared the top 100 co-expressed human and mouse genes and identified 18 common
103 genes, of which 13 encode proteins with known cilia associations (**Supplementary Table 3**). For
104 the remaining 5 genes, we examined the genomes of ciliated species lacking a TZ
105 (*Physcomitrella patens*, *Selaginella moellendorffii, Plasmodium falciparum* and *Giardia*
106 *intestinalis*[22]) and found a TZ phylogenetic distribution for *TMEM107* (**Supplementary Fig.**
107 **1b**). Although this gene was previously shown to regulate mammalian ciliogenesis and Shh
108 signaling[23], the basis of these functions is unknown.
109
110 In *C. elegans*, the TMEM107 orthologue (F39B2.9; TMEM-107) is exclusively expressed in
111 ciliated cells and localises at the TZ (**Fig. 1c** and **Supplementary Fig. 2a**). Like other TZ genes,
112 worm and human *TMEM107* possess X-box promoter motifs and nematode *tmem-107* expression
113 is exclusively dependent on the RFX transcription factor DAF-19 (**Supplementary Fig. 2b, c**).
114 TMEM107 possesses four predicted transmembrane helices and short cytosolic N- and C-termini

115 (**Supplementary Fig. 1c, d**). Disruption of the nematode transmembrane domain linkers or
116 cytosolic N- or C- termini did not affect localisation, suggesting that TMEM-107's TZ
117 association depends on its transmembrane helices (**Supplementary Fig. 2d**). Similar to *C.*
118 *elegans* TMEM-107, human TMEM107::GFP also localises at the TZ (**Fig. 1d**).
119
120 The TZ localisation, combined with reported ciliogenesis and signaling roles[23], makes
121 *TMEM107* a strong ciliopathy gene candidate. Indeed, similar to observations for other
122 ciliopathy gene disruptions[24] and the *Tmem107 Schlei* mouse[23], *Tmem107*-depleted IMCD3 cells
123 display reduced ciliation and lumen size in 3D spheroids (**Fig. 2a, b**). To explore disease
124 associations, *TMEM107* exons were screened for mutations in 238 JBTS or OFD type VI
125 individuals using Sanger and next generation sequencing. We identified a homozygous missense
126 variant (NM_183065: g.8079298T>C; p.Glu45Gly) in OFDVI female twins with
127 consanguineous parents, and a compound heterozygous mutation comprising a frameshift
128 deletion (NM_032354.3: g.8077560delT; p.Leu134Phe*fs**8) and an in-frame codon deletion
129 (NM_032354.3: g.8077890_8077893delGAA; p.Phe106del) in a JBTS male (**Fig. 2c** and
130 **Supplementary Fig. 3a,b**). All patients possess similar neurological disturbances, retinopathy,
131 and the JBTS-associated molar tooth sign (**Fig. 2d-g** and **Supplementary Fig. 3c**). The females
132 were diagnosed with OFDVI because they also displayed hamartoma and frenulae phenotypes
133 (**Supplementary Fig. 3c**). All three variants segregated in the families, were predicted to be
134 pathogenic by Human Splicing Finder and PolyPhen2, and were absent in the heterozygous and
135 homozygous state in human variation databases (NHLBI Exome Variant Server, ExAC). In
136 further support of pathogenicity, fibroblasts from the JBTS male displayed reduced ciliation, and
137 those cilia that formed were abnormally long (**Fig. 2h-j**). The latter contrasts with the short cilia

of hypomorphic *Schlei* mouse fibroblasts, carrying a missense mutation (E125G) in Tmem107[23]; which may be explained by cell type, species or allelic distinctions. We could not use complementation experiments to investigate how the patient mutations affect TMEM107 function because even moderate TMEM107(WT) over-expression was toxic to fibroblast cilium formation. However, TMEM107(E45G) and TMEM107(F106del) retained the ability to localise at the TZ, indicating that these mutations exert their pathogenicity by disrupting TMEM107 functions at the TZ, rather than a gross effect on TMEM107 localisations or stability (**Supplementary Fig. 2e**).

Next, we investigated the role of TMEM107 in *C. elegans* TZs. Since loss-of-function alleles were unavailable, we used CRISPR-Cas9[25] to isolate *oq100*, a 27-bp deletion + 14-bp insertion (**Fig. 3a**). This mutation causes a frame-shift and premature stop, which disrupts transmembrane domains 3 & 4, and is likely a null allele as *tmem-107(oq100)::GFP* is not expressed (**Fig. 3b**). *tmem-107(oq100)* mutants appear normal for dye-filling (indirect measure of cilium integrity[26]), cilium length and morphology, and amphid cilium ultrastructure (**Fig. 3c, d** and **Supplementary Fig. 4a**). Also, *tmem-107* mutants possess normal cilia-related chemoattraction and foraging sensory behaviours (**Fig. 3e**). Thus, *tmem-107* loss does not grossly affect cilium structure and function.

We examined if *tmem-107* functions redundantly with other TZ genes and found a synthetic genetic interaction with *nphp-4*. Like worms with mutations in a *C. elegans* MKS *and* an NPHP module gene[10, 15], *tmem-107;nphp-4* double mutants display defects in dye-filling, cilium length and morphology and sensation (**Fig. 3c-e**). The synthetic cilium structure defects were confirmed

by TEM, which showed that 3-5 amphid pore cilia are truncated or missing in *nphp-4;tmem-107* mutants, whereas most or all axonemes are present in the corresponding single mutants (**Supplementary Fig. 4a**). Conversely, *tmem-107* does not synthetically interact with *mksr-1* as assessed by dye filling (**Fig. 3c**), suggesting that *tmem-107* is part of the MKS module whose genes synthetically interact with NPHP module genes but not with each other[10, 15, 27]. Importantly, transgenic expression of *tmem-107(WT)::gfp* in *tmem-107;nphp-4* worms rescued the cilium structure and function phenotypes, confirming phenotypic linkage to *oq100* (**Fig. 3c, e**). Surprisingly, *tmem-107* constructs mimicking the human TMEM107 patient mutations also rescued the *tmem-107;nphp-4* Dyf phenotype (**Supplementary Fig. 4c**), and in agreement with these variants retaining functionality, the encoded proteins localised normally (**Supplementary Fig. 2b**).

TZ ultrastructure was also disrupted in *tmem-107;nphp-4* worms compared to single mutants of *tmem-107* (unaffected) and *nphp-4* (modestly affected; discussed below). In double mutants, Y-links were reduced or missing, and in agreement with loss of these membrane-microtubule connectors, many TZs were undocked from the plasma membrane, frequently extending from ectopic positions within the distal dendrite (**Supplementary Fig. 4b**). Also, *tmem-107;nphp-4* mutant cilia often displayed abnormal vesicle accumulations in the TZ and PCMC regions (**Supplementary Fig. 4b**). Consistent with these structural defects, MKS-2 TZ localisation is disrupted in *tmem-107;nphp-4* worms, but not in single mutants (**Fig. 3f**; *nphp-4* data shown in ref [15]). Furthermore, *tmem-107;nphp-4* neuronal dendrites (phasmids) are frequently short, indicating dendritic tip anchoring defects during dendrite elongation (**Fig. 3g**)[10, 17, 27]. Thus, like

known MKS module components, TMEM-107 functionally interacts with NPHP-4 to facilitate TZ formation, composition and dendrite formation.

As mentioned above, *nphp-4* single mutants display modest ultrastructural TZ defects, despite previous reports that TZs were normal in these worms[16]. In *nphp-4* worms carrying *tm925* (deletion) or *gk529336* (nonsense) mutations, Y-link densities were frequently reduced and sometimes missing, and undocked TZs were observed in ADF and ADL neurons (**Supplementary Fig. 4b**). Thus, *C. elegans nphp-4* alone regulates aspects of Y-link assembly or stability, which agrees with the mild cilium structure defects in corresponding mutants (**Supplementary Fig. 4a**; also ref [16]). We conclude that NPHP and MKS modules are not fully redundant in building TZs in at least some amphid channel neurons.

Similar to known MKS module gene mutants[10, 15, 19], TRAM-1 abnormally leaks into *tmem-107* mutant cilia, demonstrating membrane diffusion barrier defects and altered ciliary composition in these worms (**Fig. 3h**). However, membrane-associated RPI-2 does not leak into *tmem-107(oq100)* cilia indicating the barrier is selectively disrupted (**Fig. 3h**), which contrasts with other MKS module gene mutants, where both TRAM-1 and RPI-2 leak into cilia[10].

To further investigate TMEM107 associations with MKS module proteins, and the evolutionary conservation of these interactions, three complementary approaches were taken. First, we employed protein localisation dependency assays. In *C. elegans*, the MKS module is proposed to assemble hierarchically[10, 13, 15, 19]. In this model, MKS-5 occupies the root of the hierarchy (Layer 1), required for TZ localisation of all MKS module components. In contrast, MKS-3, JBTS-14, MKS-6 and TMEM-17 occupy a peripheral level (Layer 3) not required for localisation of other

206 proteins, whereas MKSR-1, MKSR-2, TMEM-231 and MKS-2 form an intermediate level
207 (Layer 2), required for Layer 3 but not Layer 1 protein targeting. To evaluate if TMEM-107 is
208 part of this model, TMEM-107::GFP localisation was assessed in MKS module gene mutants,
209 and MKS module protein localisations were analysed in *tmem-107(oq100)* mutants. We found
210 TMEM-107 is not required for Layer 1 and most Layer 2 protein localisations, but is required for
211 the organisation of Layer 3 proteins (JBTS-14, TMEM-17), as well as MKS-1 (unassigned layer)
212 and TMEM-231 (layer 2) (**Fig. 4a**). In reverse experiments, TMEM-107 localisation depends on
213 Layer 1, 2 and MKS-1 proteins, but not Layer 3 proteins (**Fig. 4a**). Furthermore, TMEM-107 and
214 NPHP module proteins are not localisation interdependent (**Fig. 4a).** These data show that
215 TMEM-107 recruits an MKS submodule of proteins (TMEM-17, TMEM-231, JBTS-14, MKS-
216 1) and suggest that TMEM-107 occupies an intermediate layer, connecting Layer 2 to a Layer 3
217 subset (**Fig. 4b**). Interestingly, TZ recruitment of TMEM-17 and TMEM-231 by TMEM-107 is
218 independent of its short cytosolic N- and C- termini, suggesting this function is orchestrated by
219 the transmembrane helices or interhelical linkers (**Fig. 4c**). In agreement with our nematode data,
220 depletion of mammalian *Tmem107* also alters the localisation of some MKS module proteins
221 (Tmem231 and Tmem237) but not others (Rpgrip1l) (**Fig. 4d**).
222
223 In the second approach, co-immunoprecipitation (coIP) assays were performed to determine if
224 mammalian TMEM107 biochemically associates with MKS module proteins. GFP-tagged
225 TMEM107 was exogenously co-expressed with FLAG-tagged TMEM216, TMEM231,
226 TMEM17 or TMEM237, or with myc-tagged MKS1, and assessed for associations in IMCD3
227 cells using binding conditions optimized for membrane proteins[28]. We found TMEM107
228 immunoprecipitates full-length TMEM216, TEMM231, TMEM 237 and MKS1, but not

TMEM17 (**Fig. 4e**). For TMEM237, TMEM107 did not detectably interact with an N-terminal cytoplasmic domain of this protein (TMEM237Nt), indicating the TMEM237 association depends on its transmembrane helices or C-terminal cytoplasmic domain (**Fig. 4e**).

In the third approach, we used differential Dollo parsimony[29] to obtain co-evolutionary relationships for MKS module proteins. In agreement with the nematode hierarchy[10, 15], Layer 2 orthologues TMEM216, B9D1 and B9D2 form a co-evolving MKS core (**Fig. 4f**). This 'core' co-evolves with Layer 3 orthologues TMEM67 and CC2D2A, consistent with localisation dependencies in mammalian systems[5], though not in worms (**Fig. 4f**). We also identified a co-evolving TMEM107, TMEM17 and TMEM231 submodule, again agreeing with nematode localisation dependencies (**Fig. 4f**). Unexpectedly, this submodule does include TMEM237 because orthologues are missing in stramenopiles, alveolates and excavates (**Supplementary Fig. 1b**), indicating late incorporation into the MKS module. Also surprising was that RPGRIP1L only marginally co-evolves with the MKS module, despite its central role in module organisation[10, 19] (**Fig. 4f**). This may be explained by low RPGRIP1L sequence conservation and difficulties in assigning orthologues. Finally, the interdependent MKS1 and TMEM107 localisations are not reflected in the co-evolution data. Thus, our evolutionary findings support conserved roles for *C. elegans* TMEM107 in organising an MKS submodule with TMEM231 and TMEM17, but differences in the modular arrangements of TMEM107 with other MKS proteins might exist between species.

The specific TZ localisation of MKS module proteins, together with their requirement for Y-link assembly with NPHP proteins, could suggest this module interacts with Y-links or associated

structures. If true, MKS module proteins should be immobile. Using fluorescence recovery after photobleaching (FRAP) assays in *C. elegans*, photobleaching one half of the TZ signal for TMEM-107::GFP, MKS-2::GFP and MKS-6::GFP resulted in no signal recovery (30 minutes), indicating that the non-bleached TZ pool is immobile (**Fig. 5a**). Also, no recovery was observed when entire TZ signals were photobleached (MKS-2::GFP), demonstrating that MKS module proteins possess slow TZ entry kinetics (**Supplementary Fig. 5a**). This immobility depends on other MKS module proteins because MKS-2 is highly mobile in *mksr-1* (B9D1) mutants (**Fig. 5b**). Therefore, we conclude that transmembrane (TMEM-107, MKS-2) and membrane-associated (MKS-6) proteins are anchored at the TZ membrane. Furthermore, at least for MKS-2, anchoring requires an intact MKS module.

Next, we used stimulated emission depletion (STED) super-resolution microscopy to further investigate *C. elegans* MKS (TMEM-107, MKS-2 and TMEM-231) and NPHP (NPHP-1) module protein distributions within TZs. Side view TZ images (axial orientation) revealed that these proteins are periodically distributed along the axial plane, frequently appearing as multiple independent rings (partial or complete), or possibly spiral structures **(Fig. 5c** and **Supplementary Fig. 5b)**. However, in some images with slightly better resolution, individual dots of signal were evident within ring-like domains (**Fig. 5c**; arrowheads), suggesting these domains consist of multiple discrete protein clusters. Punctate signals were also observed for NPHP-1, suggesting MKS and NPHP module proteins possess similar axial TZ distributions (**Fig. 5c**). We were unable to determine protein distributions from true traverse (radial) views of the TZ because of the orientation of worms (and cilia) on the imaging slide.

We also performed STED imaging on endogenous RPGRIP1L and TMEM67 in human cells (RPTEC-TERT1). Imaging of TZs in radial orientation revealed that these proteins can form discrete clusters arranged as a complete or near complete single shallow ring **(Fig. 5d** and **Supplementary Fig. 5c)**. Although the number of resolved clusters was variable, rings with 7-8, or possibly 9 clusters were observed. In addition, RPGRIP1L ring diameters were significantly narrower than TMEM67 ring diameters (**Fig. 5d**). As an alternative approach, we imaged endogenous RPGRIP1L using direct stochastic optical reconstruction microscopy (dSTORM). Similar to our STED findings, dSTORM imaging of RPGRIP1L revealed a single shallow ring of TZ signal, comprised of at least 7-8 independent punctae **(Fig. 5e** and **Supplementary Fig. 5d)**. Thus, both mammalian RPGRIP1L and TMEM67 appear to be organised as discrete clusters within ring-like domains of differing diameters, indicating distinct radial positioning at the TZ membrane (TMEM67) and core (RPGRIP1L). The periodicity of the clusters approaches the 9 fold symmetry of Y-links, suggesting possible association with these structures.

Together, our super-resolution imaging indicates that MKS module proteins occupy periodic radial and axial subdomains of the TZ core and membrane (**Fig. 5f**). The nematode axial pattern is reminiscent of the ciliary necklace, a conserved TZ membrane specialisation comprised of 1-7 rows of intramembrane leaflet particles identified almost 40 years ago[11, 12] **(Fig. 5f)**. In further support of a necklace association, *C. elegans* MKS module proteins are anchored at the TZ membrane, and the periodic radial distribution of mammalian TMEM67 at the TZ membrane approaches the periodicity of Y-links implicated in necklace formation[11]. In one model, Y-links would anchor MKS module proteins at the necklace (**Fig. 5f**). However, a tomographic (TEM) reconstruction of a portion of the *C. elegans* TZ indicates that Y-links are continuous structures

298 running the TZ length (**Supplementary Video 1** and **Fig. 5f;** also ref [18]); thus, any association
299 of MKS proteins with Y-links would occur at various axial positions along continuous Y-link
300 sheets (**Fig. 5f**). Unfortunately, we could not identify a necklace pattern for mammalian MKS
301 module proteins because of spatial constraints and resolution limits conferred by the short TZ
302 (<200 nm). Future super-resolution imaging on mammalian cells with longer TZs should clarify
303 the nematode observations.
304
305 In summary, we have identified TMEM107 as a ciliopathy TZ protein and conserved MKS
306 module component. Causality of the mutations identified here to JBTS and OFDVI is supported
307 by very recent reports of additional *TMEM107* mutations linked to MKS and OFD [30, 31].
308 Furthermore, we show that ciliopathy proteins are anchored at the TZ membrane, and display
309 periodic radial and axial distributions at the TZ core and membrane. In addition, our finding that
310 MKS-2 is mobile in *mksr-1* (B9D1) mutant cilia indicates that protein anchoring at the TZ is
311 important for barrier functions. Strikingly, membrane diffusion barriers of the axon initial
312 segment (AIS) that limit free exchange of phospholipids also depend on anchored membrane
313 proteins[32], and sodium channels within the AIS membrane are coordinately localised with evenly
314 spaced (180-190 nm) actin-spectrin cytoskeletal rings[33]. Thus, the TZ and AIS cellular
315 compartments may share comparable molecular organisation underpinning common mechanisms
316 of barrier function. Finally, our bioinformatics analysis to discover genes of the TZ compartment
317 goes beyond most comparative genomics studies that focus on the entire cilium. Indeed, distinct
318 phylogenetic distributions exist for other ciliary modules such as IFT-A, IFT-B and BBSome
319 assemblies[34]. Thus, exploitation of genomics data can help to disentangle ciliary modules,
320 ultimately leading to greater understanding of ciliary transport, signaling and disease.

ACKNOWLEDGEMENTS

This work was funded via the European Community's Seventh Framework Programme FP7/2009 (SYSCILIA grant agreement 241955; to OEB, MAH, RHG, CAJ, and Gencodys; to MAH), Science Foundation Ireland (11/PI/1037 to OEB), the Dutch Kidney Foundation CP11.18 "KOUNCIL" (to RHG), the GIS-Institut des Maladies Rares (HTS; to C. T-R), the French Fondation for Rare Disease (to CTR), the Virgo consortium (FES0908; to MAH), the Netherlands Genomics Initiative (050-060-452; RvdL; to MAH), the French Ministry of Health (PHRC national 2010-A01014-35 and 2013; to CTR), the Fondation pour la Recherche Médicale (DEQ20130326532; to SS), the Regional Council of Burgundy (to CTR), a Sir Jules Thorn Award for Biomedical Research (JTA/09; to CAJ), and the UK Medical Research Council (MR/K011154/1 to CAJ, and MR/K015613/1; to MP). We thank the patients and their families for their participation. We also thank the NHLBI GO Exome Sequencing Project and its ongoing studies which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926) and the Heart GO Sequencing Project (HL-103010). We thank Michel Leroux (Simon Fraser University, Canada), Brad Yoder (University of Alabama, USA), the *Caenorhabditis elegans* genetic center (Minnesota, USA), the National Bioresource project (Tokyo, Japan), the International *C. elegans* gene knockout consortium, and the *C. elegans* Million Mutation Project for nematode reagents. We are grateful to Christian Eggeling and Christopher Lagerholm (Weatherall Institute of

Molecular Medicine and the Wolfson Imaging Center, Oxford, UK) for assistance with STED microscopy, Dimitri Scholz and Tiina Toivonen (UCD Conway Institute imaging facility, Dublin, IRL) for imaging support, and Remko Dijkstra (Scientific Volume Imaging bv, Hilversum, NL) for assistance with STED image deconvolution. We also thank Alexa Cleasby (Faculty of Biological Sciences, University of Leeds, Leeds, UK) for help with developing the dSTORM technique, Dr. Ben Chih (Genentech Inc., South San Francisco, CA, USA) for the kind gift of polyclonal anti-TMEM17 and TMEM231 antibodies, and Dr. Tara McMorrow (University College Dublin, Dublin, Ireland) for the generous gift of the RPTEC/TERT1 cells. We thank Diana Rodriguez (Trousseau hospital, Paris) for assistance with analysis of brain MRIs. The dSTORM microscope was generously funded by alumnus Michael Beverly, in support of the University of Leed's 'making a world of difference campaign'.

AUTHOR CONTRIBUTIONS

N.J.L., J.E.K., K.G and O.E.B. performed and interpreted experiments with *C. elegans*. T.J.P.v.D., R.v.d.L. and M.A.H. performed all bioinformatics analyses. A-L.B., L.B., D.D., T.A-B., S.S., and C.T-R collected, purified patient samples, performed exome sequencing and analysed sequencing data. N.J.L., S.K., G.J.M. performed the STED imaging. A.C., M.P and C.A.J conducted the dSTORM imaging and processing. K.S., S.K, G.G.S., K.M.W., and R.H.G. conducted transfection and immunofluorescence microscopy in mammalian cells. K.S. and C.A.J. contributed the co-immunoprecipitation experiments. J-B.R., L.F. and C.T-R. diagnosed and referred patients. The co-corresponding authors shared supervision of the work. M.A.H. uncovered TMEM107 as a candidate ciliary gene, and directed the bioinformatics work. C.T-R. collated JBTS and OFD patient samples, performed clinical characterisation and directed the sequencing. O.E.B. directed research, analysed and collated data for the manuscript. O.E.B., M.A.H, R.H.G., and C.A.J. conceived and executed the study, and O.E.B., N.J.L., T.J.v.D., and M.A.H. wrote the manuscript.

COMPETING FINANCIAL INTEREST STATEMENT

The authors declare no competing financial interest in relation to this work.

FIGURE LEGENDS

Figure 1. A weighted co-expression approach to discover TZ genes identifies TMEM107 as a TZ protein. (a) Frequency histogram of binned human gene co-expression scores, derived from weighted analyses of gene expression datasets using a training set of 20 known TZ genes (Supplementary Table 1). Frequencies normalised to compare different distributions. SYSCILIA gold standard genes[21] in yellow; TZ gene training in blue; all other genes in grey hatched. Box-plots display median and quartiles for histogram distributions. Whiskers (hashed lines) denote the minimum and maximum extent of the dataset. **(b)** Recall performance (also known as sensitivity) of the co-expression approach retrieves known TZ (blue lines) and ciliary (yellow lines) genes. The graph shows that TZ genes can be retrieved compared to ciliary genes. Inset: recall performance for top 200 ranked genes. Ciliary genes taken from the SYSCILIA gold standard[21]. **(c)** *C. elegans* TMEM-107::GFP localises at the TZ. Shown are fluorescence images from worms expressing TMEM-107::GFP alone (left panels) or together with an ARL-13::tdTomato reporter (right panels). Left panels; accumulation of TMEM-107::GFP at the ciliary base region of 12 bilateral amphid cilia (amp; brackets), labial and CEP cilia (subset denoted by arrowheads), bilateral phasmid cilia (arrowheads) and the right-sided PQR cilium (asterisk) in the tail. Note that head schematic only shows a subset of the hermaphrodite's ciliated head neurons. Right panels; TMEM-107::GFP localises immediately proximal to middle

segment (ms)-restricted ARL-13::tdTomato. Image shows all four phasmid cilia (left and right). Schematic denotes major subcompartments in phasmid cilia with microtubule doublets (only two shown) in the TZ and middle segments, and microtubule singlets in the distal segment (ds). den; dendrite. Bars; 2 μm (left two images), 1 μm (right images). **(d)** Human TMEM107 localises at the TZ. Shown are images of hTERT-RPE1 cells stably expressing GFP-tagged human TMEM107 (green) at a low level, costained with antibodies for ciliary axonemes (polyglutamylated tubulin; PolyGluTub) and the TZ (RPGRIP1L, TMEM67). Bars; 5 μm.

Figure 2. *TMEM107* regulates mammalian ciliogenesis and is mutated in OFDVI and JBTS individuals. (a) IMCD3 cells transfected with *Tmem107* siRNA possess reduced *Tmem107* mRNA expression (vs scrambled siRNA control; qPCR data) and reduced mean ciliary frequency. Data represents mean ± S.E.M (n=350 cells, 1 experiment). *$p<0.05$ (unpaired t-test; vs control). **(b)** When grown in 3-D culture, IMCD3 cells transfected with *Tmem107* siRNA form spheroids with a reduced mean size. Cilia (orange) stained for acetylated alpha-tubulin; cell junctions (green) stained for beta-catenin. Data represents mean ± S.E.M (n=25 spheroids pooled from 2 independent experiments). *$p<0.05$ (unpaired t-test; vs control). Bar; 5 μm. **(c)** Schematic of human TMEM107 protein showing the position of identified patient mutations. Grey boxes correspond to the transmembrane domains. Mat; maternal, Pat; paternal, NA; not available. **(d)** Brain MRIs (axial views) showing the molar tooth sign, linked to elongated, thick and mal-oriented superior peduncles (white arrows) and hypoplastic vermis. **(e)** Brain MRI showing a dysplastic and highly hypoplastic vermis in sagittal view. A secondary enlargement of the fourth ventricle with displacement of the fastigium is also evident. **(f)** Brain MRI (axial view) showing heterotopias, enlarged lateral ventricles and polymicrogyria. (**g**) Brain MRI (sagittal view)

showing enlarged posterior fossa (asterisk) with a cystic dilation of the fourth ventricle, a severe midbrain dysplasia and a thin corpus callosum with enlarged ventricles. **(h-j)** Shown in h are fibroblasts derived from skin biopsies of healthy control (wild type; WT) and patient 3 (JBTS) immunostained for cilia using antibodies against ARL13B (red; ciliary membrane) and acetylated tubulin (white; axonemal microtubules). Compared to control cells, JBTS cell cilia possess reduced lengths (i) and frequencies (j). Data represents mean ± S.E.M (n=10 (i) and 25 (j) cells};data represent 1 of 3 independent experiments)**.** * $p<0.05$ (unpaired t-test; vs WT), ** $p<0.01$ (unpaired t-test; vs WT), bars; 5 μm.

Figure 3. *C. elegans tmem-107* controls diffusion barrier integrity and functions with *nphp-4* to regulate ciliary and TZ structural integrity. (a) Schematic of *oq100* Indel mutation in *tmem-107* gene. Exons denoted by grey boxes (numbers; nucleotide positions). del; deletion, ins; insertion. **(b)** *oq100* mutation disrupts TMEM-107 expression. Shown are amphid cilia TZs in worms expressing GFP-tagged wild-type or mutant (*oq100*) TMEM-107. Bar; 2 μm (images identically scaled). **(c)** Dye filling assay (measure of cilium integrity for 6 amphid (head) and 2 phasmid (tail) ciliated neurons) showing dye-filling defects (Dyf) in *tmem-107(oq100);nphp-4(tm925)* double mutants, but not single mutants, or a *tmem-107(oq100);mkrs-1(tm3083)* double mutant. Dyf phenotype is rescued by expression of wild-type *tmem-107* (GFP-tagged; see Figure 1c). Bars; 10 μm. **(d)** Images of ASER neuronal cilia from worms expressing a *gcy-5p::gfp* that stains the ASER neuron. Numbers refer to cilium length measurements; mean ± S.E.M (n=28 (N2), 44 (*tmem-107*), 46 (*nphp-4*) and 81 (*tmem-107;nphp-4*) cilia). Brackets denote ciliary axonemes (cil). Arrowhead; occasional break in GFP staining observed only in double mutant. den; dendrite. * $p<0.01$ (unpaired t-test; vs WT), ** $p=0.01$ (unpaired t-test; vs *nphp-4*), Bars; 3

μm. **(e)** *tmem-107(oq100);nphp-4(tm925)* double mutants possess defects in cilia-related behaviours. Shown are population assays of isoamyl alcohol (IAA) attraction and single worm foraging assays. Data represents mean ± S.E.M. For IAA assays, n=30 (N2), 20 (*tmem-107*), 22 (*nphp-4*) and 29 (*tmem-107;nphp-4*); For foraging assays, n= 44 (N2), 43 (*tmem-107*), 63 (*nphp-4*), 54 (*tmem-107;nphp-4*) and 37 (*tmem-107;nphp-4;Ex[tmem-107(wt)]* indpependent experiments, respectively* $p<0.01$ (unpaired t-test; vs WT), ** $p<0.01$ (unpaired t-test; vs *tmem-107;nphp-4*). CI; chemotaxis index. **(f)** TZ composition is altered in *tmem-107;nphp-4* double mutants. Shown are phasmid cilia from worms expressing TZ-localised MKS-2::GFP and periciliary membrane-localised, TRAM-1::tdTomato (asterisk). Bars; 2 μm. **(g)** *tmem-107(oq100);nphp-4(tm925)* double mutants possess short phasmid (PHA/B) dendrites and misplaced cilia. Neurons stained with OSM-6(IFT52)::GFP. Cil; ciliary axonemes, den; dendrite, cb; cell bodies (also denoted by asterisks). Brackets denote PHA/B cilia. Bars; 5 μm. **(h)** TZ membrane diffusion barrier is selectively disrupted in *tmem-107(oq100)* mutants. Shown are phasmid cilia from worms expressing TRAM-1::tdTomato (and MKS-2::GFP; marks TZ) (left images) or RPI-2::GFP (and XBX-1::tdTomato; marks cilia) (right images). TRAM-1 (translocon subunit) and RPI-2 (retinitis pigmentosa 2) are excluded from wild-type (WT) cilia, whereas TRAM-1 (but not RPI-2) leaks into *tmem-107(oq100)* cilia. Asterisk; TZ localization of MKS-2, pcm; periciliary membrane, cil; ciliary axoneme. Bars; 2 μm.

Figure 4. Evolutionary conserved association of TMEM107 with the TZ-localised MKS module. (a) Phasmid TZ localisations of GFP-tagged MKS and NPHP module proteins in WT and *tmem-107(oq100)* mutant worms, and TMEM-107::GFP in MKS and NPHP mutants. Bar; 1 μm (all images similarly scaled). mis-loc.; mislocalised. **(b)** Schematic summarising TZ

528 localisation dependencies in (a). TMEM-107 positioned at an intermediate level within a
529 hierarchical three layer (L1-3) MKS module assembly model (drawn based on refs[10, 13, 15, 19];
530 MKS-1 'unassigned' because hierarchical analysis has not yet been conducted using an *mks-1*
531 null allele). Human orthologues denoted in brackets. **(c)** Expression of TMEM-107::RFP with
532 disrupted cytosolic N- or C- termini (nTMEM-107, cTMEM-107; see methods) rescues
533 mislocalised TMEM-17::GFP and TMEM-231::GFP in *tmem-107(oq100)* mutants. Shown are
534 phasmid cilia TZs. Bar; 0.5 μm. **(d)** *Tmem107* depletion (siRNA) in IMCD3 cells disrupts
535 relative localisations of endogenous MKS module proteins. Cells double-stained as indicated and
536 colocalisation determined as an R_{total} Pearson correlation value (FIJI "Colocalization Threshold"
537 plugin). In *Tmem107* depleted cells, Rpgrip1l localisation is unaffected (relative to basal body
538 (BB) γ-tubulin), whereas Tmem231and Tmem237 proteins shift (black arrows) relative to γ-
539 tubulin or Rpgrip1l. Data in graph represents mean ± S.E.M (n=150 cells pooled from 3
540 independent experiments). siScr; siRNA scrambled control. ** $p<0.01$, *$p<0.05$ (unpaired t-test;
541 vs siScr control). Bar; 1 μm. **(e)** Coimmunoprecipitation (coIP) assays in IMCD3 cells. Upper
542 panels, lanes 1-4: input material from whole cell extracts (WCEs) transfected with the indicated
543 constructs and immunoblotted (IB) with anti-GFP or anti-FLAG. Lanes 5-8: proteins
544 immunoprecipitated (IP) by an irrelevant antibody (irr. Ab; anti-MICU3) or anti-GFP, and then
545 immunoblotted for FLAG or GFP. IgG heavy chain (HC) and light chain (LC) in coIPs are
546 indicated. Asterisks (*) mark non-specific proteins. Lower panels, lanes 9-12: input WCE
547 showing expression of FLAG-TMEM231, FLAG-TMEM17 and c myc-MKS1. Lanes 13-21: IPs
548 with antibodies against MKS1 (lane 14), TMEM231 (231; lane 17) and TMEM17 (17; lane 20)
549 and then immunoblotted as indicated. Note that although TMEM107 co-IP's TMEM231,
550 TMEM231 does not co-IP detectable levels of TMEM107. **(f)** Co-evolution relationships

between MKS components using differential Dollo parsimony that counts along a phylogenetic tree how often two genes are lost independently from each other. Thickness and color gradient indicate strong co-evolution. Edges with differential Dollo parsimony scores >11 are not shown. Dashed box: co-evolving MKS submodule.

Figure 5. Anchoring and periodic distributions of MKS module proteins within the TZ. (a) GFP-tagged TMEM-107, MKS-2 and MKS-6 are immobile within the *C. elegans* TZ. Shown are fluorescence recovery after photobleaching (FRAP) curves and representative time-lapse images after photobleaching one half of a TZ signal (boxed region). Data points represented as mean ± S.E.M. (n=3 (MKS-6) or 4 (TMEM-107, MKS-2) independent experiments). Bar; 500 nm. **(b)** *C. elegans* MKS-2 immobility depends on MKS module proteins. Shown is a FRAP curve and representative time-lapse images (phasmid cilia) after photobleaching MKS-2::GFP signals (boxed region) in an *mksr-1* mutant. Asterisk; periciliary membrane. Data points represented as mean ± S.E.M. (n=4 independent experiments). au; arbitrary units, Bar; 2 μm. **(c)** Arrowheads; independent signal clusters within a ring-like domain. Bars; 200 nm (high magnification images), 500 nm (low magnification images). **(d)** STED images of endogenous human RPGRIP1L and TMEM67 in renal RPTEC cells showing clusters (arrowheads) of protein in a single ring of differing diameters (mean ± S.D.) at the TZ. Corresponding confocal images co-stained for cilia with polyglutamylated tubulin antibody. *p=0.001 (unpaired t-test; vs TMEM67). Bars; 100 nm. **(e)** dSTORM of human RPGRIP1L (visualised with AlexaFluor647) with 10 nm binning, image smoothing and contrast enhancement in FIJI (raw images shown in Supplementary Figure 5d), showing periodic localisation (arrowheads) in a loose ring at the TZ. Image depth-coded by colour. Z-axis scale bar (nm) on right. Bar; 100 nm. **(f)** Models. MKS module proteins (and *C.*

574 *elegans* NPHP-1) occupy periodic radial and axial TZ subdomains. Mammalian RPGRIP1L and
575 TMEM67 localise as independent clusters, forming a single ring domain at the TZ core
576 (RPGRIP1L) or membrane (TMEM67). *C. elegans* MKS and NPHP proteins also localise as
577 discrete independent clusters, forming multiple ring domains (or possible spiral domains) along
578 the TZ length. The nematode axial distribution may correspond to the ciliary necklace (TEM
579 example from ref[12]). Periodicity and immobility of MKS module proteins suggests association
580 with Y-links, which form extended sheets in *C. elegans* (Supplementary Video 1) and are
581 implicated in necklace formation.
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596

REFERENCES

1. Reiter, J., Blacque, O. & Leroux, M. The base of the cilium: roles for transition fibres and the transition zone in ciliary formation, maintenance and compartmentalization. *EMBO reports* **13**, 608-618 (2012).
2. Goetz, S.C. & Anderson, K.V. The primary cilium: a signalling centre during vertebrate development. *Nat Rev Genet* **11**, 331-344 (2010).
3. Blacque, O.E. & Sanders, A.A. Compartments within a compartment: what C. elegans can tell us about ciliary subdomain composition, biogenesis, function, and disease. *Organogenesis* **10**, 126-137 (2014).
4. Hsiao, Y.C., Tuz, K. & Ferland, R.J. Trafficking in and to the primary cilium. *Cilia* **1**, 4 (2012).
5. Chih, B. *et al.* A ciliopathy complex at the transition zone protects the cilia as a privileged membrane domain. *Nature cell biology* **14**, 61-72 (2011).
6. Craige, B. *et al.* CEP290 tethers flagellar transition zone microtubules to the membrane and regulates flagellar protein content. *The Journal of cell biology* **190**, 927-940 (2010).
7. Garcia-Gonzalo, F.R. *et al.* A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. *Nature genetics* **43**, 776-784 (2011).

8. Hu, Q. *et al.* A septin diffusion barrier at the base of the primary cilium maintains ciliary membrane protein distribution. *Science (New York, N.Y* **329**, 436-439 (2010).
9. Kee, H.L. *et al.* A size-exclusion permeability barrier and nucleoporins characterize a ciliary pore complex that regulates transport into cilia. *Nature cell biology* **14**, 431-437 (2012).
10. Williams, C.L. *et al.* MKS and NPHP modules cooperate to establish basal body/transition zone membrane associations and ciliary gate function during ciliogenesis. *The Journal of cell biology* **192**, 1023-1041 (2011).
11. Gilula, N.B. & Satir, P. The ciliary necklace. A ciliary membrane specialization. *The Journal of cell biology* **53**, 494-509 (1972).
12. Heller, R.F. & Gordon, R.E. Chronic effects of nitrogen dioxide on cilia in hamster bronchioles. *Experimental lung research* **10**, 137-152 (1986).
13. Roberson, E.C. *et al.* TMEM231, mutated in orofaciodigital and Meckel syndromes, organizes the ciliary transition zone. *The Journal of cell biology* **209**, 129-142 (2015).
14. Cevik, S. *et al.* Active Transport and Diffusion Barriers Restrict Joubert Syndrome-Associated ARL13B/ARL-13 to an Inv-like Ciliary Membrane Subdomain. *PLoS genetics* **9**, e1003977 (2013).
15. 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. *American journal of human genetics* **89**, 713-730 (2011).
16. Jauregui, A.R., Nguyen, K.C., Hall, D.H. & Barr, M.M. The Caenorhabditis elegans nephrocystins act as global modifiers of cilium structure. *The Journal of cell biology* **180**, 973-988 (2008).

17. Williams, C.L., Winkelbauer, M.E., Schafer, J.C., Michaud, E.J. & Yoder, B.K. Functional redundancy of the B9 proteins and nephrocystins in Caenorhabditis elegans ciliogenesis. *Mol Biol Cell* **19**, 2154-2168 (2008).

18. Schouteden, C., Serwas, D., Palfy, M. & Dammermann, A. The ciliary transition zone functions in cell adhesion but is dispensable for axoneme assembly in C. elegans. *The Journal of cell biology* **210**, 35-44 (2015).

19. Jensen, V.L. *et al.* Formation of the transition zone by Mks5/Rpgrip1L establishes a ciliary zone of exclusion (CIZE) that compartmentalises ciliary signalling proteins and controls PIP2 ciliary abundance. *The EMBO journal* (2015).

20. Baughman, J.M. *et al.* A computational screen for regulators of oxidative phosphorylation implicates SLIRP in mitochondrial RNA homeostasis. *PLoS genetics* **5**, e1000590 (2009).

21. van Dam, T.J., Wheway, G., Slaats, G.G., Huynen, M.A. & Giles, R.H. The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. *Cilia* **2**, 7 (2013).

22. Barker, A.R., Renzaglia, K.S., Fry, K. & Dawe, H.R. Bioinformatic analysis of ciliary transition zone proteins reveals insights into the evolution of ciliopathy networks. *BMC genomics* **15**, 531 (2014).

23. Christopher, K.J., Wang, B., Kong, Y. & Weatherbee, S.D. Forward genetics uncovers Transmembrane protein 107 as a novel factor required for ciliogenesis and Sonic hedgehog signaling. *Dev Biol* **368**, 382-392 (2012).

24. Giles, R.H., Ajzenberg, H. & Jackson, P.K. 3D spheroid model of mIMCD3 cells for studying ciliopathies and renal epithelial disorders. *Nature protocols* **9**, 2725-2731 (2014).

25. Friedland, A.E. *et al.* Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nat Methods* **10**, 741-743 (2013).

26. Starich, T.A. *et al.* Mutations affecting the chemosensory neurons of Caenorhabditis elegans. *Genetics* **139**, 171-188 (1995).

27. Williams, C.L., Masyukova, S.V. & Yoder, B.K. Normal ciliogenesis requires synergy between the cystic kidney disease genes MKS-3 and NPHP-4. *J Am Soc Nephrol* **21**, 782-793 (2010).

28. Valente, E.M. *et al.* Mutations in TMEM216 perturb ciliogenesis and cause Joubert, Meckel and related syndromes. *Nature genetics* **42**, 619-625 (2010).

29. Kensche, P.R., van Noort, V., Dutilh, B.E. & Huynen, M.A. Practical and theoretical advances in predicting the function of a protein by its phylogenetic distribution. *Journal of the Royal Society, Interface / the Royal Society* **5**, 151-170 (2008).

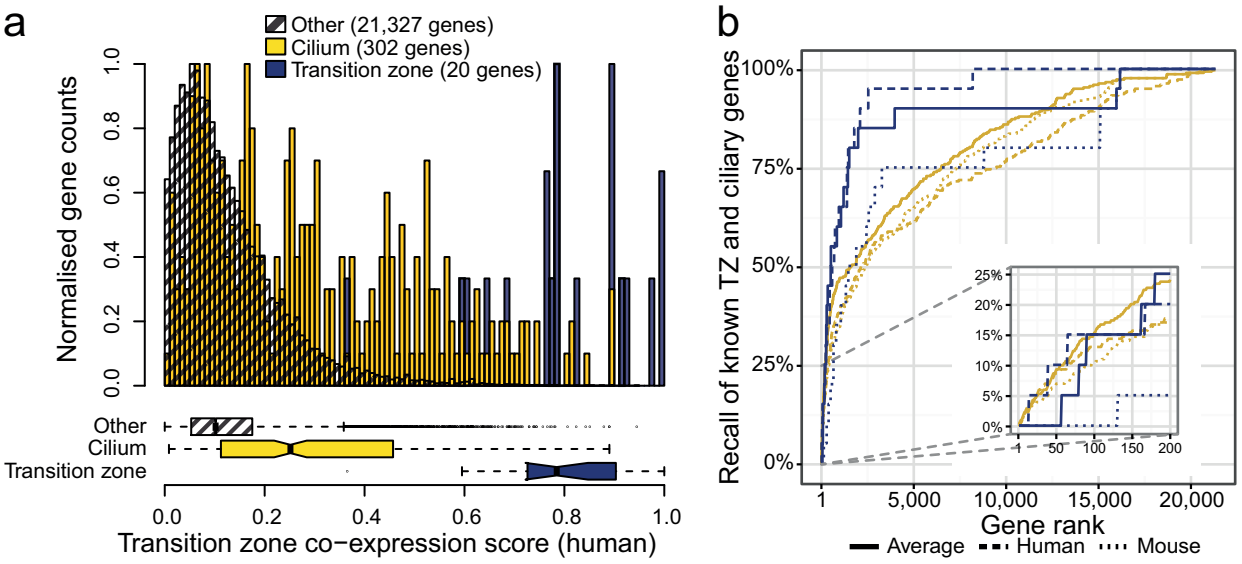
30. Iglesias, A. *et al.* The usefulness of whole-exome sequencing in routine clinical practice. *Genetics in medicine : official journal of the American College of Medical Genetics* **16**, 922-931 (2014).

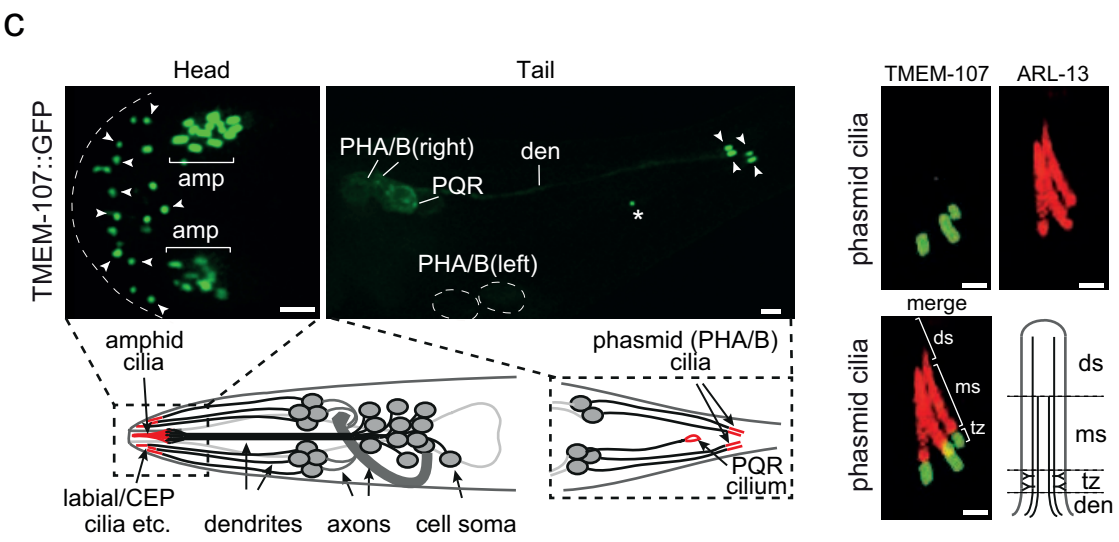
31. Shaheen, R. *et al.* Identification of a Novel MKS Locus Defined By TMEM107 Mutation. *Human molecular genetics* (2015).

32. Nakada, C. *et al.* Accumulation of anchored proteins forms membrane diffusion barriers during neuronal polarization. *Nature cell biology* **5**, 626-632 (2003).

33. Xu, K., Zhong, G. & Zhuang, X. Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons. *Science (New York, N.Y* **339**, 452-456 (2013).
34. van Dam, T.J. *et al.* Evolution of modular intraflagellar transport from a coatomer-like progenitor. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 6943-6948 (2013).

Figure 1-Lambacher *et al.*





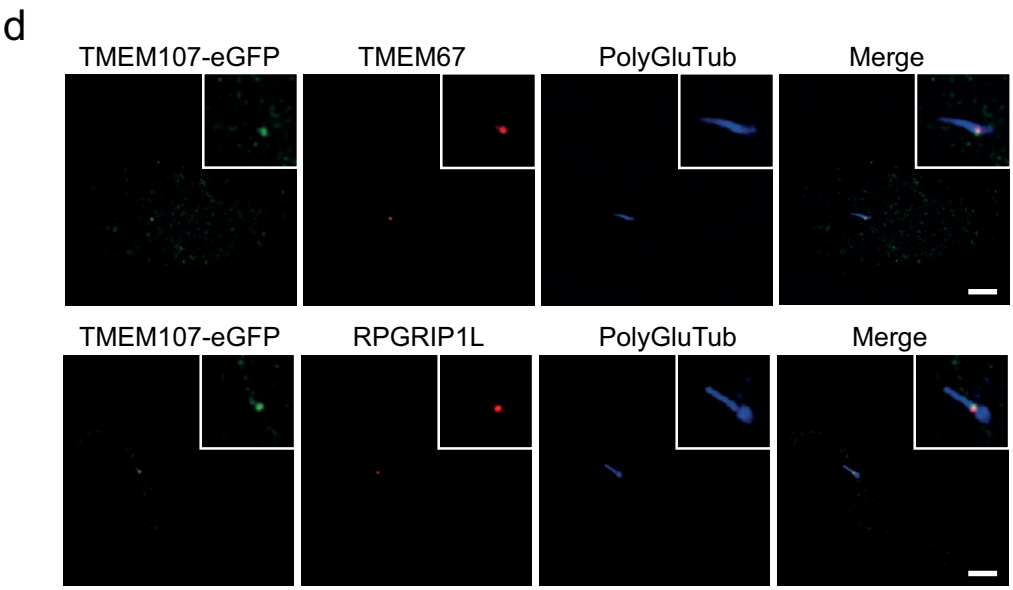


Figure 2-Lambacher *et al.*

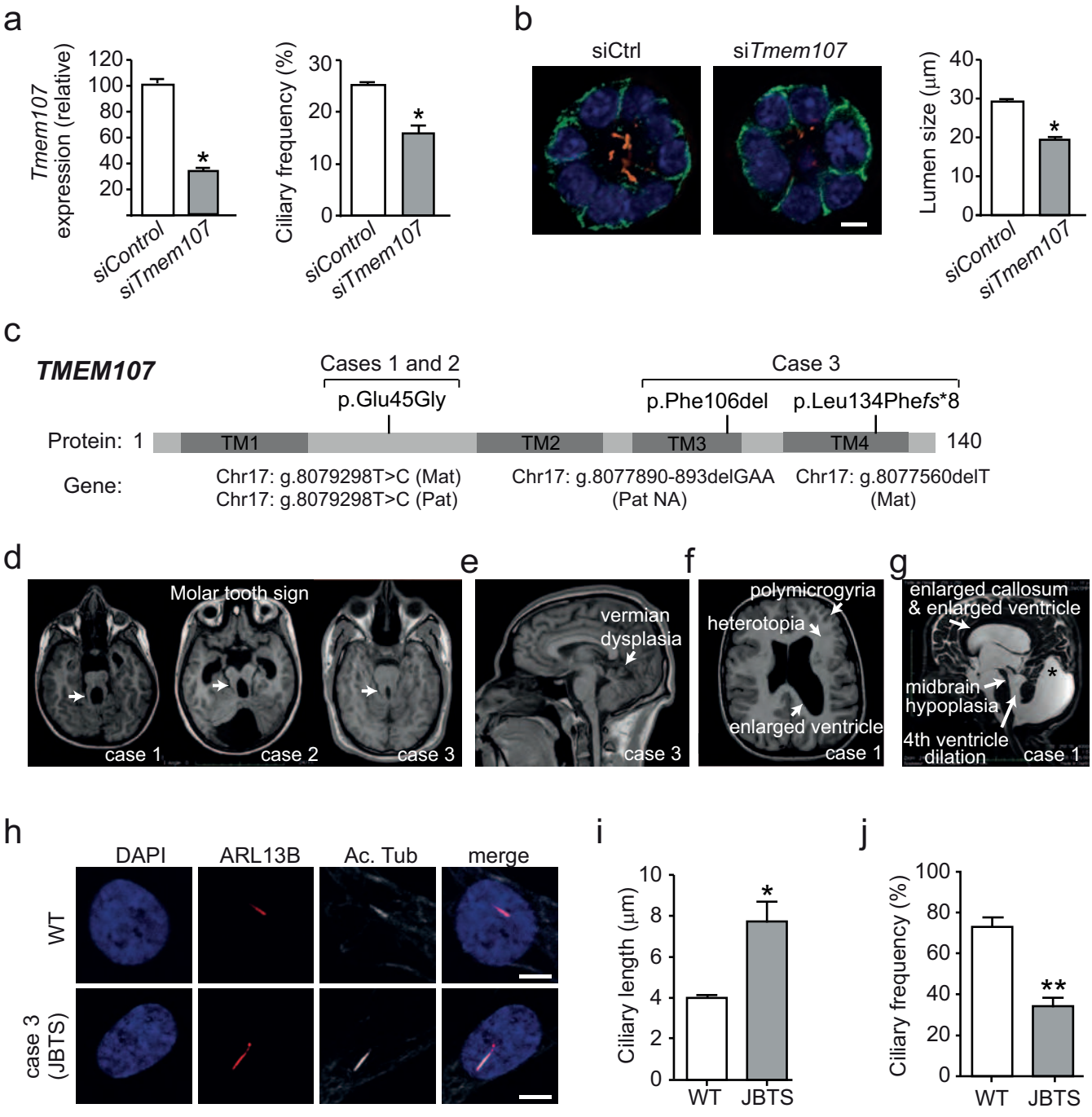


Figure 3-Lambacher *et al.*

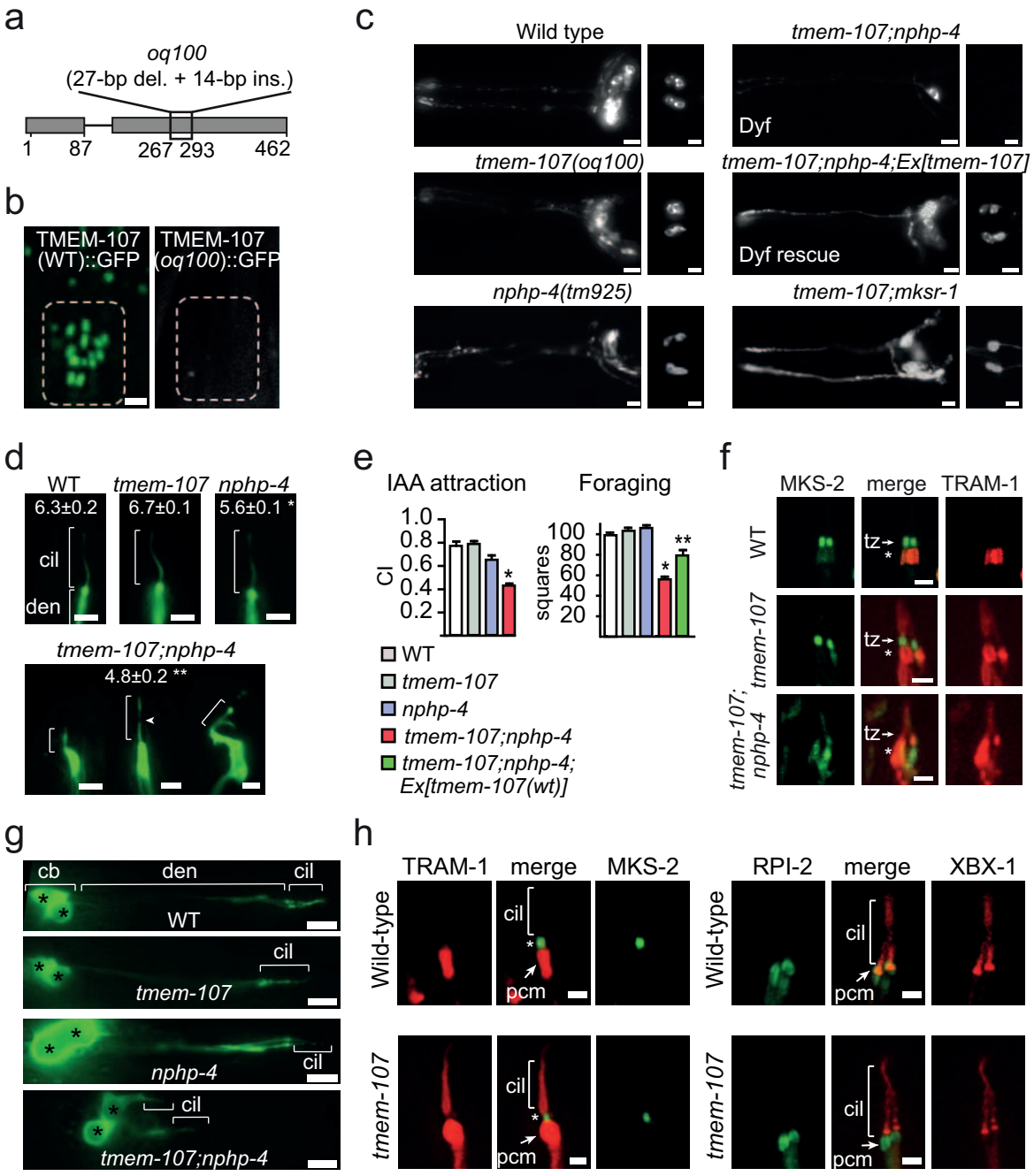


Figure 4-Lambacher *et al.*

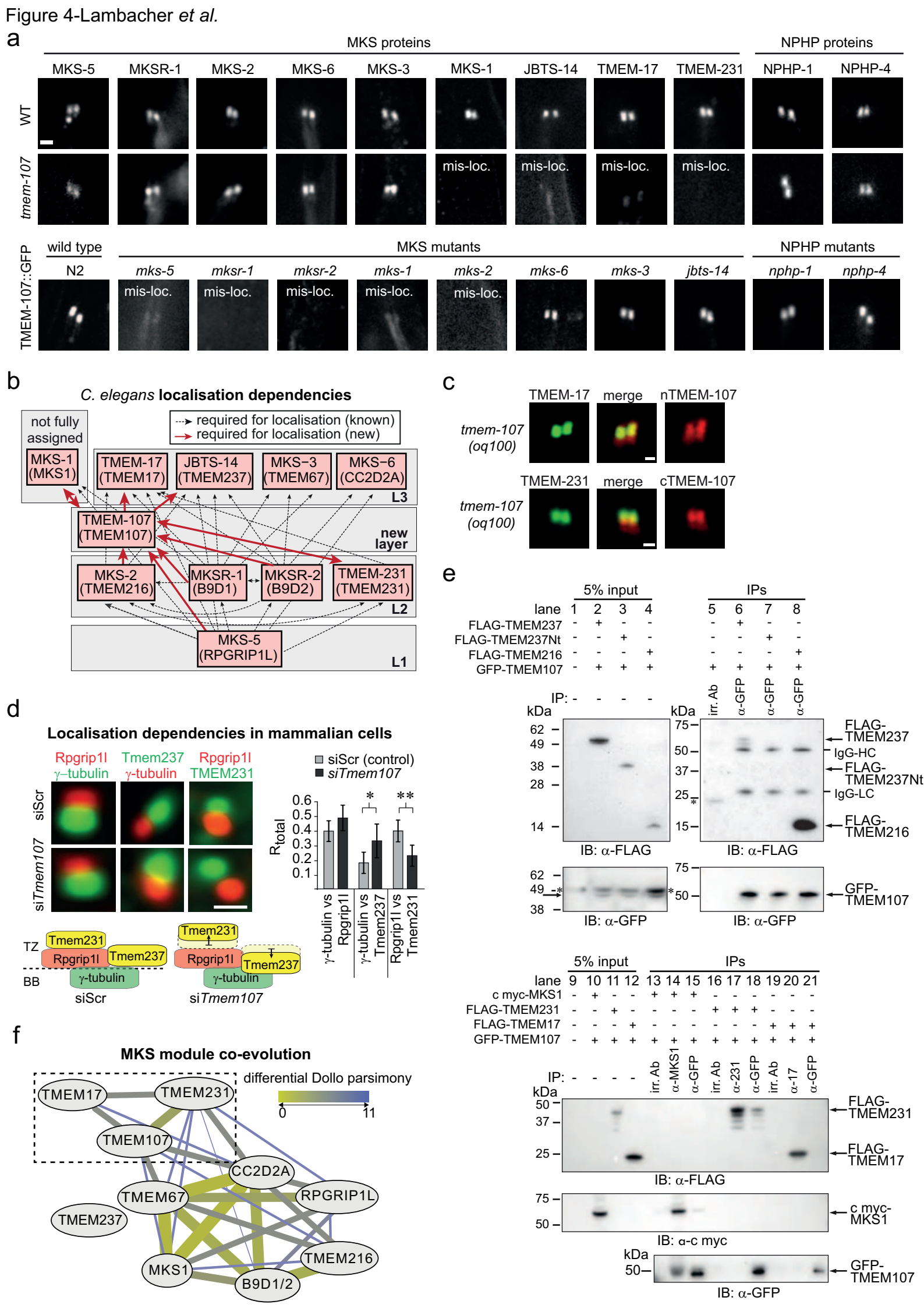
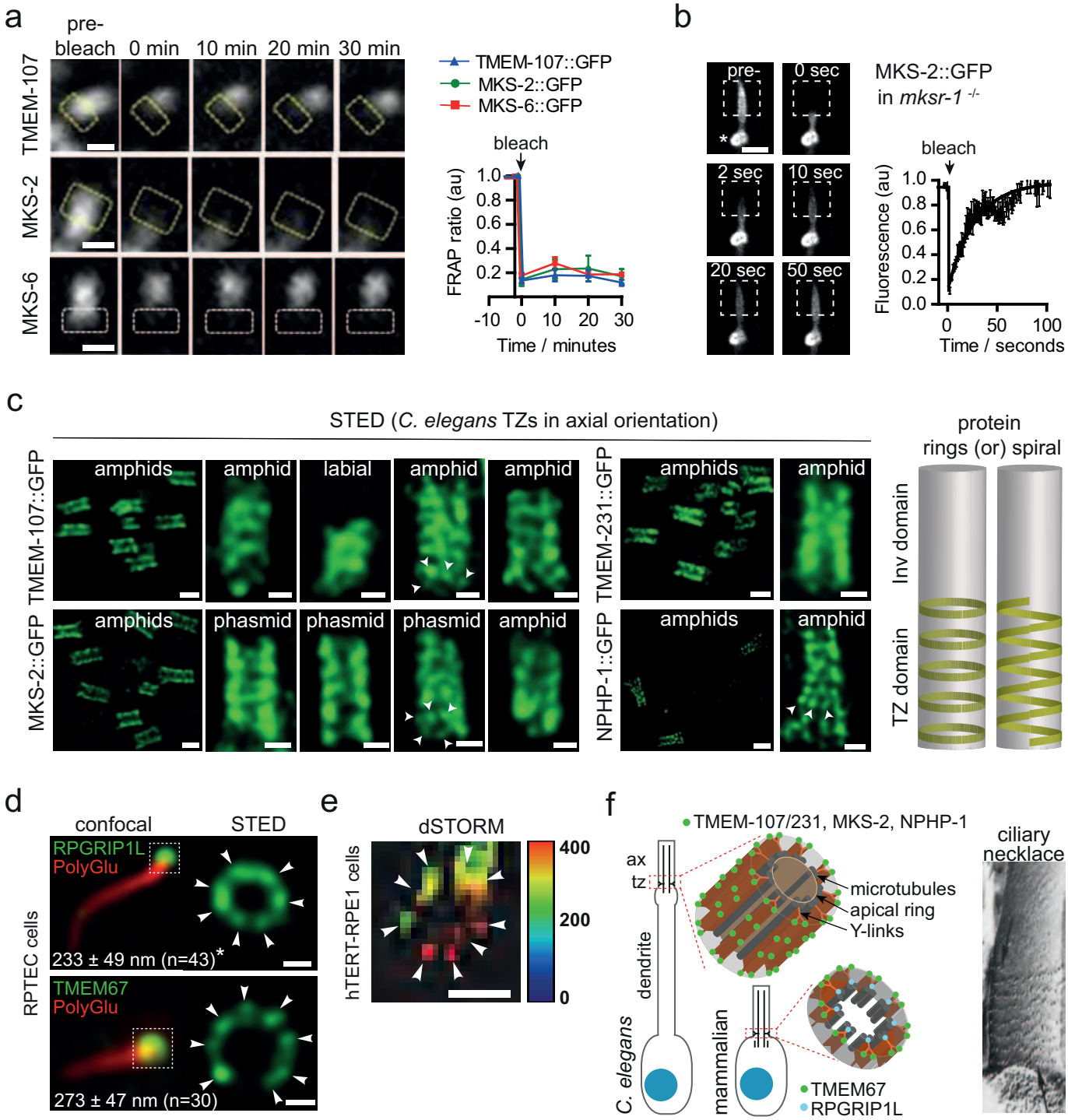


Figure 5-Lambacher *et al.*



ONLINE METHODS

Co-expression analyses

We collected a set of 20 genes known to localise at the TZ[1]. Nucleoporins were excluded because of their additional role in nuclear transport and are therefore likely to have substantially different expression profiles compared to other TZ proteins. The SYSCILIA gold standard (SCGSv1; http://www.syscilia.org/goldstandard.shtml)[21] was used to evaluate the approach for retrieving cilia related genes. A large collection of publicly available microarray data sets was obtained from the NCBI gene expression omnibus (GEO) database: 809 data sets for mouse (~13,000 individual experiments) and 868 data sets for human (~22,000 individual experiments); the full list of the human and mouse GEO (microarray) datasets (GDS) employed are shown in Supplementary Table 2. Each expression data set was assessed for its potential to find novel TZ related genes by determining the expression coherence of the 20 genes known to localise to the TZ. Experiments where known TZ genes show coherent expression obtain a high weighting, and contribute more to the co-expression calculation than experiments with less coherent expression. These weightings are used to calculate an integrated probability for each gene, according to how much its expression profile correlates with that of TZ genes across the expression data sets. We calculated TZ co-expression separately for mouse and human, obtaining scores for ~15,000 mouse genes and ~21,000 human genes.

Sequence analysis and evolutionary analysis

Worm and human MKS and NPHP module protein sequences were extracted from NCBI. Orthologous sequences were detected by BLAST and PSI-BLAST and extracted from an in-house protein sequence and orthology database[34]. Retrieved sequences were manually

verified by reverse BLAST searches. Hidden Markov Models (HMMs) were created for each protein family by first aligning the sequence using MAFFT (version v6.884b, options *--localpair --maxiterate 1000*). Highly fragmented sequences were removed from the alignment. Custom HMMs were created using HMMer 3.0 and searched against our in-house protein sequence database again for sequences missed by PSI-BLAST. Finally we searched the genomic and EST sequences using TBLASTN to find orthologues not found by gene prediction algorithms. For MKS-5 (RPGRIP1L), we also employed sequential sequence searches using NCBI PSI-BLAST using stepwise more diverse orthologous sequences as seed (e.g. we used the *Phytopthora infestans* sequence to find the *Chlamydomonas reinhardtii* sequence and used that sequence to find the *Volvox carterii* ortholog). Discovered sequences were confirmed by reverse PSI-BLAST searches and sequence alignments. All presences and absences were noted for each species considered (**Supplementary Figure 1**). The TMEM107 alignment was made using MAFFT (options *--globalpair --maxiterate 1000*) and edited to fix minor alignment errors. For the differential Dollo parsimony (diff Dollop)[29] analysis, which counts the number of independent losses of proteins along a phylogenetic tree, we used a script provided by Philip Kensche[29]. The network was constructed and visualised in Cytoscape.

Targeted next generation and exome sequencing analysis

Next generation sequencing analyses were performed on 5 µg of DNA sample from each of 198 JBTS and OFD individuals. In 192 JBTS individuals, simultaneous target sequencing was performed using a panel of 25 candidate or causal JBTS genes on an Illumina MiSeq in accordance with the manufacturer's recommendations. In 6 OFD individuals, exome capture was achieved using the SureSelect Human All Exon 50Mb kit (Agilent). The resulting libraries underwent 2×100-bp paired-end sequencing on an Illumina HiSeq 2000 in

accordance with the manufacturer's recommendations. Reads were aligned to the human reference genome (GRCh37/hg19) with the Burrows-Wheeler Aligner (BWA.0.5.6) and potential duplicate paired-end reads were removed using picardtools.1.22 (http://picard.sourceforge.net/). The Genome Analysis Toolkit (GATK) 1.0.57 was used for base quality score recalibration and indel realignment, as well as for single-nucleotide variant and indel discovery and genotyping using standard hard filtering parameters. Homozygous variants with quality scores of >30, sequencing depth of >4, quality/depth ratio of >5.0 and strand bias of <-0.10 were retained for subsequent analyses. Coverage was assessed with the GATK Depth of Coverage tool by ignoring reads with mapping quality of <20 and bases with base quality of <30. Candidate events were inspected using Integrative Genomics Viewer (IGV). Variants were excluded when the frequency was >1/1000 in the Exome Variant Server, NHLBI (http://evs.gs.washington.edu/EVS/). The sequencing data has been deposited in the European Nucleotide Archive (http://www.ebi.ac.uk/ena); accession number PRJEB11176.

Sanger sequencing

TMEM107 (NM_183065) mutation screening was performed by direct sequencing of PCR products (coding exons and adjacent intronic junctions) in 40 additional JBTS individuals, 54 MKS patients (mutation-negative for known genes), and 32 other patients with a clinical diagnosis consistent with a ciliopathy. PCR primers (sequences available upon request) were designed with Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/primer3/). PCR products were purified using the Exo-SAP cleanup kit (USB). Sequencing was performed using the ABI BigDye Terminator Cycle Sequencing kit (v3.1) (Applied Biosystems) following the manufacturer's instructions in an ABI 3130 sequencer 7 (Applied Biosystems). Sequence data were analyzed with SeqScape v2.7 (Applied Biosystems). The impact of the missense

mutation was assessed using PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/). Written consent was obtained from patients for generation and use of their molecular data.

C. elegans strains, maintenance and crossing

C. elegans were maintained and cultured at 20°C using standard techniques. All strains employed shown in Supplementary Table 4. Standard genetic crossing techniques were used to make double mutants and to introduce transgenes into genetic backgrounds. Genotyping was performed using PCR (primer sequences available upon request). The *daf-19(m86)* mutation was followed using the dye-filling assay[35].

Generation of the *oq100* mutation by CRISPR

sgRNA template and Cas9 plasmids were obtained from Addgene (46168, 46169). The *tmem-107*-sgRNA was generated following Friedland *et al.*[25], targeting the sequence ATAGAGATCGAGACGGCGAC. sgRNA and Cas9 constructs were each injected at 125ng/µl, together with the pCeh361 (*dpy-5(+)*) construct at 100ng/µl, into *dpy-5*(*e907*) hermaphrodites. 204 *dpy-5* rescued F1 worms were screened for a *tmem-107*-associated Hpy99I (NEB R0615) restriction fragment length polymorphism (RFLP). 3 independent lines with an RFLP were subjected to Sanger sequencing to identify the *tmem-107* mutations. Two worms had the same in-frame 6bp deletion (P74S75del), predicted to cause no serious effects and were not investigated further. The 3rd worm possessed a 27bp deletion in exon 2 (genomic breakpoints: I:14758954/14758955 - 14758981/14758982) replaced with a 14bp insertion (GTGACAACGTGGAA). This allele designated *oq100* features a frame shift leading to a predicted truncated protein of 72 native amino acids followed by an ectopic 17 amino acids and premature stop.

101 ***C. elegans* reporters**

102 All constructs were generated by fusion PCR as previously described[36]. For the
103 transcriptional (promoter) *tmem-107p*::GFP reporter, GFP amplified from pPD95.67 was
104 fused with 559bp of 5'UTR sequence along with the first 15bp of exon 1 (with the adenine of
105 the start codon mutated to cytosine). For the translational (protein) *tmem-107*::*gfp* reporter,
106 the entire exonic and intronic sequence of each gene, together with upstream promoter
107 sequence (see above), was fused in frame with GFP amplified from pPD95.77. For
108 substituting extramembranous TMEM-107 sequences Polyphobius software
109 (http://phobius.sbc.su.se/poly.html) was employed to determine TMEM-107 residue topology
110 twice: (i) based on *C. elegans* only, and (ii) based on sequence conservation in *C. elegans*
111 close relatives. Subsequently, a consensus prediction was compiled, establishing that aa1-
112 9(N-terminus), 30-47(linker 1), 73-78 (linker 2) and 100-103 (linker 3) are very likely
113 extramembraneous. Next, we searched the literature for tetraspan transmembrane proteins not
114 associated with ciliary TZs and identified with the software Phobius (http://phobius.sbc.su.se)
115 similar sized linker sequences, suitable for replacing TMEM107 linkers. aa 131-148 of SNG-
116 1 to substitute linker 1, aa 74-79 of SPE-38 to replace linker 2, and aa 49-52 to substitute
117 linker 3. In addition, we generated three constructs replacing the TMEM-107 linkers (1-3)
118 with sequences containing scrambled versions of the same residues. The TMEM-107 C-
119 terminal truncation construct TMEM-107(aa1-129)::GFP – called cTMEM107 in Figure 4c -
120 was similarly amplified except that the C-terminal 7 amino acid encoding sequence was not
121 included. The TMEM-107 N-terminal substituted construct (SPH-1(aa1-11)::TMEM-107
122 (10-136)::GFP) – called nTMEM-107 in Figure 4c - was amplified to contain the N-terminal
123 11 amino acids of SPH-1 fused to TMEM-107 lacking its N-terminus. The SPH-1 N-terminus
124 was chosen as replacement because SPH-1 and TMEM-107 are topologically similar; both
125 are predicted tetraspan transmembrane protein with cytosolic N-terminal tails (albeit of

different sequence). Note that full length SPH-1::GFP, SNG-1::GFP and SPE-38::GFP translational reporters (under the control of an *arl-13* promoter active in most ciliated cells) do not localise to ciliary TZs (data not shown). For mimicking the TMEM-107 patient variants, PCR was used to engineer E46G (equivalent to human HsE45G) and F96del (equivalent to HsF106del) mutations into nematode TMEM-107. For simulating the HsL134F*fs**8 mutation, TMEM-107 was truncated at L120, and nucleotides corresponding to the human ectopic residues (FSSPSLG) were added. Transgenic worms expressing above constructs were generated using gonadal transformation via microinjection. The transcriptional construct was injected into *dpy-5(e907)* worms at 50ng/µl, together with the *dpy-5(+)*-containing rescuing construct (at 50 ng/µl), pCeh361. Translational constructs were injected into N2 worms typically at 5ng/µl, together with a coelomycete cell-expressed (*unc-122p*::*gfp* or *unc-122p*::*dsRed*) co-injection marker at 100 ng/µl.

C. elegans immobilization for microscopy

Live worms were immobilized on 10% agarose pads using microbeads (Polysciences #00876-15) or with 40mM tetramisole (Sigma#L9756).

C. elegans fluorescence imaging and FRAP assay

Imaging was performed on an epifluorescence-fitted upright Leica DM5000B compound microscope or an inverted Nikon Eclipse Ti microscope with Yokogawa spinning disk unit (Andor Revolution). Image analysis and formatting was conducted using Image J software (NIH). Fluorescence recovery after photobleaching (FRAP) assays were performed using the above confocal system, with an attached FRAPPA unit. Samples were imaged pre-bleach, and then bleached using a single pulse of the 488 nm laser at 100% with a dwell time of 100 µs. Images were recorded immediately post-bleach (i.e., 0 minute time point) and at varying

time points post bleach until recovery plateaued. All images were taken using the same camera settings (exposure time, gain etc.) and images were subsequently stacked to ensure identical background intensities. Using ImageJ, photobleached and non-photobleached regions were selected and intensity (grey value) measured at each timepoint. After background subtraction, ratios of bleached:non-bleached regions were calculated, and ratios were normalised to a pre-bleach ratio of 1.0.

Dye-filling, chemosensory, and roaming (foraging) assays

Assays for dye filling (DiI), roaming and chemoattraction to a volatile odorant (isoamyl alcohol) were performed as previously described[35]. A chemotaxis index was calculated at 30 minutes and 60 minutes. For the roaming assay, single worms were placed for 16 hours onto seeded plates and track coverage assessed using a grid reference.

Transmission electron microscopy

Young adult worms were fixed, sectioned and imaged using previously reported methodology[35]. For the TEM electron tomography, 200nm thick sections were collected onto formvar carbon coated slot grids and a tilt series acquired from -60 to +60 in 1 degree increments using FEI software. The tilt series was converted to a z stack using serial EM and then visualised using Amira software (FEI).

Cloning of mammalian constructs

Full-length *TMEM107* was cloned into the pcDNA3.0 vector (Invitrogen Inc.), and then shuttled into mCherry-, EGFP-, and FLAG- containing vectors. Mutations were introduced into TMEM107-pEGFP-N3 by QuickChange mutagenesis (Stratagene).

Immortalised mammalian cell lines and antibodies

Mouse inner medullary collecting duct (IMCD3), human retinal pigmented epithelial (hTERT-RPE1) and human embryonic kidney (HEK293) cells were derived from the American Type Culture Collection (ATCC). The genomic status was assessed by array CGH and karyotyping (May 2013) and tested every three months for mycoplasma. Cells were maintained in DMEM/Ham's F12 medium supplemented with 10% foetal calf serum (FCS), under standard conditions (37^{o}C, 5% CO_2) at low passages (<25). Human renal proximal epithelial cells (RPTEC-TERT1) were provided by Tara McMorrow (University College Dublin, Ireland). 3D spheroids (IMCD3 cells) were grown and scored as previously described. Primary antibodies: mouse monoclonal anti-denatured GFP (Sigma-Aldrich cat. #G6539, clone GFP-20), rabbit polyclonal anti-native GFP (A.V. peptide, "Living Colors", Clontech cat# 632377), mouse-anti-γ-tubulin, mouse anti-acetylated-tubulin (Sigma-Aldrich cat. # T7451, clone 6-11B-1), rabbit-anti-γ-tubulin and mouse anti-β actin (Abcam Ltd. cat. # ab6276, clone AC-15). Guinea pig-anti-RPGRIP1L[37], rabbit-anti-TMEM237[15], and rabbit anti-TMEM231[5] have been previously described. Secondary antibodies: Polyclonal Alexa-Fluor 488- Alexa-Fluor 594- and Alexa-Fluor 568- conjugated goat anti-mouse IgG, and goat anti-rabbit IgG (Molecular Probes; cat. # A-11001/A-11005/A-11031/A-110034).

Ciliogenesis assays with human primary fibroblast cells

Patient fibroblasts were stored in the CRB Ferdinand Cabanne Biobank (Dijon, France). Fibroblasts were grown from skin biopsies in DMEM supplemented with 10% FCS and 1% P/S. Cells were incubated at 37°C in 5% CO2 to approximately 90% confluence. Fibroblasts were serum starved for 24 hours prior to fixation with 4% PFA for 5 minutes at room temperature followed by ice cold methanol for 3 minutes and blocked in PBS containing 1% BSA and 0.1% triton X-100 for 30 minutes. Fixed cells were incubated in primary antibodies

diluted in block solution (mouse anti-acetylated tubulin, Sigma T6793, 1:10000, rabbit anti-ARL13b, ProteinTech 17711-1-AP, 1:400) for 90 minutes at room temperature and Alexa Fluor conjugated secondary antibodies from Life Technologies (donkey anti-rabbit 568, 1:400, donkey anti-mouse 647, 1:400) and Hoechst333 for 60 minutes at room temperature. Coverslips were mounted using Fluoromount G. Confocal imaging was performed using a Zeiss LSM700. Approximately 70 events per condition were scored. GraphPad Prism 5.0 was used to perform two-tailed Student's t tests.

Immunofluorescence assays with *Tmem107*-depleted IMCD3 cells

For co-localisation assays (Figure 4d), murine IMCD3 cells were seeded at $2.5x10^5$ cells/well on sterile glass coverslips in six-well plates. Lipofectamine RNAiMAX (LifeTechnologies) was used to transfect cells with the siRNA SMARTpool targeting *Tmem107* (Dharmacon) according to the manufacturer's protocol. After 72 hr, cells were fixed in ice-cold methanol, blocked in 1% non-fat milk and processed for immunofluoresence microscopy using standard methods[38] using AlexaFluor-488 or AlexaFluor-568-conjugated secondary antibodies (1:500, LifeTechnologies). Confocal images were obtained using 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. Co-localization analyses were performed using the FIJI software plug-in "Co-localization Threshold". For knockdown experiments in Figure 2a,b, murine IMCD3 cells were transfected with a siRNA SMARTpool targeting *Tmem107* (Dharmacon), and spheroids grown and scored as previously described[24].

Co-immunoprecipitation assays

For transfection with plasmids, cells at 90% confluency were transfected using Lipofectamine 2000 (Invitrogen Inc.) according to the manufacturer's instructions. Cells were incubated for 24 to 72 hrs prior to lysis or immunostaining. Co-immunoprecipitation was performed as described previously, except that 10mM CHAPS was used as zwitterionic detergent in the lysis buffer[28]. Whole cell extracts (WCE) were prepared from confluent IMCD3 cells transiently transfected with 1.0 µg plasmid constructs in 90mm tissue culture dishes, or scaled down as appropriate. WCE supernatants were processed for immunoprecipitation experiments by using 5 µg affinity-purified mouse anti-GFP ("Living Colors", Clontech Inc.), or 5 µg MAbs, or 5-10 µg purified IgG fractions from rabbit polyclonal antisera, coupled to protein G- and/or protein A-sepharose beads (GE Healthcare UK Ltd.). The affinity-purified rabbit polyclonal anti-TMEM17 and TMEM231 have been described previously[5]. Immunoprecipitations were performed in reduced salt incubation buffer (20 mM Tris, pH7.5, 25 mM NaCl, 2 mM EDTA, 0.5 mM EGTA, 0.02% [w/v] NaN_3, 10% [v/v] glycerol, 10% [v/v] ethanol, 0.1% [v/v] protease inhibitor cocktail) containing 1mM CHAPS. Proteins were analysed by SDS-PAGE (using 4-12% polyacrylamide gradient gels) and western immunoblotting according to standard protocols using either rabbit polyclonal antisera (final dilutions of 1:200-1000) or mAbs (1:1000-5000). Appropriate HRP-conjugated secondary antibodies (Dako Inc.) were used (final dilutions of 1:10000-25000) for detection by the enhanced chemiluminescence "Femto West" western blotting detection system (Pierce Inc.)

Stimulated emission depletion (STED) microscopy

STED imaging was performed on a Leica DMI6000 SP8X CW gated STED system with a 592 nm depletion laser and a HCX PL APO 100x NA1.40 oil objective. Samples were excited at 488nm and emission was detected between 510 and 540nm. Pixel size was < 20 nm

(typically ~15 nm). Images were deconvolved using Huygens Professional software (Scientific Volume Imaging B.V), which is part of the standard configuration of the Leica STED system. For deconvolution a signal to noise ratio of 7 was employed and the 'classic' algorithm selected. All other parameters used were default values (contained in the meta data for each image file), as defined by the Huygen's 'deconvolution wizard.'

Mammalian cell preparation for STED microscopy

RPTEC-TERT1 cells were seeded on glass coverslips and cultured for at least 7 days upon reaching confluence to induce primary cilia formation. Cells were fixed with 3% PFA for 5 minutes at room temperature followed by ice cold methanol for 5 minutes. After blocking with 3% BSA in 0.2% Triton X-100/PBS for 30 min, cells were stained for TZ proteins (guinea pig anti-RPGRIP1L, 1:500, or rabbit anti-TMEM67, 1:200) and the ciliary axoneme (mouse anti–polyglutamylated tubulin, 1:1000; Adipogen; cat. # AG-20B-0020-C100; clone GT335) at room temperature for 2 h. Primary antibodies were detected with polyclonal Alexa-Fluor conjugated secondary antibodies (1:100 dilution; 1hr at room temperature): goat anti-guinea pig 488 (Molecular Probes; cat. # A11073), goat anti-rabbit 488 (Molecular Probes; cat. # A11008) or goat anti-mouse 568. All antibodies were diluted in blocking solution. Coverslips were mounted on glass slides in ProLong Diamond (Life Technologies).

Direct stochastic optical reconstruction microscopy (STORM) system

The dSTORM system was based on the 3D PALM system of York et al.[39]. We used an inverted microscope (Olympus, IX81 with additional side-port ILL100-TIRZD) fitted with an automated *x-y* stage with additional piezoelectric adjustment in *z* (Applied Scientific Instrumentation, PZ-2000). The objective lens was a 60x, 1.2 NA, water immersion lens (Olympus, UPLSAPO60XW). The system included a focus locking device that reduced *z*-

drift during calibration and data acquisition (Mad City Labs, C-focus) and a cylindrical lens (Thorlabs, f = 100 mm, LJ1567RM-A). Images were captured by a back-thinned, electron-multiplying CCD camera, cooled to −80°C (Andor Technology, iXON Ultra, model DU-897U-CSO-#BV), using published scripts called from the camera interface (Andor Technology, SOLIS). This also converted SIF images generated by the camera to DAT format for processing.

Preparation of secondary antibody and samples for dSTORM

Affinity-purified guinea pig anti-RPGRIP1L has been described previously[37], and was used at 1:200 dilution. 100 µg polyclonal donkey anti-guinea pig IgG (H+L) secondary antibody was labelled ("AffiniPure", Jackson ImmunoResearch Europe cat. #706-005-148) in PBS containing 120 mM $NaHCO_3$ with 2 µg carboxylic acid succinimidyl ester of the photoswitchable dye AlexaFluor647 (A37573, Life Technologies Inc.) Labelling was for 30 min. at room temperature, and unincorporated dye was removed by gel filtration through NAP-5 columns (17-0853-02, GE Healthcare) according to the manufacturer's protocol. Antibody:dye labelling ratios of approximately 1:1 were confirmed by measured absorbances in a spectrophotometer, with a final concentration of 0.3 µg/µl. Secondary antibodies were used at titres of 1:100.

Human hTERT-RPE1 cells were seeded on cleaned coverslips (#1.5, 25-mm diameter CS-25R15; Warner Instruments) at $2.5x10^5$ cells per well on the coverslips in six-well plates and serum starved in normal media with 0.2% FCS for 48 hours to induce ciliogenesis. Coverslips were processed for immunofluorescence staining using standard methods, and post-fixed with 3% para-formaldehyde/0.1% gluteraldehyde in PBS. Coverslips were then incubated with 0.01% poly-L-lysine (Sigma-Aldrich, P4707) for 10 min followed by a suspension of 100 nm

gold nanoparticles (1:10, Sigma-Aldrich 724031, in PBS) for use as fiducials. Calibration series for depth information (see below) were taken at this point, using the relevant excitation laser(s). Data was acquired in the presence of fluorescence quenching buffer consisting of 10 mM Tris-HCl (pH 8.0), 50 mM NaCl, 5 mg/ml glucose, 114 mM β-mercaptoethanol, 0.5 mg/ml glucose oxidase and 40 µg/ml catalase.

dSTORM image acquisition, processing and analysis

Imaging software employed was described previously[39] (see https://github.com/AndrewGYork/palm3d for further details). Calibration images were taken to calculate the point spread function (PSF) in steps of 50nm over a 4µm range, using fluorescence of selected fiducials. Labels were then excited until emission was quenched and then stochastically re-activated using a 405 nm laser initially at low power (0.4 mW), followed by data collection. Fluorescence emission events were localized in *x*-*y* and *z* using the PSFs captured in the calibration series[39]. Drift was recorded and corrected using images of the fiducials. Emission events were binned into a histogram for display and to correct for distortion by the cylindrical lens. Binning was at either 20 nm or 10 nm, with *x*-*y*-*z* localisation precisions of 20 x 30 x 50 nm for fiducial markers[39]. The software applied smoothing to reflect the limits of localisation precision, and *z*-stacks were displayed in FIJI.

Statement of image representation and reproducibility

Representative images are shown for all worm and mammalian cell imaging. The following states how many samples were imaged in each figure panel where a representative image is shown: Fig. 1c (>30 worms), Fig. 1d (30 cells), Fig. 2b (50 spheroids), Fig. 2h (75 cells), Fig. 3b (>20 worms for each strain), Fig. 3c (>50 worms for each strain), Fig. 3d (>20 worms for each strain), Fig. 3f (>30 worms for each strain), Fig. 3g (>50 worms for each strain), Fig. 3h

(>30 worms for each strain), Fig. 4a (>40 worms for each strain), Fig. 4c (>20 worms for each strain), Fig. 4d (50 cells for each experiment), Fig. 5a (>20 worms for each strain), Fig. 5b (10 worms), Fig. 5c (>100 worms imaged per strain), 5d (>30 cells per experiment), Fig. 5e (>20 TZs for each marker), Suppl. Fig. 2a (>20 worms), Suppl. Fig. 2b (>40 worms), Suppl. Fig. 2d (>20 worms), Suppl. Fig. 2e (>30 cells), Suppl. Fig. 4a (2-4 amphid pores), Suppl. Fig. 4b (>20 transition zones), Suppl. Fig. 4c (>30 worms per strain), Suppl. Fig. 5a (>10 sets of amphid transition zones).

35. Sanders, A.A., Kennedy, J. & Blacque, O.E. Image analysis of Caenorhabditis elegans ciliary transition zone structure, ultrastructure, molecular composition, and function. *Methods Cell Biol* **127**, 323-347 (2015).

36. Hobert, O. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. *BioTechniques* **32**, 728-730 (2002).

37. Arts, H.H. *et al.* Mutations in the gene encoding the basal body protein RPGRIP1L, a nephrocystin-4 interactor, cause Joubert syndrome. *Nature genetics* **39**, 882-888 (2007).

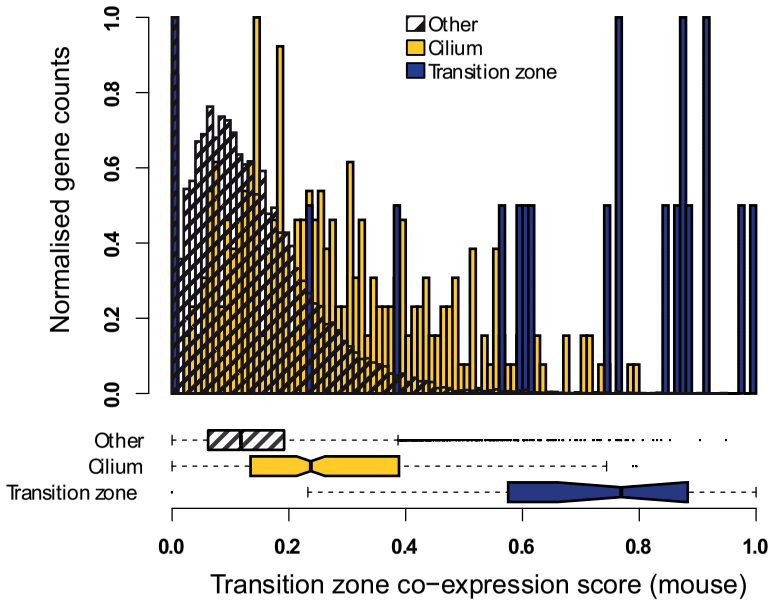
38. Dawe, H.R. *et al.* Nesprin-2 interacts with meckelin and mediates ciliogenesis via remodelling of the actin cytoskeleton. *Journal of cell science* **122**, 2716-2726 (2009).

39. York, A.G., Ghitani, A., Vaziri, A., Davidson, M.W. & Shroff, H. Confined activation and subdiffractive localization enables whole-cell PALM with genetically expressed probes. *Nat Methods* **8**, 327-333 (2011).

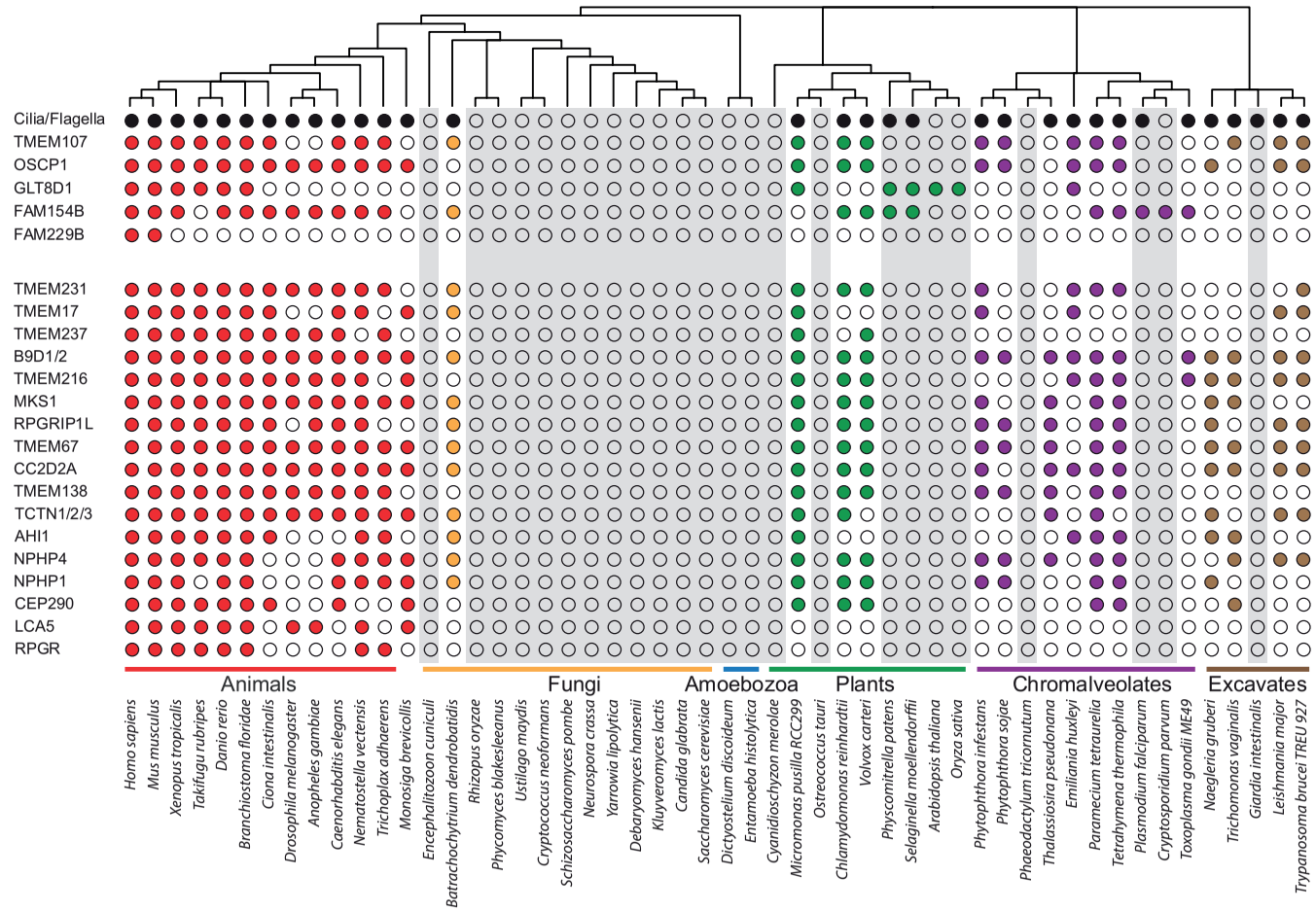
40. Darmency-Stamboul, V. *et al.* Detailed clinical, genetic and neuroimaging characterization of OFD VI syndrome. *European journal of medical genetics* **56**, 301-308 (2013).

Supplementary Figure 1-Lambacher *et al.*

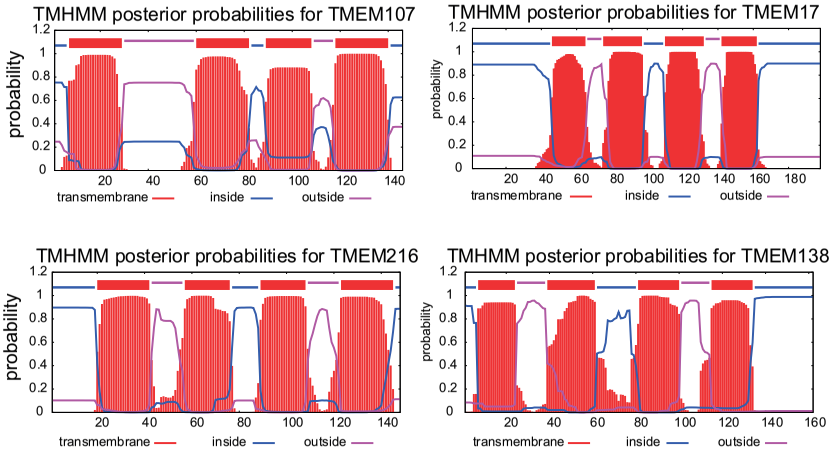
a



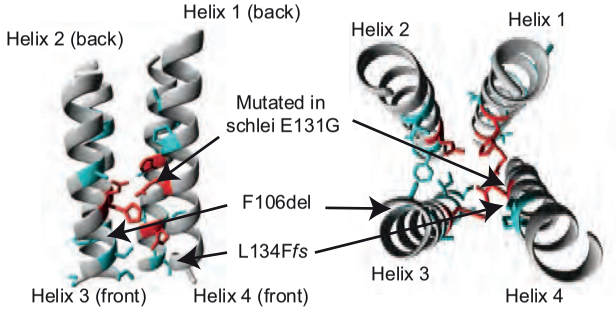
b



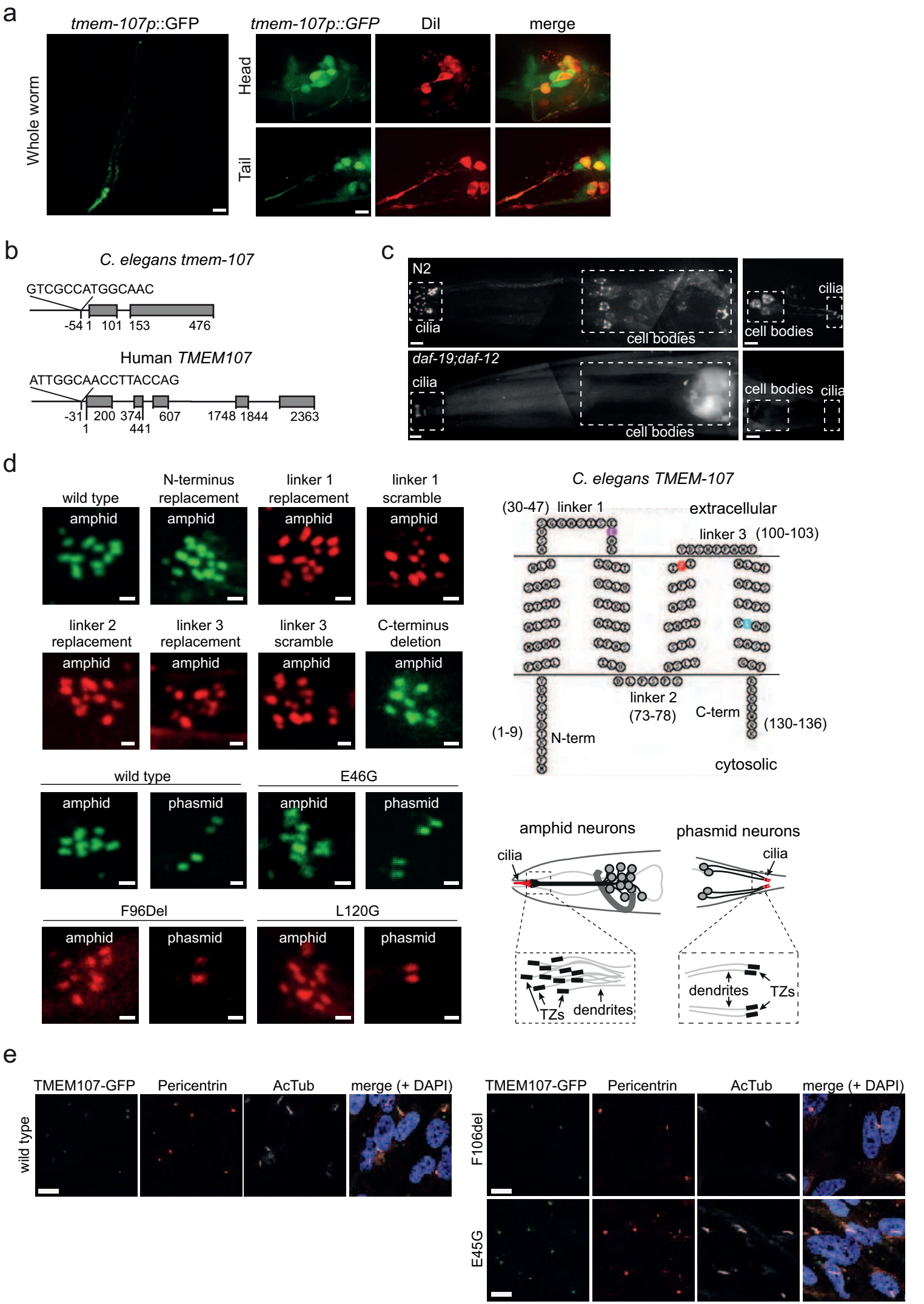
c



d

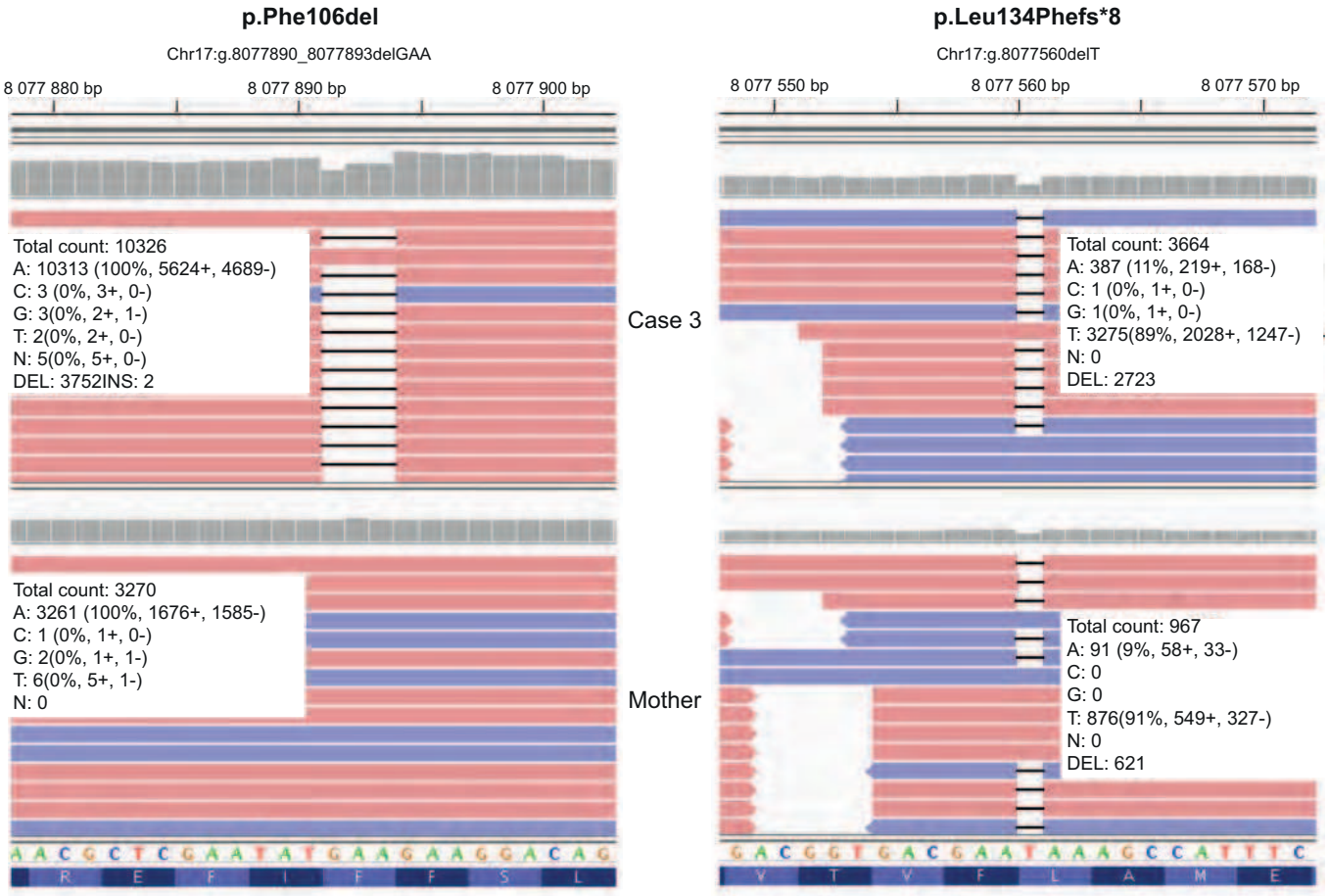


Supplementary Figure 2-Lambacher *et al.*

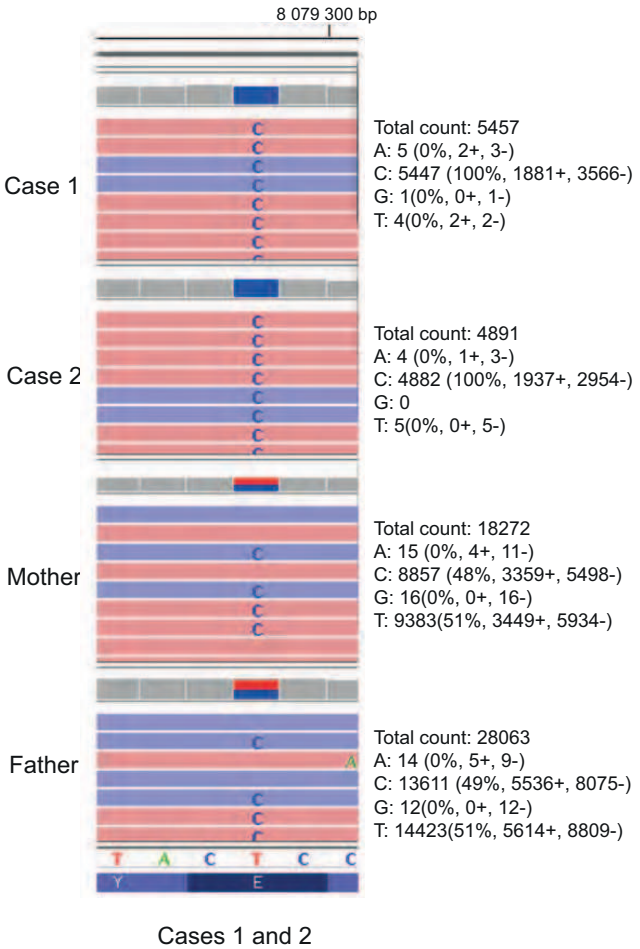


Supplementary Figure 3-Lambacher *et al.*

a



b

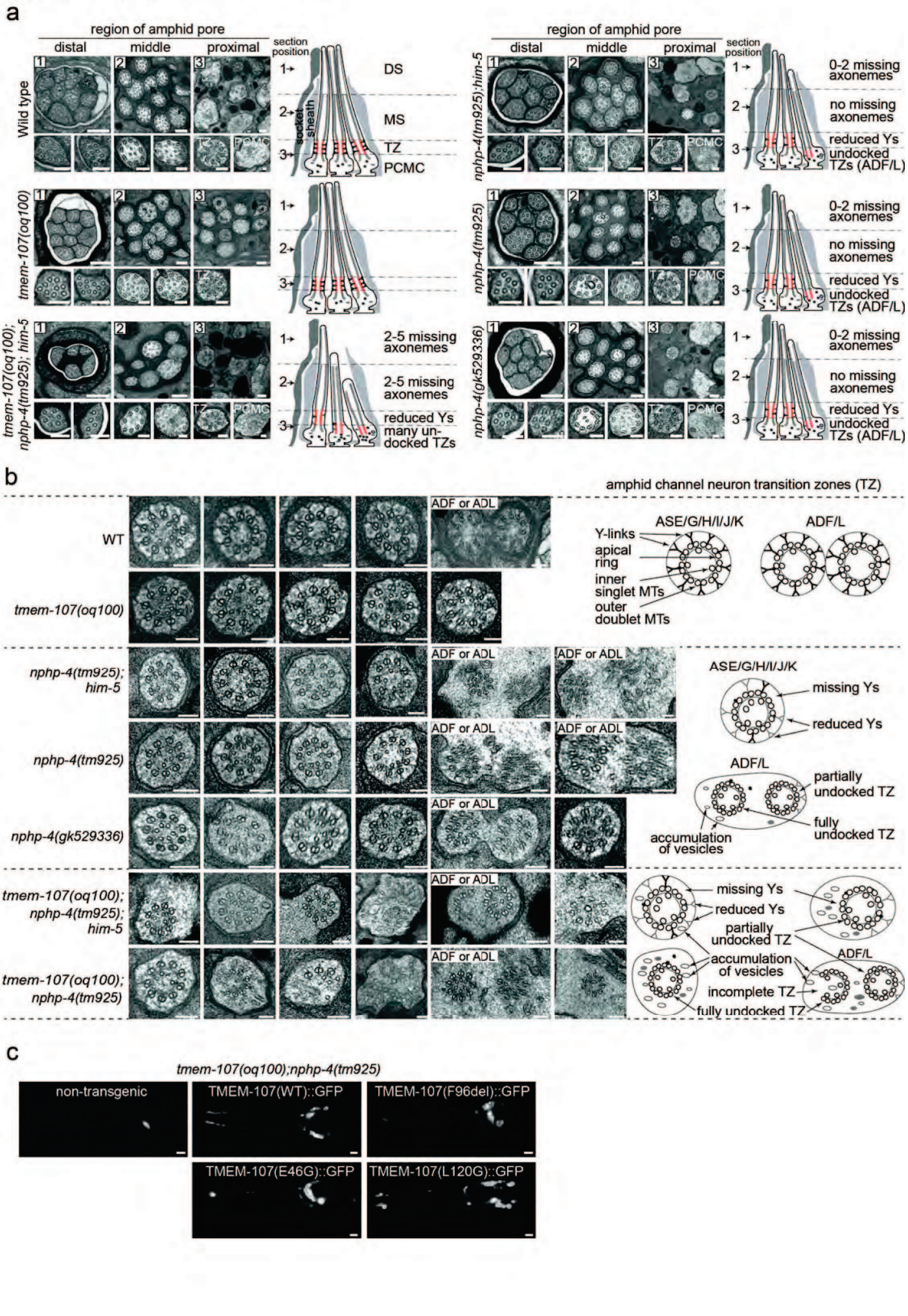


c

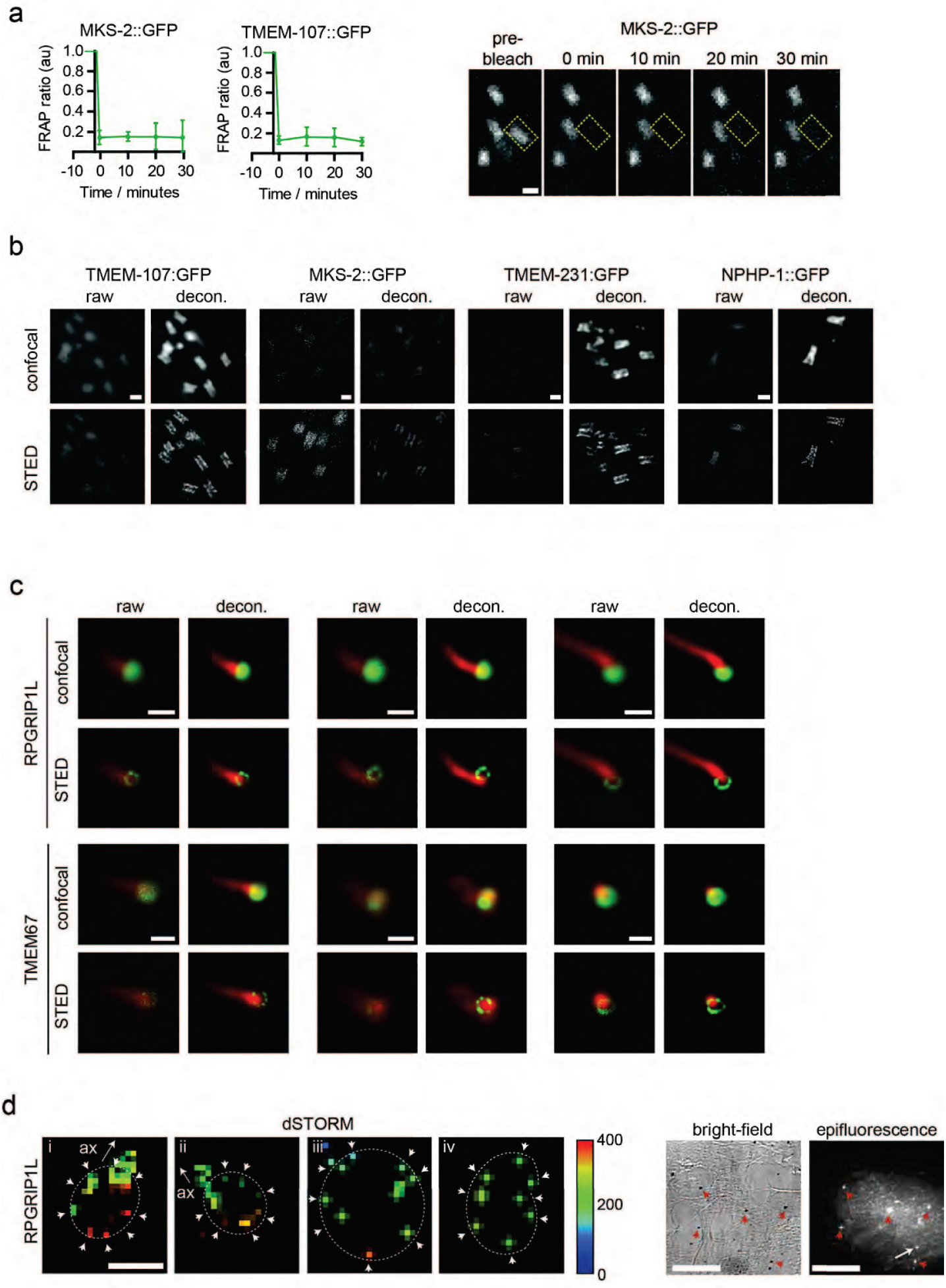
Cases	**1***	**2***	**3**
Sexe	F	F	M
Age at last follow-up	9y	9y	22y
Origin	Turkey	Turkey	Caribbean
Consanguinity	+	+	-
Oral abnormalities			
Abnormal frenulae	+	+	-
Lobulated tongue	-	-	-
Lingual harmatomas	+	+	-
Cleft palate	-	-	-
Tooth abnormalities	-	-	-
Facial dysmorphism			
Hypertelorism	+	+	-
Low-set ears	+	+	-
Cleft lip	-	-	-
Pseudocleft of the upper lip	-	-	-
Micro/retroagnathia	-	-	-
Hand abnormalities			
Brachydactyly	-	-	-
Syndactyly	-	-	-
Polydactyly	Post axial	Post axial	-
Foot abnormalities			
Brachydactyly	-	-	-
Syndactyly	-	-	-
Polydactyly	Post axial	Post axial	-
Neurological features			
Ataxia	+	+	+
Oculomotor apraxia	+	+	+
Psychomotor delay / ID	+	+	+
Brain abnormalities			
Cerebellar hypoplasia	+	+	+
Molar Tooth Sign	+	+	+
Heterotopiae (periventricular, subcortical)	+	+	-
Polymicrogyria	-	+	-
Hypothalamic hamartoma	-	-	-
Other features			
Ventilatory disorders (apnea, hyperpnea)	+	+	-
Renal malformations	NA	NA	-
Retinopathy/ Pathologic ERG	+	+	+
Cardiac malformations	-	+	-
Deafness	-	-	-
Liver involvement	-	-	+
Diagnosis	OFDVI	OFDVI	JBS

NA: Not Available; F: female ; ERG: electroretinogramm ; ID : Intellectual disability ; JBS : Joubert syndrome ; M: Male; y: years; OFDVI : oral-facial-digital syndrome type VI. * cases 3 and 4 from Darmency-Stamboul et al., 2013.

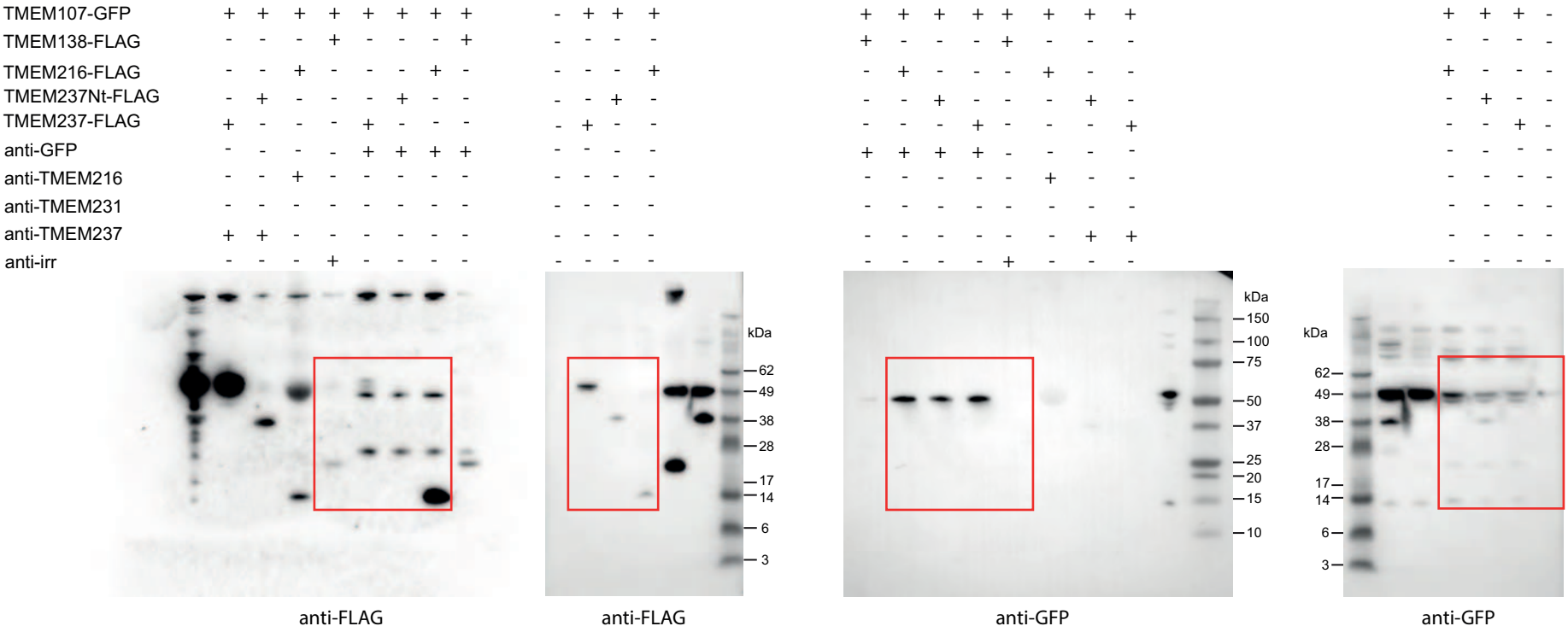
Supplementary Figure 4-Lambacher *et al.*

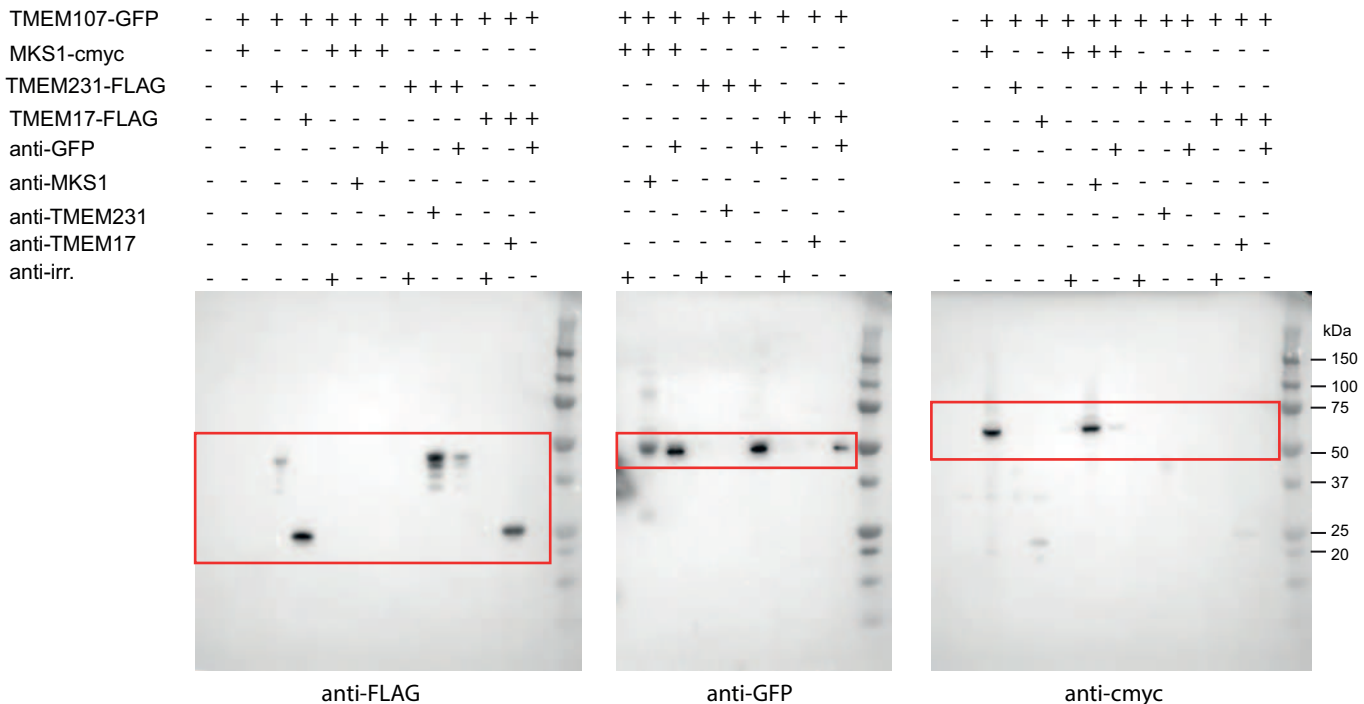


Supplementary Figure 5-Lambacher *et al.*



Lambacher *et al.* Supplementary Figure 6





SUPPLEMENTARY TABLE AND FIGURE LEGENDS

Supplementary Table 1. Training set of TZ genes used for co-expression screening.

Supplementary Table 2. Lists of all human and mouse NCBI GEO microarray datasets used for co-expression screening. GSE reference accessions and GEO dataset (GDS) identifiers are provided for each dataset as well as the number of samples per dataset.

Supplementary Table 3. Human and mouse genes co-expressed with the TZ gene training set (Supplementary Table 1). Sheet 1: the top 500 best scoring genes, based on the average of the human and mouse co-expression scores (ranks), which serves as a resource for finding new cilium genes. Also shown is published information of protein localisations, ciliary function and disease, and presence or absence from the SysCilia gold standard[21]. Sheet 2: top 100 best scoring human co-expression hits. Sheet 3: top 100 best scoring mouse co-expression hits. Sheet 4: list of genes common to the top 100 human and mouse genes (sheets 2, 3).

Supplementary Table 4. *C. elegans* strains used in this study.

Supplementary Figure 1. Phylogenetic and bioinformatics screening data of candidate TZ genes. (a) Frequency histogram of binned mouse gene co-expression scores, derived from weighted analyses of gene expression datasets using a training set of 20 known TZ genes (Supplementary Table 1). This graph is the equivalent of the human gene co-expression dataset presented in Figure 1a. Frequencies normalised to compare different distributions. Grey hatched; all human genes, yellow; ciliary genes in the SysCilia gold standard[21], blue;

TZ gene training set. Box-plots display median and quartiles for histogram distributions. **(b)** Presence and absence of candidate and known TZ genes in 52 eukaryotic species. The presence of orthologues for the 20 TZ training set genes and the five candidate TZ genes were determined by bi-directional best hits using BLAST and PSI-BLAST, as well as custom built hidden Markov models, HHPred, and intermediate sequence searches using PSI-BLAST and TBLASTN. Species are ordered according to their phylogenetic relationship as shown by the phylogenetic tree at the top. The top row indicates which species possess cilia or flagella. Grey columns indicate species lacking a (canonical) TZ. Ciliated species that have lost MKS genes appear to lack well defined Y-shaped linkers[22]. **(c)** Model of the four transmembrane helix topology of human TMEM107. Predicted transmembrane regions for TMEM107 and three known TZ proteins (TMEM216, TMEM138, and TMEM17) using TMHMM2.0 (http://www.cbs.dtu.dk/services/TMHMM). Alignment of TMEM107 sequences to the homologous TMEM216, TMEM138, and TMEM17 suggests TMEM107 is homologous to these three TZ proteins (not shown). **(d)** To model the transmembrane helices we used a standard existing helix obtained from the PDB. We swapped the amino acid side chains one by one using YASARA. The transmembrane topology of TMEM107 was predicted with TMHMM2.0. Helices are ordered anti-clockwise, starting with helix 1 in the right-rear, (bottom to top), helix 2 at the left-rear (top to bottom), helix 3 at the front-left (bottom to top) and helix 4 at the front-right (top to bottom). On the right side the four helices are depicted from a downwards viewpoint. The evolutionary conserved, charged residues (in red, a histidine and an arginine in helix 1, a glutamate in helix 2, a histidine in helix 3 and a glutamate in helix 4) are at the same height in the four helices, suggesting interactions, and therefore a four helix bundle model of the protein's transmembrane structure. The conserved non-charged residues are in cyan. The mouse *Schlei* (E125G) mutation (the human equivalent is E131G)[23] and the human F106Del and L134F*fs* mutations found in this study are indicated

by arrows. E45G lies within the extracellular loop between helix 1 and helix 2 and is not depicted here.

Supplementary Figure 2. Expression and localisation analyses of *C. elegans* and human TMEM107 constructs (wild type and variants) (a) *C. elegans tmem-107* is expressed exclusively in ciliated sensory neurons. Shown are fluorescence images of worms expressing a transcriptional *tmem-107p*::GFP reporter (P). DiI costain identifies a subset of ciliated neurons, namely 6 amphid cells (ADL, ASH, ASJ, ASK, AWB and ASI (not shown)) and both phasmid cells (PHA/B). Bars; 25 μm (large whole worm panels), 6 μm (small head and tail panels). den; dendrites, cil; cilia. **(b)** Schematics showing candidate X-box sequences in the promoters of human and nematode *TMEM107*. **(c)** DAF-19 RFX transcription factor is required for TMEM-107::GFP expression in *C. elegans*. Shown are head (left panels) and tail (right panels) regions of N2 (wild-type) and *daf-19(m86);daf-12(sa204)* double mutant worms expressing a translational *tmem-107::gfp* transgene (see Figure 1c). Bars; 6 μm. **(d)** Analysis of TMEM-107 (wild type and variants) localisation in *C. elegans*. Shown are fluorescence images of the amphid and phasmid TZ regions (see also bottom schematic) in worms expressing various GFP tagged (C-terminus) TMEM-107 proteins. Top schematic shows the predicted topology of the tetraspan TMEM-107 *C. elegans* protein and indicates the disrupted domains and sequences. Linker 1 replacement sequence taken from SNG-1, and linker 2 and 3 replacement sequences taken from SPE-38 (see methods section for further details). The coloured residues denote amino acids mutated in the *TMEM107* patients (see methods section for descriptions). TZ; transition zone. Bars; 1μm. **(e)** Analysis of human TMEM107 patient variant protein localisation. Images of hTERT-RPE1 cells expressing GFP-tagged human TMEM107(E45G) or TMEM107(F106del), costained with antibodies for ciliary axonemes (acetylated tubulin; AcTub) and basal bodies (pericentrin). Bars; 10 μm.

Supplementary Figure 3. Sequencing details for the three cases of mutated *TMEM107* and clinical details of *TMEM107* patient phenotypes. Integrative genomics viewer data showing: **(a)** compound heterozygous *TMEM107* mutations in case 3 consisting of one frameshift deletion (NM_032354.3: g.8077560delT; p.Leu134Phefs*8) and one in-frame deletion (NM_032354.3: g.8077890_8077893delGAA; p.Phe106del), and **(b)** homozygous *TMEM107* missense variant (NM_183065: g.8079298T>C; p.Glu45Gly) in cases 1 and 2. Clinical details of the three *TMEM107*-mutated cases are presented in **(c)**, leading to OFDVI and JBTS diagnoses. Cases 1 and 2 had previously been reported[40].

Supplementary Figure 4. Effect of *tmem-107* mutations on cilium ultrastructure and function. (a) Cilium ultrastructure is highly disrupted in *tmem-107;nphp-4* double mutants. Low (large panels) and high (small panels) magnification TEM images of cilia from serial cross sections taken from the distal (1), middle (2) and proximal (3) regions of the amphid pore (position of section in pore denoted by numbers in schematic). Wild-type pores consist of 10 ciliary axonemes (only three shown in schematics for simplicity), each consisting of a distal segment (DS; singlet A microtubules), a middle segment (MS; doublet A/B microtubules), a transition zone (TZ; with membrane-microtubule connecting Y-links) and a periciliary membrane compartment (PCMC). In *tmem-107(oq100);nphp-4(tm925)* double mutants (also harbours the *him-5(e1490)* mutation linked to *nphp-4*), multiple axonemes are missing in the middle and distal pore regions, TZ Y-links (Y's) are reduced or missing, and vesicles frequently accumulate in the TZ and PCMC regions. Also, the majority of double mutant TZs are partially or fully disconnected (undocked) from the ciliary membrane, extending from ectopic anterior positions within the PCMC. In contrast, most or all of the ciliary axonemes are present in *tmem-107(oq100)* and *nphp-4* single mutants. However,

nphp-4 worms carrying the *tm925* deletion (with or without *him-5(e1490)*) or the *gk529336* nonsense mutation show consistent defects in Y-link integrity and TZs are undocked in two neurons (ADF, ADL). Images are representative of at least 4 analysed amphid pores for all strains except *nphp-4(tm925) a*nd *nphp-4(gk529336)* where 2 pores were analysed. Bars; 200 nm (low magnification images), 100 nm (high magnification images). **(b)** Compendium of TZ images and associated schematics showing the TZ defects outlined above in (a). Bars; 100 nm. **(c)** Dye filling assay (DiI) of *tmem-107(oq100);nphp-4(tm925)* worms transgenically expressing various GFP-tagged TMEM-107 constructs (wild type, E46G, F96del, L120G). Shown are fluorescence images of the head region. Non-transgenic worms are strongly dye-filling defective, whereas dye filling is restored in worms expressing TMEM-107 constructs (wild type or mutant versions). Bars; 10 µm.

Supplementary Figure 5. Supplementary FRAP and super resolution imaging data. (a) FRAP curve and representative time lapse images following quenching of 100% of MKS-2::GFP and TMEM-107::GFP signals at the TZ (boxed region shows the bleached TZ of an amphid channel cilium). Data points represented as mean ± S.D. n=3 (MKS-2::GFP) or 4 (TMEM-107::GFP) independent experiments. Bar; 500 nm. **(b)** Raw and deconvolved (decon.) STED and confocal images of *C. elegans* MKS and NPHP module proteins (GFP-tagged). Bars; 500 nm. **(c)** Raw and deconvolved (decon.) STED and confocal images of renal RPTEC cells stained for polyglutamylated tubulin (ciliary axonemes; red; confocal only) and either endogenous human RPGRIP1L or TMEM67 (green; confocal and STED). STED imaging reveals that RPGRIP1L and TMEM67 form clusters of discrete signals arranged as a hollow ring at the TZ. Bars; 500 nm. **(d)** Super-resolution dSTORM microscopy of RPGRIP1L in the ciliary transition zone of human hTERT-RPE1 cells. The loose, tilted ring TZ organisation of RPGRIP1L shown in Figure 5e (i) and examples from

additional cells (ii-iv). dSTORM image reconstruction used 10 nm histogram bins, Gaussian image smoothing in the *palm3d* reconstruction output and contrast enhancement in FIJI. Dashed circles and ovals circumscribe TZ localisations which form the identified hollow loose ring structure, with discrete clusters of protein denoted by white arrowheads. Localisation density at individual points on the ring varied between samples, and was highest in (i). The distribution of signals in iv deviates significantly from an oval and could represent a partial spiral or helical arrangement. In some images (I, ii), some signal appears to enter the ciliary axoneme (ax) distal to the TZ. Images depth-coded by colour, with the *z* axis scale bar in nm indicated on the right. Representative bright-field and epifluoresence images from cells stained for RPGRIP1L and TMEM67, with the transition zone acquired and reconstructed in (iv) indicated by the white arrow. Red arrowheads indicate fiducials. Scale bars; 100 nm (dSTORM images; all images identically scaled), 10 μm (bright-field and epifluorescence images).

Supplementary Figure 6. Uncropped scans of western blots shown in Figure 4e. Red boxes denote the cropped regions shown in Figure 4e.

Supplementary Video 1. Electron Tomogram of the *C. elegans* TZ. Reconstruction derived from a 200 nm section of a *C. elegans* amphid channel ciliary TZ. Arrow denotes a Y-link density throughout the tomogram, indicating that the Y-link structures are continuous sheets along the axial plane. Bar; 100 nm.