Structure of the polycystic kidney disease

TRP channel Polycystin-2 (PC2)

Mariana Grieben[1,5], Ashley CW Pike[1,5], Chitra Shintre[1,5], Elisa Venturi[2], Sam El-Ajouz[2], Annamaria Tessitore[1], Leela Shrestha[1], Shubhashish Mukhopadhyay[1], Pravin Mahajan[1,4], Rod Chalk[1], Nicola A Burgess-Brown[1], Rebecca Sitsapesan[2], Juha T Huiskonen[3], Elisabeth P Carpenter[1].

[1]Structural Genomics Consortium, University of Oxford, Oxford, UK.

[2]Department of Pharmacology, University of Oxford, UK.

[3]Oxford Particle Imaging Centre, Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

[4]Present address: Astex Pharmaceuticals, Cambridge, UK.

[5]These authors contributed equally to this work

Correspondence should be addressed to J.T.H (juha@strubi.ox.ac.uk) and E.P.C (liz.carpenter@sgc.ox.ac.uk)

Abstract

Mutations in either polycystin-1 (PC1/PKD1) or polycystin-2 (PC2/PKD2/TRPP1) cause autosomal dominant polycystic kidney disease (ADPKD) by unknown mechanisms. Here we present the structure of human PC2 in a closed conformation, solved by electron cryo-microscopy at 4.2 Å resolution. The structure reveals a novel polycystin-specific "Tetragonal Opening for Polycystins" (TOP) domain, which is tightly bound to the top of a classic transient receptor potential (TRP) channel structure. The TOP domain is formed from two extensions to the voltage sensor-like domain (VSLD); it covers the channel's ER lumen/extracellular surface and encloses an upper vestibule, above the pore filter, without blocking the ion conduction pathway. The TOP domain fold is conserved among the polycystins, including the homologous channel-like region of PC1, and is the site of a cluster of ADPKD-associated missense variants. Extensive contacts between the TOP domain subunits, the pore and the VSLD provide ample scope for regulation by physical and chemical stimuli.

Introduction

Autosomal dominant polycystic kidney disease (ADPKD) is the most common monogenic genetic disease, affecting at least 1:800 live births[1]. Progressive formation and enlargement of fluid-filled cysts cause loss of kidney function and end stage renal disease in half of patients over 60 years old[2]. Cysts also form in the pancreas and liver, and there is an increased risk of cardiovascular dysfunction[3], hypertension[4], cerebral aneurysms and inflammation[5]. 15% of cases of ADPKD are caused by mutations in the *PKD2* gene, which encodes the polycystin-2 (PC2) protein[6] and 85% by mutations in the related *PKD1* gene, which encodes the polycystin-1 (PC1) protein[7]. ADPKD is thought to be a ciliopathy, a disease associated with changes in cilia function. PC2 is also required for the correct establishment of left-right asymmetry in response to cilia-driven leftward flow of fluid in the embryonic node during embryogenesis. Left-right patterning defects are seen in *Pkd2*[-/-] mouse embryos[8,9] and in some patients with ADPKD[10]. PC1 and PC2 are found in cilia[11,12], but the role they play in this organelle is debated. A PC1–PC2 complex could act as a mechanosensitive cation channel in cilia[13], providing Ca^{2+} influx when cilia are bent[14], although this view has been challenged recently[15,16]. PC2 also acts as a calcium regulated cation channel in the endoplasmic reticulum (ER)[17], and is found in the plasma membrane[18] and lateral cell-cell junctions, where it helps to mediate cell-cell contacts, cellular polarity, cell division and apoptosis.

PC2 is a transient receptor potential (TRP) non-selective, Ca^{2+}-permeable cation channel[19] that can be regulated by diverse stimuli including divalent cations[17,20], pH[21], voltage[21] and phosphorylation[22]. PC1 has a channel-like domain that is homologous to PC2[6] (Supplementary Fig. 1). PC1 and PC2 form a complex, possibly by association of their C-terminal coiled-coil domains, and this complex facilitates trafficking to cilia[23]. PC2 is also thought to form complexes with several other channels and cytoskeletal proteins[22]. Mutations in the *PKD1* and *PKD2* genes in patients with ADPKD occur throughout both genes and range from deletions, insertions, frameshifts and splice variants, to nonsense mutations and missense variants (http://pkdb.mayo.edu/). Although ADPKD inheritance is dominant, ADPKD is believed to be caused by loss of function mutations[24], with a "second hit", such as mutation of the second allele, or tissue damage initiating cyst formation[25,26]. A total of 27 missense variants have been identified in the channel region in patients with ADPKD, but the mechanisms by which these variants impact PC2 function remain unclear.

Results

PC2 structure determination

To understand the structural basis for ADPKD, we determined the structure of the PC2 channel and the associated domain that resides in the ER lumen or cell surface, using electron cryo-microscopy (cryo-EM). We selected a truncated PC2 construct (spanning residues Pro185 to Asp723) after extensive optimisation for crystallisation, that had given crystallographic data to 4.2 Å resolution (Online methods). This construct, which comprises the S1-S6 transmembrane helices (TMHs) and the 'extracellular loop' regions between the voltage sensor helices, was used for cryo-EM data collection (Online methods and Table 1). We reconstructed the PC2 tetramer from 19,546 particles, selected from an initial dataset of 150,882 particle images, recorded on a direct electron detector. From these particles we obtained a map of PC2 by gold-standard refinement with an overall resolution of 4.2 Å, as estimated by Fourier shell correlation. At this resolution, we were able to build an atomic model for the S1-S6 region and the ER/extracellular domain *de novo*, without reference to other TRP channel structures (Online methods, Fig. 1, Supplementary Figs. 2 and 3). We confirmed cysteine positions in the TM region using the location of anomalous difference peaks in a map calculated from an 8.5Å mercury derivative X-ray dataset (Supplementary Fig. 3g; online methods). The PC2$^{P185-D723}$ construct behaved as a voltage-sensitive, cation-selective ion channel, exhibiting complex sub-conductance gating behavior when reconstituted into artificial membranes under voltage-clamp conditions (see Methods, Fig. 1d, e, and Supplementary Fig. 4 for further details). It also retained functional properties such as sensitivity to block of K$^+$ current by Ca^{2+} and amiloride that have been observed in the full length channel [20,27,28] (Supplementary Fig. 4h-j). When control liposomes without PC2$^{P185-D723}$ were incorporated into the bilayer, we observed no single channel-like events (n=7; Supplementary Fig. 4g). Taken together, these functional data confirmed that the truncated PC2$^{P185-D723}$ construct retains many of the characteristics reported for the full length channel.

Overall architecture of the PC2 structure

As anticipated, PC2 has a classic TRP channel fold with a polycystin-specific domain locked onto the ER lumen or extracellular surface of the channel (Fig. 1). The first four TMH (S1-S4) form a voltage sensor-like domain (VSLD), followed by a pore domain with two TMH (S5-S6) interspersed by a pore helix (PH1), the selectivity filter and second short pore helix (PH2). The VSLD interacts with the pore domain of the adjacent chain in the tetramer. This architecture is common to many voltage-gated ion channels (VGICs)[29] and TRP channels[30,31]

(Supplementary Fig. 5). The PC2 VSLD forms a classic VGIC four helix bundle but, like other TRP channels, it lacks the four arginine residues (R1-R4) on S4 (Supplementary Fig. 5e, f) that are responsible for voltage sensing in classic VGICs[29]. Two conserved lysine residues (Lys572 and Lys575) at the C-terminal end of S4, the first of which occupies the R4 position, may contribute to the PC2 voltage sensitivity[21] (Fig. 1e and Supplementary Fig. 4).

The TOP domain

PC2 is unusual among the TRP channels in that it has two extensions to the VSLD loops, referred to as "loop 1" and "loop 3" and the structure and role of these loops was unclear. We have solved the structure of this region, revealing a novel, highly structured domain that we designated as the Tetragonal Opening for Polycystins (TOP) domain (referred to as the polycystin domain in Shen et al., [32]). PC2 is present in a variety of membranes, including the ER, plasma membrane and cilia. When PC2 is in the ER, the TOP domain will reside in the ER lumen, and when it is in the plasma membrane or the cilia, it will be on the extracellular surface. In kidney tubule epithelial cells it will reside in the renal tubule lumen. The TOP domain comprises loop1, a 219-residue insertion (Ser244 to Leu462) between S1 and S2 of the VSLD, packed against a 20-residue extension to the S3-S4 loop (Val532 to Leu550) (Fig. 2a-c). It contains a central five-stranded antiparallel β-sheet, with three α-helices (α1 - α3) on one face forming the upper surface of the domain. The lower surface of the β-sheet, facing the pore domain surface, is covered by two elements, the VSLD S3-S4 loop extension (including helix S3b) and a β-hairpin (Fig. 2c). A notable feature of all polycystins is the "polycystin motif"[33]; which lies in the core of the TOP domain, forming strand β1 and the start of the first β-hairpin in the TLC (Fig. 2d). The β-hairpin is part of an unusual, extended structure that resembles a three leaf clover (TLC), consisting of three irregular β-hairpins arranged at approximately 90˚ to each other, which we called the TLC extension (Fig. 2c, e). The three TLC β-hairpins may be involved in cooperativity between channel subunits; as they form connections with different parts of the channel, the first β-hairpin (TLC1) connects adjacent TOP domains, extending as far as the S3-S4 extension of the adjacent TOP domain; (Fig. 3), the second (TLC2) wraps around the edge of the β-sheet and the third (TLC3) extends towards the pore. The three TLC β-hairpins therefore provide links between adjacent TOP domains, the S3-S4 extension and the pore turret.

Glycosylation of the TOP domain is important for stability and trafficking of PC2[34]. Despite extensive enzymatic deglycosylation, three of the five proposed N-glycosylation sites (Asn328, Asn362 and Asn375) have one or two N-acetyl glucosamine (NAG) units resolved;

Asn299 is in the Ser298 – Ala303 region, which is not resolved in this structure and Asn305 is adjacent to the disordered region and its glycosylation status is unclear (Supplementary Fig. 3f). The significance of the glycosylation sites in PC2 are discussed in Wilkes *et al.*,[35].

The extensive interactions between the TOP domains and the channel surface are clearly important for the stability of the PC2 structure and they may also play a role in cooperativity between subunits. Four TOP domains assemble to form a tightly-packed tetrameric ring that extends 35 Å above the membrane, covering the luminal surface of the VSLD and the pore (Fig. 3a). Viewed from above, the β-strands and α-helices form a square around a wide central opening. Each TOP domain forms extensive interactions with adjacent TOP domains and with the upper face of the channel (Fig. 3b, c). The extensive (1,340 Å^2) interface between TOP domains comprises interactions between the loops and TLC extension of one TOP domain and the loops at the other end of the β-sheet of the adjacent TOP domain. The first TLC β-hairpin contacts not only the β-sheet of the adjacent domain, but also the S3b helix at the top of the adjacent VSLD and this interaction allows communication between one TOP domain and the S3-S4 extension in the adjacent chain, thus facilitating inter-subunit cooperativity. Many of the residues involved in these inter-domain contacts are highly conserved (as calculated from an alignment of 177 PC2 sequences using the CONSURF webserver[36], Supplementary Figs. 1 and 6). The tetrameric ring of TOP domains shows extensive interactions distal to the pore axis. Proximal to the pore axis, however, there are relatively few interactions and a groove separates the individual TOP domains, providing a route for ions to approach the top of the pore filter.

There are extensive interactions between the TOP domain and the top of the channel region of PC2, which will be important for regulation of channel function by the TOP domain (and also interactions with lipids[35]). The TOP domain directly contacts both the pore turret of the adjacent chain and the VSLD of its own chain, providing opportunities for the TOP domain to influence the opening of the pore filter and thus to control channel activity. The substantial interface (770 Å^2) between the TOP domain β-sheet and the extended loop between S3 and S4 has several conserved residues and could be involved in regulation of the channel by the TOP domain (Fig. 3c, Supplementary Fig. 6b). Asn375 on the TLC extension is glycosylated and the oligosaccharide chain is packed against S3b-S4 region, suggesting glycosylation may be important for this interdomain interaction. The turret region of the pore of the adjacent protein chain (the S5-PH1 and the PH2-S6 loops) interacts with the TOP domain helix α1 and the β5 strand, forming a contact surface that is also partially conserved (Supplementary Figs. 1 and 6d). Such interactions provide a connection between the voltage sensors and the turret region

of the adjacent pore domains, thus providing a mechanism for cooperativity between neighboring chains.

Below the TOP domain there is a wide upper vestibule that is accessible to the ER lumen or extracellular environment through an apical opening and four lateral portals (Fig. 4a, b). The 9 Å wide apical opening on the pore axis is at the center of the TOP domain ring and is formed by the α1-α2 loop (EL1). Although the Thr265 and Lys267 sidechains project into the opening they would not impede cation movement (Fig. 3a). It is possible that movement of the TOP domain towards the pore axis or movement of the α1-α2 loop could block the ion path. However, the loop residues are not conserved, so it seems unlikely that they are critical for channel gating. The lateral portals may provide an alternative route for ion entry into the upper vestibule. The 40 Å wide and 20 Å high upper vestibule has a highly negatively charged surface, with each protomer contributing seven acidic residues on the pore domain upper surface and one on the TOP domain (Fig. 4c, d). Mutation of two of these negatively-charged residues (Asp643 and Glu648), results in changes in relative cation permeability[28]. The negative charge distribution in the upper vestibule would be expected to encourage cations to accumulate near the pore filter.

A pore domain with two closed gates

TRP channels have two gates, one at the selectivity filter and the other at the lower gate, formed by S6; in this structure of PC2 both gates appear to be closed (Fig. 4e,f). The PC2 S5-Pore-S6 domain has a classic P-loop structure between S5 and S6, with a pore helix (PH1), a two-residue pore filter (Leu641 to Gly642), followed by a three residue strand and a short second pore helix (PH2) (Fig. 5d). This conformation resembles the closed conformations of other TRP channels[30,37] (Fig. 5a-c). The narrowest parts of the pore filter, defined by opposing mainchain carbonyl oxygens of Leu641 and Gly642, are only 4.8 and 4.2 Å wide. Therefore, the available radius of the pore filter is only 0.56 Å at Gly642 and 0.88 Å at Leu641 (taking into account the van der Waals radii of the carbonyl oxygen atoms, as calculated by HOLE[38]). Not only would hydrated Na^+, K^+ and Ca^{2+} ions be too large to pass through, having radii of more than 3.3 Å, but even dehydrated ions are too large to permeate, Na^+, K^+ and Ca^{2+} ions having radii of 1.2, 1.5 and 1.0 Å respectively. The narrow pore filter is the result of PH1 being tilted upwards by 10° compared to TRPV1, so that the pore filter is moved 3-4 Å up the pore axis (Fig. 5b,c). External Ca^{2+} inhibits monovalent cation currents in PC2 and the residue positioned at the entrance of the pore filter, Asp643, has been implicated in this process[28]. Asp643 could coordinate a Ca^{2+} ion thus blocking monovalent cation flux, as is seen for the

Ca$_v$Ab and TRPV6 channels[39,40], although in the closed PC2 structure we do not see ordered Ca^{2+} ions. The TRPV6 structure also shows a partially hydrated Ca^{2+} ion within the selectivity filter[40], and this requires a pore radius of >2 Å at the selectivity filter, compared to 0.56 Å in PC2, further confirming that in this structure the selectivity filter is closed.

In addition to the closed selectivity filter, there are two further constrictions below the selectivity filter that are too narrow for hydrated ions to pass through, suggesting that the lower gate is also closed (Fig. 4e, f). There is a narrowing of the central vestibule at Phe669 and a lower gate at the S6 helix bundle crossing below the central vestibule, involving residues Leu677 and Asn681 (Fig. 4e). S5 is bent between residues Gly599 and Phe600 and S6 has a π-helix forming a bulge in the α-helix between Phe669 and Ile671, as is seen in TRPV1[30]. Mutation of Phe604 to proline caused constitutive PC2 channel activation in a model system, whereas the Phe604Ala and Phe604Ile mutations did not[28]. Phe604 lies near the S5 helix bend, with its sidechain at the center of a cluster of hydrophobic sidechains from S6 of the same chain (Met668, Leu672, Phe676), and the adjacent chain (Phe666 and Phe670) (Fig. 5e). The Phe604Pro mutation could constitutively activate the channel by disrupting both the S5 helix structure and the hydrophobic cluster that stabilizes the distortions of S5 and S6. Straightening of S5 and S6 could lead to movement of the lower gate residues Leu677 and Asn681, thus opening this gate, an activation mechanism similar to that seen in other TRP channels[30].

The lower gate could also be controlled by changes in the cytoplasmic N- and C-terminal domains and their interactions with other proteins. The N-terminus includes a ciliary targeting motif[41]. The C-terminus contains two EF-hand motifs, that may act as a calcium sensor[42], and a coiled-coil domain. The latter is believed to be required for hetero-oligomerization with the interaction partners, including PC1 (reviewed in Yang and Erlich[43]). The N- and C-termini of PC2 also mediate interactions with cytoskeletal proteins[44] and they contain three phosphorylation sites[45-47] which may affect both interactions between individual PC2 domains and interactions with other proteins. The structure described here is truncated at both N- and C-termini, so these domains are not present. Interactions between the EF-hand and the cytoplasmic face of PC2 are described in Wilkes *et al.*[35]. In general however the N- and C-terminal domains of PC2 may not be rigidly connected to the channel part of the structure[32,35], at least in some conformations adopted by PC2. The details and extent of these interactions, and the interactions between PC2 and other proteins, remain to be determined.

ADPKD-associated missense variants

Sequencing the *PKD2* and *PKD1* genes of patients with ADPKD has identified at least 278 changes in the *PKD2* gene and 2,323 in the *PKD1* gene (http://pkdb.mayo.edu/). These changes result in truncations, insertions, deletions and frame shift mutations in the PC2 and PC1 proteins. There are also a series of single amino acid substitutions which we will refer to as missense variants, rather than mutations, except in cases where a change in activity has been demonstrated. At least 27 germ-line missense variants in the PC2 channel and TOP domain have been identified in patients with ADPKD, but not in the general population (from the NCBI SNP[48] and ExAC databases[49]). These missense variants are distributed throughout the structure, but are more frequent in the TOP domain, which makes up only 40% of the PC2 channel region sequence, but has 60% of the ADPKD-associated missense variants in this region (Fig. 6a, Supplementary Table 1, Supplementary Fig. 7).

Many of the TOP domain variants occur in residues that form the core of the TOP domain fold and changes in this region are likely to be destabilizing (Fig. 6b,c; Supplementary Fig. 7). These TOP domain variants include changes to sidechains packed between the β-sheet and α-helices (Trp280Arg, Tyr292Cys and Trp414Gly); on the lower surface of the β-sheet (Arg322Trp and Arg440Ser) and on the TLC3 β-hairpin that packs against the β-sheet (Gly390Ser and Ser378Del). The guanidinium moiety of Arg322 and Arg440 sidechains are 4 Å apart and they will interact via the adjacent acidic sidechains of Asp425 and Glu559. In addition Arg440 is adjacent to the proposed lipid binding site described in the accompanying paper (Wilkes et al.,[35]) and could stabilize the lipid head groups. The Trp414Gly mutation is intriguing as it shows reduced activity with effective trafficking in oocytes (in a model system with a gain-of-function mutation[28]), but loss of trafficking with conserved channel function in pig kidney epithelial cells[50]. Therefore the Trp414Gly mutation is capable of affecting both trafficking and channel function in an unpredictable way.

A total of 13 ADPKD-associated missense variants map to the conserved interface between adjacent TOP domains (Fig. 6b,c; Supplementary Figs. 6 and 7). Arg420 lies directly between the subunits and the pathogenic mutation Arg420Gly causes loss of ion currents[28]. In contrast, Arg325 is closer to the surface, where it forms a π-orbital stacking interaction with the sidechain of Phe358. The Arg325Pro variant has little effect on ion flux[28]; it may instead be important for trafficking of the PC1–PC2 complex. Arg420, Tyr345 and the adjacent Gly418 seem to be key residues, as they stabilize the intersection of the three TLC loops and the TOP-TOP interface, such that any mutation of these residues (e.g. Tyr345Cys and Gly418Val) would disrupt the TOP-TOP interface and interfere with the TOP domain movements discussed in Wilkes et al.[35] (Fig. 6b,c).

In addition to mutations in the TOP domain, there are four missense variants that fall within the VSLD and six that lie in the pore domain (Supplementary Table 1, Supplementary information). The well-characterized Asp511Val channel-inactivating mutation[17,28] lies on S3, with its sidechain facing into the space between the S1-S4 helices (Fig. 6d). It is central to a hydrogen-bonding network involving two intramembrane lysines that may be involved in voltage sensing (Lys572 and Lys575), as well as Tyr487, Glu491 and Asn508. Loss of the acidic group at the center of the H-bond network would disrupt the interhelix contacts within the VSLD and would affect changes in conformation[35]. Changes in the S5 helix also appear to affect channel function, notably in the region of Phe604 and Phe605. Site-directed mutagenesis of Phe604Pro leads to constitutive activation of the channel. Some patients with ADPKD have a deletion of Phe605 and site directed mutagenesis to remove Phe605 in a Phe604Pro background causes loss of channel function[28], potentially by disrupting opening of the lower gate. Five missense variants found in patients with ADPKD lie directly within the pore domain region. Substantial changes in the charge and/or the size of sidechains before (Ala615Thr (S5)), on PH1 (Phe629Ser, Cys632Arg, Arg638Cys) or at the N-terminus of S6 (Leu656Trp) are likely to be pathogenic, as they will disrupt the pore architecture and movements that are required to open the pore filter (Fig. 6e). For further information on individual mutations, see Supplementary Table 1 and Supplementary Note.

85 % of patients with ADPKD have mutations in the *PKD1* gene, which encodes a 4,303 residue protein, PC1, with a PC2-homologous, channel-like domain between Arg3666 and Glu4123 (Supplementary Fig. 1). Modeling of the PC1 channel-like domain based on the PC2 structure confirms that PC1 could form a similar overall fold to PC2 with a VLSD, a TOP domain and the pore region. However, PC1 is not thought to form a functional channel and the presence of two arginine residues (Arg4100 and Arg4107) on S6, whose sidechains are predicted to lie on the pore axis by the homology model, may contribute to the lack of channel function. Like PC2, the PC1 channel-like domain has many potentially pathogenic variants in the TOP, VSLD and pore domains (Fig. 6f and Supplementary Fig. 7b), suggesting that changes in the PC1 TOP domain and S1-S6 structure disrupt PC1 function. In addition to PC1 and PC2, there are six other polycystin-like proteins (PKD1L1, PKD1L2, PKD1L3, PKD2L1, PKD2L2 and PKDREJ), with TOP domain structures that can now be modelled based on the PC2 structure.

Discussion

Many intriguing questions remain regarding the role that the channel, VSLD and TOP domains play in regulating PC2 function, and the differences in function of the PC2 homotetrameric channel in the ER and a PC1–PC2 complex in cilia and other membranes[23]. PC2 is regulated by diverse stimuli, including internal and external Ca^{2+}, voltage, pH, membrane stretch and phosphorylation, and interactions with other channels and cytoskeletal proteins. Our structure of PC2, together with published mutagenesis data[28] suggests that external Ca^{2+} would reduce monovalent cation currents by binding to Asp643 at the top of the pore filter. The structure of PC2 with an open pore filter, presented in Wilkes et al., 2017[35], provides an understanding of how Ca^{2+} ions can act as permeant ions and how the cytoplasmic EF hand could regulate channel activity. Membrane stretch could affect the pore filter through movements of the S5-S6 helices, as seen in the K2P potassium channels[51], or through displacement of the TOP, and N- or C-terminal domains; TOP domain movements have been observed between the closed and the two open conformations and these changes may be involved in mechanosensitivity, as well as Ca^{2+} signaling[35]. Voltage sensitivity[21] (Fig. 1e and Supplementary Fig. 4) could be mediated directly at the pore filter[52] or through lysine residues between the helices on the VSLD. Our structure also provides an understanding of how activation of the channel could occur, since an activatory mutation, Phe604Pro[28] on S5 would disrupt a hydrophobic cluster between S5 and S6 on the two chains, thus allowing straightening and movement of the helices and opening of the lower gate, a mechanism similar to that seen for other TRP channels[31,53]. The PC2 open pore-filter structures reveal two different modes of Ca^{2+} ion binding and how the channel can be regulated by lipids.

While this article was under review at Nature Structural and Molecular Biology, a paper was published reporting the structure of truncated forms of PC2[32] (referred to as PKD2) in nanodiscs. This structure shows a closed conformation of PC2, with a similar fold and orientation for the TOP domain (referred to as the polycystin domain) compared to the closed structure reported here. This work confirms that independent structure determinations of the same protein in different laboratories, in detergent or nanodiscs, can give very similar structures.

Our structure of the PC2 channel and associated "loops" reveals a new dimension to TRP channel fold and function. The "loop" between S1 and S2 of the VSLD, is in fact a highly structured domain with extensive contacts between the four TOP domains in the tetrameric channel, the VSLDs and the pore. Many ADPKD-associated missense variants that would disrupt either the TOP domain fold or the TOP-TOP interface lead to loss of PC2 channel

activity and trafficking of the PC1–PC2 complex[28,50]. The TOP domain, VSLD and pore domain harbor 27 of the missense variants that occur in patients with ADPKD, and several of the variants have been shown to cause loss of channel activity[28]. However, we cannot rule out the possibility that some variants will increase PC2 activity, since overexpression of PC2 in mice also causes kidney cyst formation[54]. The structure of PC2 with its ER or extracellular TOP domain suggests potential channel-gating mechanisms and provides a step change in our understanding of the function of the polycystin family of proteins and their role in disease.

Accession Codes

The cryo-EM density map has been deposited in the Electron Microscopy Data Bank (EMDB) under the accession code EMD-8200. The atomic coordinates of the corresponding structure have been deposited in the Protein Data Bank (PDB) under the accession code 5K47.

Acknowledgements

Author Contributions

M.G., A.C.W.P. and C.A.S. contributed equally to this project. M.G., A.C.W.P., R.S., J.T.H. and E.P.C. designed the project. M.G. optimised and prepared protein samples for X-ray, cryoEM and electrophysiology data collection, assisted by A.T. C.A.S. and A.C.W.P. established the cryo-EM process at the SGC. M.G., C.A.S., A.C.W.P., J.T.H. and E.P.C. collected EM and X-ray data. A.C.W.P. determined the structures, assisted by M.G., C.A.S., J.T.H. and E.P.C. L.S., P.M. and S.M. performed pilot studies for protein production and produced baculovirus infected cells, and R.C. analysed the mass spectrometry data. Work by L.S., P.M., S.M. and R.C. was supervised by N.A.B-B., M.G., E.V. and S.E.A. performed the electrophysiology experiments, supervised by R.S. J.T.H. was responsible for supervising the cryo-EM work. E.P.C. was responsible for the overall project design, supervision and management. M.G., A.C.W.P., E.V., S.E.A., R.S., J.T.H. and E.P.C. analysed the data and prepared the manuscript.

Competing Financial Interests Statement

The authors declare no competing financial interests.

References

1   Wilson, P. D. Polycystic kidney disease. *N Engl J Med* **350**, 151-164 (2004).

2   Hateboer, N. *et al.* Comparison of phenotypes of polycystic kidney disease types 1 and 2. European PKD1-PKD2 Study Group. *Lancet* **353**, 103-107 (1999).

3   Du, J., Fu, J., Xia, X. M. & Shen, B. The functions of TRPP2 in the vascular system. *Acta Pharmacol Sin* **37**, 13-18 (2016).

4   Chapman, A. B., Stepniakowski, K. & Rahbari-Oskoui, F. Hypertension in autosomal dominant polycystic kidney disease. *Adv Chronic Kidney Dis* **17**, 153-163 (2010).

5   Karihaloo, A. *et al.* Macrophages Promote Cyst Growth in Polycystic Kidney Disease. *Journal of the American Society of Nephrology* **22**, 1809-1814 (2011).

6   Mochizuki, T. *et al.* PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* **272**, 1339-1342 (1996).

7   Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. The International Polycystic Kidney Disease Consortium. *Cell* **81**, 289-298 (1995).

8   Pennekamp, P. *et al.* The ion channel polycystin-2 is required for left-right axis determination in mice. *Curr Biol* **12**, 938-943 (2002).

9   McGrath, J., Somlo, S., Makova, S., Tian, X. & Brueckner, M. Two populations of node monocilia initiate left-right asymmetry in the mouse. *Cell* **114**, 61-73 (2003).

10   Bataille, S. *et al.* Association of PKD2 (Polycystin 2) Mutations With Left-Right Laterality Defects. *American Journal of Kidney Diseases* **58**, 456-460 (2011).

11   Yoder, B. K., Hou, X. & Guay-Woodford, L. M. The polycystic kidney disease proteins, polycystin-1, polycystin-2, polaris, and cystin, are co-localized in renal cilia. *J Am Soc Nephrol* **13**, 2508-2516 (2002).

12   Pazour, G. J., San Agustin, J. T., Follit, J. A., Rosenbaum, J. L. & Witman, G. B. Polycystin-2 localizes to kidney cilia and the ciliary level is elevated in orpk mice with polycystic kidney disease. *Curr Biol* **12**, R378-380 (2002).

13   Nauli, S. M. *et al.* Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. *Nat Genet* **33**, 129-137 (2003).

14   Praetorius, H. A. & Spring, K. R. Bending the MDCK cell primary cilium increases intracellular calcium. *J Membr Biol* **184**, 71-79 (2001).

15   Delling, M. *et al.* Primary cilia are not calcium-responsive mechanosensors. *Nature* **531**, 656-660 (2016).

16   Norris, D. P. & Jackson, P. K. Cell biology: Calcium contradictions in cilia. *Nature* **531**, 582-583 (2016).

17   Koulen, P. *et al.* Polycystin-2 is an intracellular calcium release channel. *Nat Cell Biol* **4**, 191-197 (2002).

18   Giamarchi, A. *et al.* The versatile nature of the calcium-permeable cation channel TRPP2. *EMBO Rep* **7**, 787-793 (2006).

19   Ramsey, I. S., Delling, M. & Clapham, D. E. An introduction to TRP channels. *Annu Rev Physiol* **68**, 619-647 (2006).

20   Gonzalez-Perrett, S. *et al.* Polycystin-2, the protein mutated in autosomal dominant polycystic kidney disease (ADPKD), is a Ca2+-permeable nonselective cation channel. *Proc Natl Acad Sci U S A* **98**, 1182-1187 (2001).

21   Gonzalez-Perrett, S. *et al.* Voltage dependence and pH regulation of human polycystin-2-mediated cation channel activity. *J Biol Chem* **277**, 24959-24966 (2002).

22   Torres, V. E. & Harris, P. C. Autosomal dominant polycystic kidney disease: the last 3 years. *Kidney Int* **76**, 149-168 (2009).

23   Hanaoka, K. *et al.* Co-assembly of polycystin-1 and -2 produces unique cation-permeable currents. *Nature* **408**, 990-994 (2000).

24   Wu, G. *et al.* Somatic inactivation of Pkd2 results in polycystic kidney disease. *Cell* **93**, 177-188 (1998).

25   Reeders, S. T. Multilocus polycystic disease. *Nat Genet* **1**, 235-237 (1992).

26   Pei, Y. A "two-hit" model of cystogenesis in autosomal dominant polycystic kidney disease? *Trends Mol Med* **7**, 151-156 (2001).

27   Zhang, P. *et al.* The multimeric structure of polycystin-2 (TRPP2): structural-functional correlates of homo- and hetero-multimers with TRPC1. *Hum Mol Genet* **18**, 1238-1251 (2009).

28   Arif Pavel, M. *et al.* Function and regulation of TRPP2 ion channel revealed by a gain-of-function mutant. *Proceedings of the National Academy of Sciences* (2016).

29   Catterall, W. A. Ion Channel Voltage Sensors: Structure, Function, and Pathophysiology. *Neuron* **67**, 915-928 (2010).

30   Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107-112 (2013).

31   Zubcevic, L. *et al.* Cryo-electron microscopy structure of the TRPV2 ion channel. *Nat Struct Mol Biol* **23**, 180-186 (2016).

32   Shen, P. S. *et al.* The Structure of the Polycystic Kidney Disease Channel PKD2 in Lipid Nanodiscs. *Cell* **167**, 763-773 e711 (2016).

33   Nomura, H. *et al.* Identification of PKDL, a novel polycystic kidney disease 2-like gene whose murine homologue is deleted in mice with kidney and retinal defects. *J Biol Chem* **273**, 25967-25973 (1998).

34   Hofherr, A., Wagner, C., Fedeles, S., Somlo, S. & Kottgen, M. N-glycosylation determines the abundance of the transient receptor potential channel TRPP2. *J Biol Chem* **289**, 14854-14867 (2014).

35   Wilkes, M. *et al.* **Accompanying paper - for back-to-back publication.**

36   Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* **38**, W529-533 (2010).

37   Paulsen, C. E., Armache, J. P., Gao, Y., Cheng, Y. & Julius, D. Structure of the TRPA1 ion channel suggests regulatory mechanisms. *Nature* **520**, 511-517 (2015).

38   Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J Mol Graph* **14**, 354-360, 376 (1996).

39   Tang, L. *et al.* Structural basis for Ca2+ selectivity of a voltage-gated calcium channel. *Nature* **505**, 56-61 (2014).

40   Saotome, K., Singh, A. K., Yelshanskaya, M. V. & Sobolevsky, A. I. Crystal structure of the epithelial calcium channel TRPV6. *Nature* **534**, 506-511 (2016).

41   Geng, L. *et al.* Polycystin-2 traffics to cilia independently of polycystin-1 by using an N-terminal RVxP motif. *J Cell Sci* **119**, 1383-1395 (2006).

42   Kuo, I. Y. *et al.* The number and location of EF hand motifs dictates the calcium dependence of polycystin-2 function. *FASEB J* **28**, 2332-2346 (2014).

43   Yang, Y. & Ehrlich, B. E. Structural studies of the C-terminal tail of polycystin-2 (PC2) reveal insights into the mechanisms used for the functional regulation of PC2. *J Physiol* **594**, 4141-4149 (2016).

44   Chen, X. Z. *et al.* Submembraneous microtubule cytoskeleton: interaction of TRPP2 with the cell cytoskeleton. *FEBS J* **275**, 4675-4683 (2008).

45   Streets, A. J., Moon, D. J., Kane, M. E., Obara, T. & Ong, A. C. Identification of an N-terminal glycogen synthase kinase 3 phosphorylation site which regulates the functional localization of polycystin-2 in vivo and in vitro. *Hum Mol Genet* **15**, 1465-1473 (2006).

46   Streets, A. J., Needham, A. J., Gill, S. K. & Ong, A. C. Protein kinase D-mediated phosphorylation of polycystin-2 (TRPP2) is essential for its effects on cell growth and calcium channel activity. *Mol Biol Cell* **21**, 3853-3865 (2010).

47   Cai, Y. *et al.* Calcium dependence of polycystin-2 channel activity is modulated by phosphorylation at Ser812. *J Biol Chem* **279**, 19987-19995 (2004).

48   Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311 (2001).

49   Exome Aggregation Consortium (ExAC), C., MA (URL: http://exac.broadinstitute.org).

50   Cai, Y. *et al.* Altered trafficking and stability of polycystins underlie polycystic kidney disease. *J Clin Invest* **124**, 5129-5144 (2014).

51   Dong, Y. Y. *et al.* K2P channel gating mechanisms revealed by structures of TREK-2 and a complex with Prozac. *Science* **347**, 1256-1259 (2015).

52   Schewe, M. *et al.* A Non-canonical Voltage-Sensing Mechanism Controls Gating in K2P K$^+$ Channels. *Cell* **164**, 937-949.

53   Cao, E., Liao, M., Cheng, Y. & Julius, D. TRPV1 structures in distinct conformations reveal activation mechanisms. *Nature* **504**, 113-118 (2013).

54   Park, E. Y. *et al.* Cyst formation in kidney via B-Raf signaling in the PKD2 transgenic mice. *J Biol Chem* **284**, 7214-7222 (2009).

Figure legends
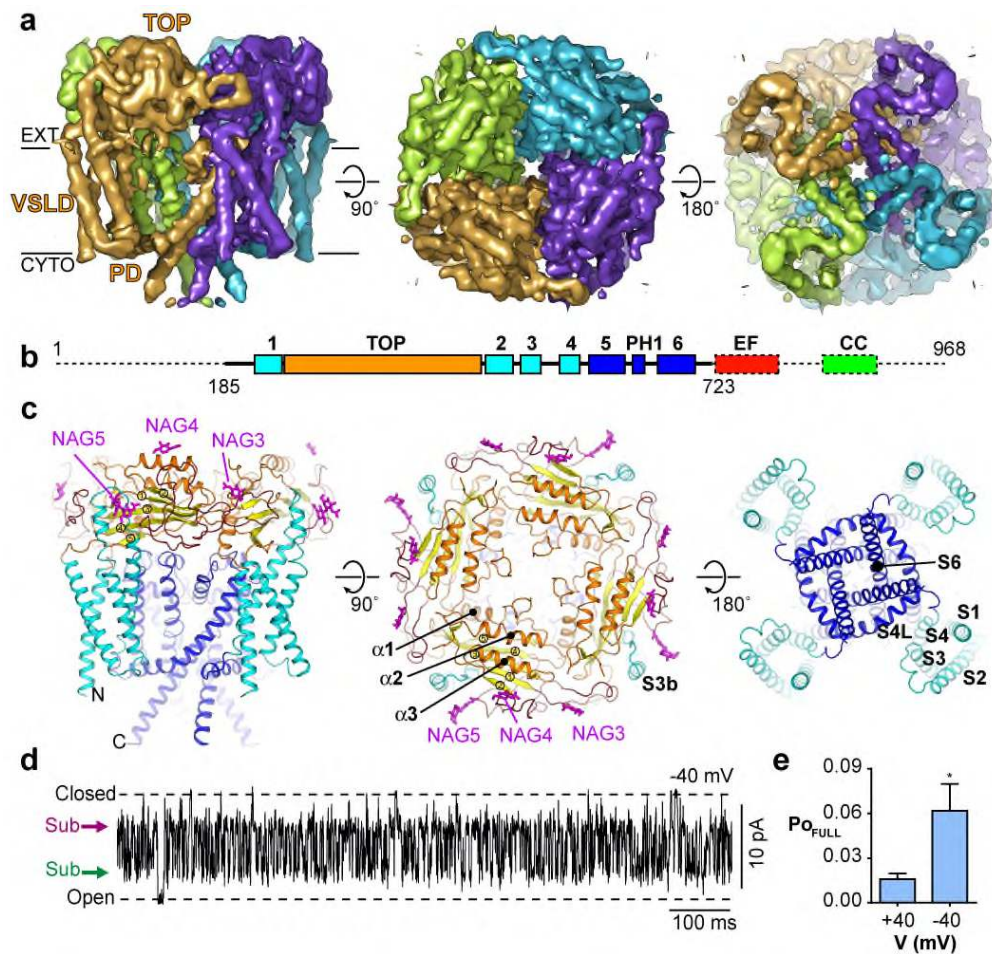
**Figure 1: PC2 structure and channel activity. a,** Reconstruction of PC2$^{P185-D723}$ at 4.2 Å resolution. Views of the side, luminal and cytoplasmic faces are shown, colored by monomer. **b,** Domain organization of PC2. The construct (PC2$^{P185-D723}$) optimized for crystallization and subsequently used for cryo-EM is indicated by non-dotted lines. (PH1, pore helix1; EF, EF-hand domain; CC, coiled coil). **c,** Ribbon representation of PC2 tetramer organization viewed corresponding to the reconstructions shown in (**a**). Chains are colored according to domain with VSLD S1-S4 helices (cyan) and TOP (orange/yellow), pore domains (PD) (S5, PH1, PH2, S6) (blue). Oligosaccharides at glycosylation sites (NAG3-5) are depicted as sticks (magenta). **d,** Typical example of the complex sub-conductance behavior of PC2$^{P185-D723}$ in symmetrical 210 mM KCl, pH 7.2 at -40 mV. Fully open and closed states are illustrated by dashed lines and the predominant sub-conductance states are indicated by arrows. **e,** The effect of holding potential on the probability of PC2$^{P185-D723}$ dwelling in the full open state (Po$_{FULL}$). Po$_{FULL}$ is significantly higher at negative potentials (n=5 independent experiments, mean ± s.e.m., Student's *t*-test, *$P<0.05$).
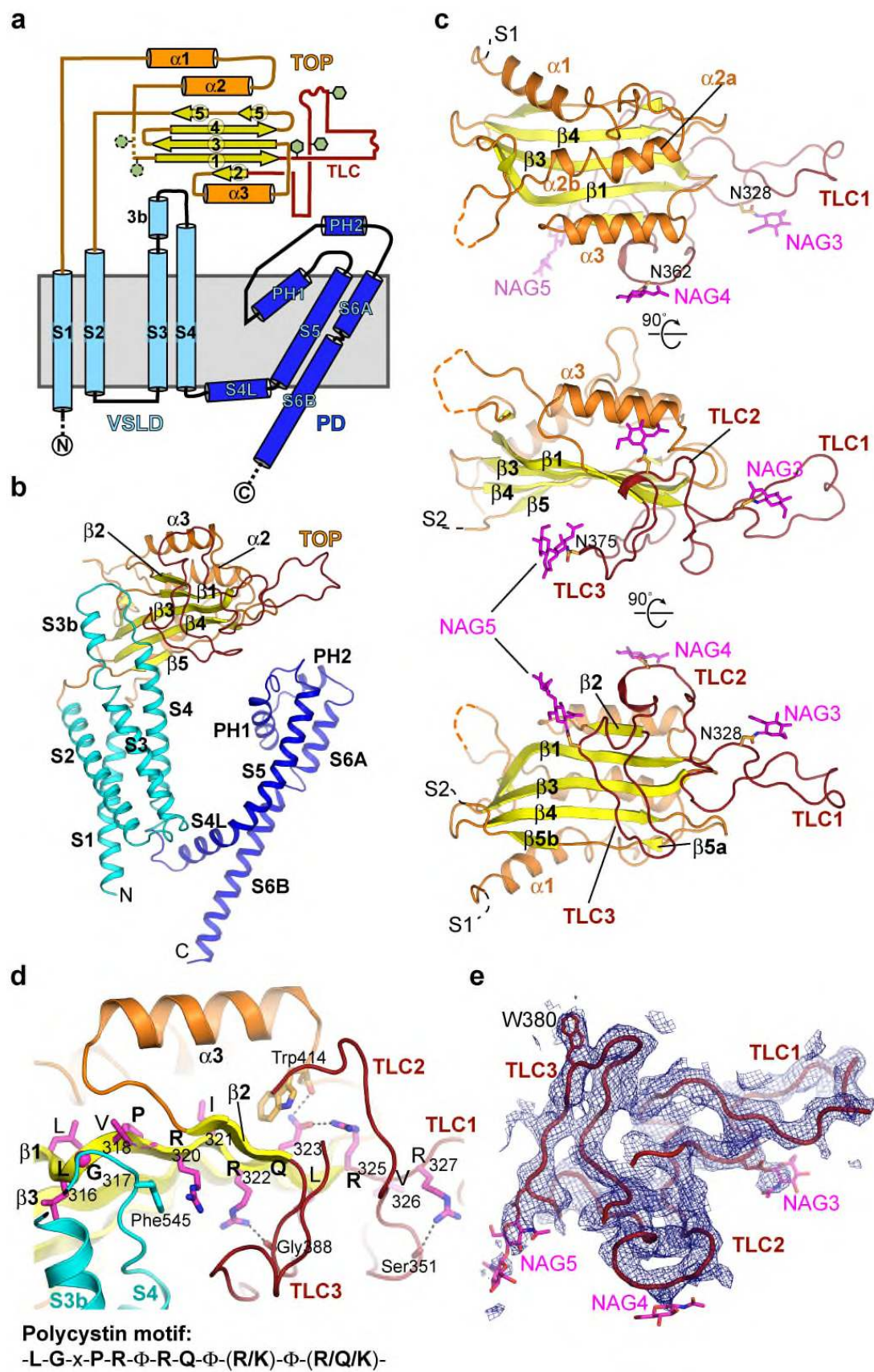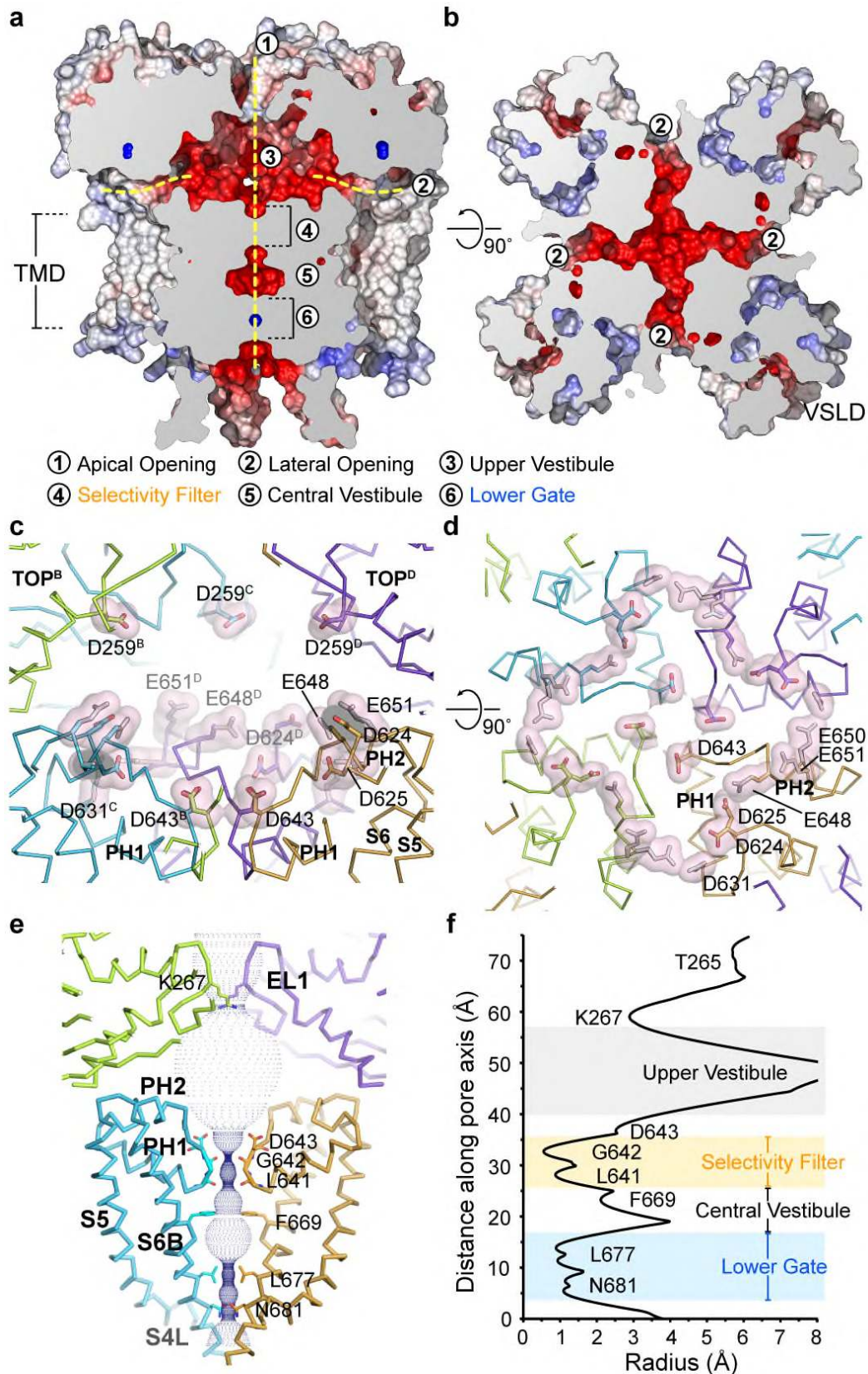
Figure 2: PC2 TOP domain structure. a. Secondary structural organization showing VSLD S1-S4 helices (cyan) and TOP (orange/yellow), pore domains (PD) (S5, PH1, PH2, S6) (blue). Cylinders, arrows and hexagons indicate helices, β-strands and glycosylation sites,

respectively. **b,** The PC2 monomer structure colored as in (**a**). **c,** Schematic representation of the TOP domain. Three perpendicular views are shown. The central five-stranded β-sheet (β1-5; yellow) is flanked on its upper face with three α-helices (orange; α1-3). A long loop between strands β1 and β2 (Asn328-Gly390) forms the 'Three Leaf Clover' (TLC) extension (dark red) comprising three β-hairpins (TLC1-3), each bearing a single glycosylation site (NAG3-5, magenta). **d,** Role of the conserved 'polycystin motif'[33]. The polycystin motif (residues Leu316-Asn328) forms the β1 strand of the TOP β-sheet. The first half of the motif forms part of the interaction site for the S3b-S4 extension of the VSLD (cyan). The conserved basic arginine residues project on both faces of the β-sheet and make interactions with other elements of the domain. **e,** Density map for the three-leaf clover (TLC) extension filtered at 4 Å and sharpened with a *B*-factor of –150 Å^2. The density map is contoured at 4.5σ. The glycosylation sites (NAG3-5; magenta sticks) are more apparent in the unsharpened map (Supplementary Fig. 3f).
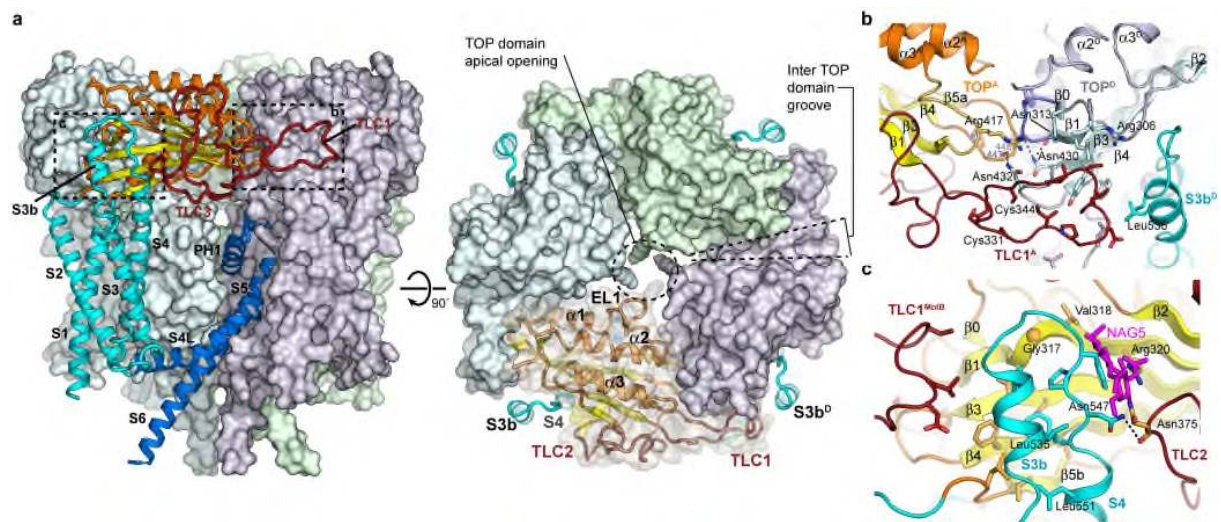
Figure 3: PC2 TOP domain interactions. a-c, Intra- and intermolecular interactions of the TOP domain ring. **a,** Perpendicular views of the PC2 tetramer looking from side or top, with the protein chain for one protomer and a molecular surface for the remaining three protomers. Regions shown in detail in **b-c** are boxed. **b,** The TLC extension intermolecular interaction with the neighboring TOP domain β-sheet (TOP^D) and S3-S4 extension. **c,** Intramolecular interaction between the lower face of the β-sheet and the S3-S4 loop extension (cyan).

① Apical Opening ② Lateral Opening ③ Upper Vestibule
④ Selectivity Filter ⑤ Central Vestibule ⑥ Lower Gate

Figure 4: Architecture of the PC2 pore. ab, Sliced molecular surfaces colored according to electrostatic potential showing pore constrictions and vestibules, viewed from the membrane (**a**) and (**b**) looking down the pore axis from the ER/renal tubule luminal face. (**c-d**) Location of negatively charged residues surrounding the pore entrance in the upper vestibule viewed

21

perpendicular **(c)** or down **(d)** the ion conduction pathway. Residues depicted by grey sticks were not resolved in the structure and have been modelled for illustrative purposes **e,** Pore profile (purple dotted surface) shown along with two diagonally opposed protomers. Residues in selectivity filter and at pore constrictions are shown as sticks. **f,** Pore radius plotted along the ion conduction pathway. Residues at the constriction points along the pore axis, where the radius narrows to 1 Å, are highlighted.
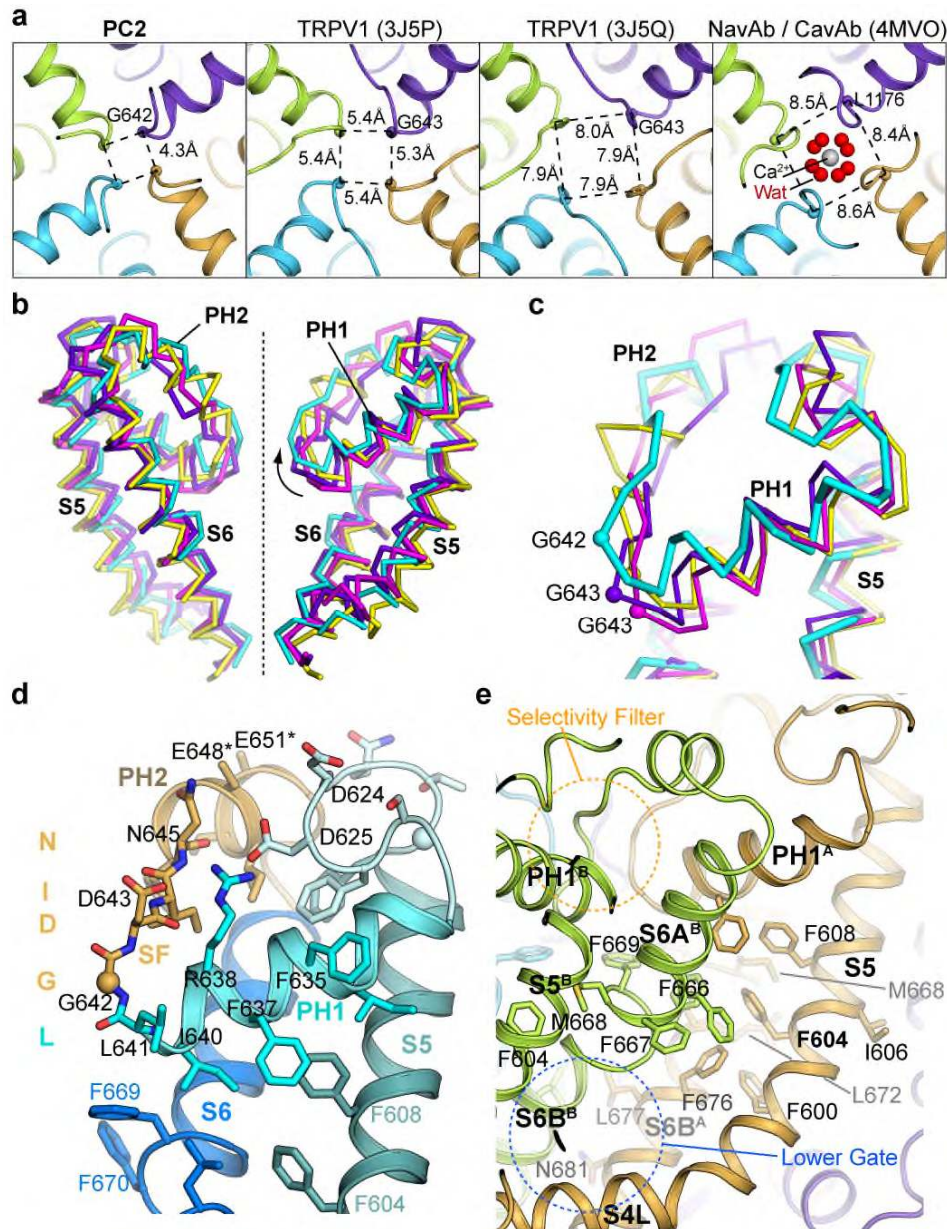
Figure 5: Comparison of the PC2 pore with pores from other VGICs. a, Pore viewed down the fourfold axis highlighting separation of the pore helices/SF for PC2 (closed filter), TRPV1 (closed filter: 3J5P; open filter: 3J5Q) and CavAb (open filter). The distance indicated by dashed lines represents the Cα-Cα separation of the first residue of the SF in each case. A hydrated Ca^{2+} ion (grey) is shown in the CavAb filter. **b,c** Superposition of pore domains for PC2 (cyan), TRPV1 (3J5P; purple), TRPV1 (3J5Q; magenta) and Ca_vAb (4MVO; yellow). **b,** two adjacent pore domains are shown viewed perpendicular to the pore axis. **c,** close-up view of PH1 region. The PH1 of PC2 is tilted by 10° resulting in the selectivity filter (SF) shifting 3-4Å up towards the ER lumen/extracellular face of the membrane. Consequently, the narrowest constriction in the filter is defined by the mainchain carbonyl of Leu641 in PC2, a residue that precedes the SF motif. The Cα position of the first residue in the SF motif is highlighted. **d,** Detailed view of the PC2 PH region. Asterisks by residue name/number indicate acidic residues whose sidechains were not modelled due to lack of sufficient density. **e**, The interface between S5, S6A/B on chain A (gold) and S6A/B on

chain B (green), with Phe604 and the surrounding cluster of hydrophobic residues. The location of the selectivity filter and the lower gate are indicated by dotted circles.
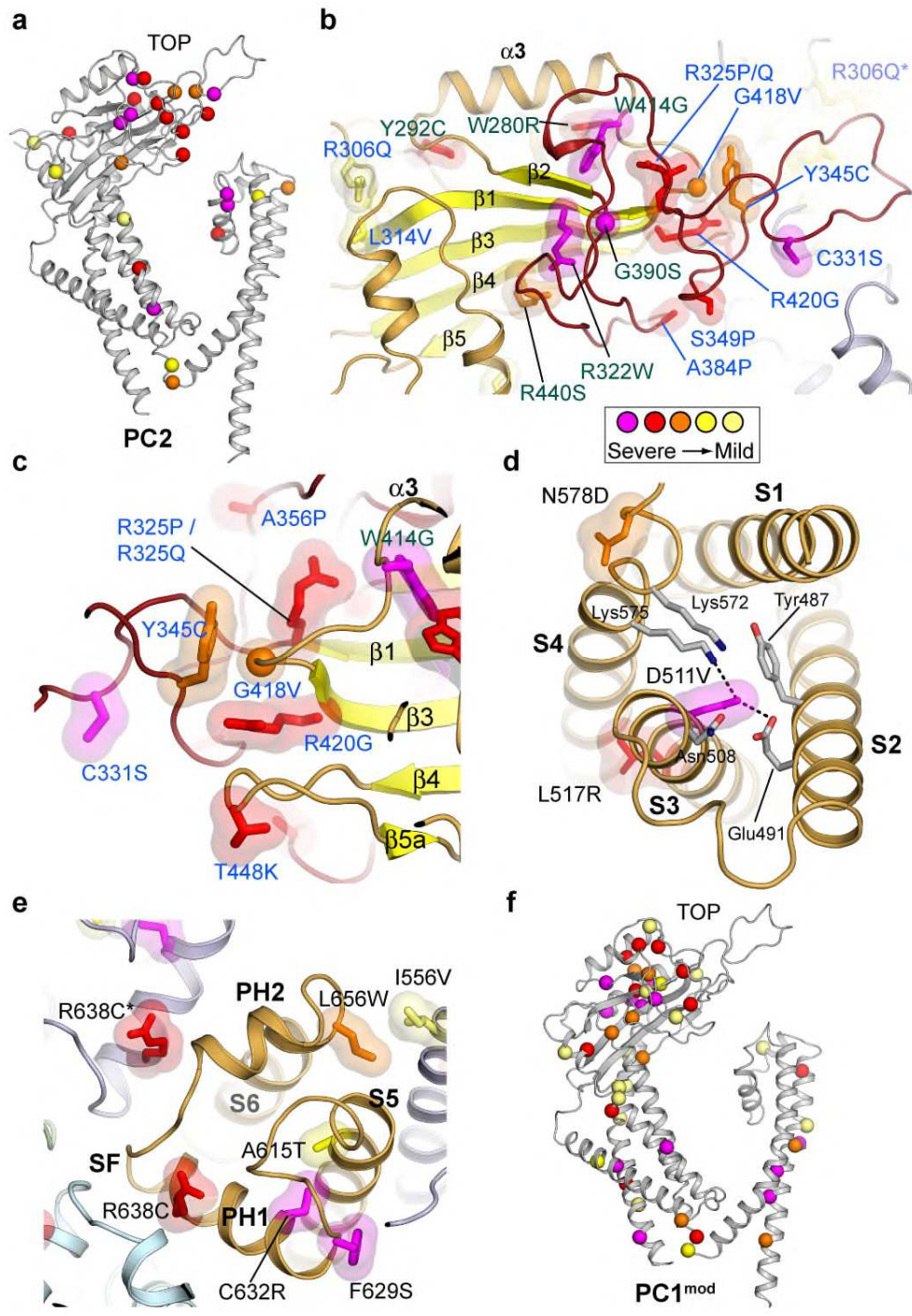
Figure 6: Locations of ADPKD-associated missense variants in PC2 and PC1. a, Location of ADPKD variants mapped onto the PC2 monomer. Variant sites are highlighted by spheres colored according to the predicted impact of the amino acid change on the channel structure (low (yellow) to high (magenta); see Supplementary Table 1 for discussion of structural impact prediction). **b,** The TOP domain and **c,** TOP-TOP interface are shown with variants that map to the core of the structure (green labels) and mutations in the TOP-TOP interface (blue. variants in **d,** the VSLD and **e,** the pore domain. **f,** PC1 channel-like domain ADPKD variants mapped onto a model of the PC1 monomer.

Tables

Table 1 | Cryo-EM data, refinement and model statistics.

Cryo-EM data	
Data Collection:	
Microscope	Tecnai Polara F30
Voltage (kV)	300
Defocus range (μm)	1.1-4.5
Pixel size (Å)	1.35
Electron dose (e$^-$/Å^2)	45
Dose rate (e$^-$/pix^2/s)	9
Particles (initial)	150,882
Particles (final, %)	19,546 (13)
Reconstruction:	
Number of particles	19,546
Symmetry	C4
Resolution (unmasked, Å)	4.82
Resolution (masked, Å)	4.22
Map sharpening B-factor (Å^2)	-198.6
Model composition:	
Non-hydrogen atoms	15,680
Protein residues	1,952
Sugars	16
Refinement:	
Resolution (Å)	4.2
Map sharpening factor (Å^2)	-150
Fourier Shell Correlation (FSC)[1]	0.8741
R-factor[2]	0.2956
Rms deviations:	
Bonds (Å)	0.0073
Angles (º)	1.036
B-factors:	
Protein (all; Å^2)	123
Sugars (Å^2)	208
Molprobity validation:	
Clashscore, all atoms	0.84
Molprobity score	1.24
Good rotamers (poor) (%)	94.96 (0)
Ramachandran Plot (% favoured)	92.08
Ramachandran Plot (% outliers)	0.42

[1] $FSC_{overall} = \Sigma(N_{shell}\ FSC_{shell})\ /\ \Sigma(N_{shell})$, where FSC_{shell} is the FSC in a given shell and N_{shell} is the number of 'structure factors' in the shell. $FSC_{shell} = \Sigma(F_{model}\ F_{EM})\ /\ (\sqrt{(\Sigma(|F_{model}|^2))}\ \sqrt{(\Sigma(|F_{EM}|^2))})$.

[2] R factor $= \Sigma||F_{obs}|-|F_{calc}||\ /\ \Sigma|F_{obs}|$

Online Methods

Cloning and expression.

The human *PKD2* gene, which encodes the PC2 (Polycystin-2, PC2 or TRPP1) protein was purchased from the Mammalian Gene Collection (MGC:138466, IMAGE: 8327731, BC112263). The construct used for structure determination consisted of residues Pro185 to Asp723 of PC2 (PC2$^{P185-D723}$), with a C-terminal purification tag with a tobacco etch virus (TEV) protease cleavage site, a 10x His purification sequence and a FLAG tag, in the expression vector pFB-CT10HF-LIC (available from The Addgene Nonprofit Plasmid Repository). The construct design and protein purification were originally optimised for structure determination by X-ray crystallography, (see additional information section). Baculoviruses were produced by transformation of DH10Bac cells. *Spodoptera frugiperda* (Sf9) insect cells in 250ml of Sf900II^{TM} serum free, protein free insect cell medium with L-glutamine (Thermofisher Scientific) were infected with recombinant baculovirus for virus amplification for 65 h at 27°C in 1 L shaker flasks. The harvested virus was then used to infect 1 L of the same cells in Insect-XPRESS™ Protein-free Insect Cell Medium with L-glutamine (Lonza) in a 3 L flask, for 65 h at 27°C.

Purification of PC2 protein for structural and functional studies.

24 L insect cell culture pellets containing heterologous expressed PC2 were resuspended in 50 ml/L of extraction buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 20 mM CaCl$_2$, 5 % glycerol, Roche protease inhibitor cocktail). The suspension was lysed twice using an EmulsiFlex-C5 homogenizer (Avestin) and the protein extracted by incubation with 1 % (w/v) DDM for 1 h at 4 °C. Detergent-solubilised protein was cleared from cell debris by centrifugation at 35,000 g, 4°C, for 1 h and further purified by immobilized metal affinity chromatography by batch binding to Co^{2+} charged TALON resin (Clontech) at 4°C for 1h. The resin was washed with 15 column volumes of wash buffer (extraction buffer supplemented with 0.01 % DDM and 30 mM imidazole pH 8.0) and, in order to exchange the detergent, with another 15 column volumes of wash buffer containing 0.035 % UDM. PC2 was eluted with the same buffer supplemented with 400 mM imidazole and the buffer subsequently exchanged using a PD10 column (GE healthcare) pre-equilibrated in 0.035 % UDM, 20 mM MES pH 6, 150 mM NaCl, 20 mM CaCl$_2$, 5% glycerol. The protein was then treated with 0.4 units of α-L-Fucosidase from bovine kidney (Sigma) at 18°C, overnight. The pH was adjusted to 7 before further enzymatic treatment at a ratio of 0.75:0.5:1 (w:w:w) TEV

protease, PNGase F, PC2 for another 24h at 18ºC. The enzymatically treated protein was separated from the 6x His-tagged enzymes and uncleaved PC2 by reverse His-tag purification and the flow-through concentrated to 0.5 ml using a 100 kDa cut-off concentrator (Vivaspin 20, Sartorius), centrifuged for 30 min at 20,817 g and further purified by size exclusion chromatography (SEC) using a Superose 6 10/300GL column (GE Healthcare) equilibrated with SEC buffer (0.035 % UDM, 20 mM HEPES pH 7.5, 200 mM NaCl, 20 mM $CaCl_2$). The peak fractions containing PC2 were combined and concentrated in a 2.5 ml 100 kDa concentrator (Centrisart 1, Sartorius) to a final concentration of 16 to 17 mg/ml and centrifuged as described above before crystallisation. For grid preparation, fresh unconcentrated protein was used at 4.5 mg/ml. For functional studies, the protein was purified as described above with 0.01 % DDM after the initial lysis step, without the addition of $CaCl_2$.

Electron cryo-EM grid preparation and data acquisition.

Aliquots (3 µl) of deglycosylated protein purified in UDM at a concentration of 4.5 mg/ml were applied to glow-discharged holey carbon grids (2 µm hole size, 1 µm spacing, 200 mesh; Protochips Inc). Grids were blotted at 80-100% humidity for 3 s and plunge-frozen in liquid ethane cooled by liquid nitrogen using a vitrification apparatus (Vitrobot Mark IV, FEI). Cryo-EM was performed using a Tecnai F30 'Polara' transmission electron microscope (FEI) operated at 300 kV and at liquid nitrogen temperature. Data were acquired using SerialEM on a direct electron detector (K2 Summit, Gatan), mounted behind an energy filter (GIF Quantum LS, Gatan) and operated at zero-loss mode (0-20 eV). Movies (22 frames, 0.4 s each) were recorded in electron counting mode (dose rate 9 e- / pix / s) with underfocus in the range from 1.1 to 4.5 µm. The calibrated magnification used was 37,037× (corresponding to a pixel size of 1.35 Å) and total dose of 45 e-/$Å^2$ per movie stack; 2.05 e-/$Å^2$ per image in the stack.

Image Processing.

Frames in each movie stack were aligned using the whole-image motion correction implemented in MOTIONCORR 2.0[55]. Aligned frames were averaged to a single micrograph for further processing. Estimation of contrast transfer function parameters was carried out using CTFFIND4 4.0.8[56] and all other processing steps were performed with RELION 1.3[57], unless stated otherwise. An initial set of 11,094 particles were manually picked and subjected to 2D class averaging. The eight most representative 2D class averages were then used as

templates for reference based particle picking[57]. To minimize reference bias, templates were low-pass filtered to 20-Å resolution prior to carrying out the auto-picking procedure. A total of 183,974 particles were automatically picked from 735 micrographs and manually curated to a set of 161,965 particles. Further bad particles were removed using reference free 2D class averaging to yield a set of 150,882 particles (Supplementary Fig. 2b). For template-based 3D classification, a smooth initial 3D template with roughly the right dimensions was created from a single 2D class average (Supplementary Fig. 2c). Briefly, a class average image, representing a side view of the molecule was positioned as a slice in an empty volume box, averaged in cylindrical fashion in Bsoft[58] and low-pass filtered to 50-Å resolution. An initial 3D classification run with no symmetry imposed using the smooth 3D template clearly indicated that PC2 formed a symmetric tetramer. Further 3D classification with C4 symmetry imposed identified a subset of 104,149 particles that were subsequently used for a first 3D 'gold-standard' refinement where two half-sets of the data were refined independently. This resulted in an initial reconstruction at resolution of 4.7 Å. After correcting for beam-induced motion particle-by-particle basis and radiation damage ('particle polishing'), the same subset of particles resulted in a reconstruction at 4.4-Å resolution. A final 3D classification run was performed without further alignments and gave the final subset of 19,546 particles. This subset was used to refine the final reconstruction with a resolution of 4.22 Å. Reported resolutions are based on FSC=0.143 threshold[59] taking into account the contribution of the applied soft-edged mask by high-resolution phase randomization[60] (Supplementary Fig. 2d). Density maps were sharpened by applying a negative B-factor (-199 Å^2) that was estimated using RELION[59]. Local resolution in different parts of the map was estimated using RESMAP[61] (Supplementary Fig. 2f-g).

Model Building.

The 4.22 Å EM map was of sufficient quality to allow unambiguous tracing of 90% of the polypeptide chain. The main chain was built *de novo* by manually fitting helix and strand poly-alanine fragments into the map. Sidechain assignment was assisted by characteristic density for bulky hydrophobic residues (Supplementary Fig. 3). Sequence register was also guided by the position of Hg atoms determined using an 8.5 Å anomalous difference Fourier map calculated from mercury (EMTS)-soaked crystals (see below for details). Five Hg-labelled cysteine residues were observed in the TMD in S1 (Cys235), S2 (Cys476 & Cys481), S3 (Cys509) and S4L (Cys593) (Supplementary Fig. 3g). The remaining four cysteines (Cys331 / Cys344 / Cys437 / Cys632) were not labelled under the conditions used to soak the

crystals. Cys437 (TOP domain; β4 strand) and Cys632 (PH1) are completely buried and presumably inaccessible to EMTS. Cys331 and Cys344, whose Cα atoms lie within 6.5 Å of each other in the first arm of the TOP domain's TLC extension, could form a disulphide bond. To investigate this possibility we alkylated the protein using iodoacetamide, digested with a series of proteases and looked for peptides containing the relevant cysteines, either disulphide linked or with free cysteine using mass spectrometry. We detected a peptide with the sequence "EC[344]YDVYSVSSEDR", suggesting that the protein did not have a disulphide between Cys331 and Cys344. The maps do not suggest that Cys331 and Cys344 form a disulphide bond in this structure. The two open filter structures[35], differ at this site; one has the disulphide bond present and the other lacks a disulphide, so clearly this bond varies between conformations.

A series of sharpened maps (sharpening B-factor applied 0, -100, -150, -200 Å^2), low pass filtered to 4Å, were used during the model building process. All model building was performed using COOT[62]. The initial, manually built model was independently refined in real-space against the experimental density map with C4 symmetry imposed using the fragment-based refinement protocols implemented in ROSETTA[63]. The output models of a series of ROSETTA refinements were ranked and the best models were used as guides to improve the manually built model. The final completed tetrameric model was refined to 4.2 Å in reciprocal space using REFMAC-EM[64] against a map filtered to 4 Å and sharpened using a B-factor of -150 Å^2. Secondary structure restraints generated by ProSMART[65] along with automatically-determined local NCS restraints between the four protein chains were also used in refinement. No additional map sharpening was applied within REFMAC and appropriate electron scattering factors were used throughout. The R factor and overall Fourier shell correlation of the model to the map were monitored during refinement (Supplementary Table 1). The final model was validated using accepted procedures[66]. Briefly, each atom in the final model was subjected to a 0.2 Å independent random shift to remove model bias and then this model was refined against one of the half-maps output by the RELION 3D gold-standard auto-refinement step. These half-maps were filtered to 4 Å and sharpened by a B-factor of -150 Å^2 (i.e. treated as for the post-refined full map used for refinement by REFMAC as described above). The Fourier shell correlations (FSC) between the resultant refined model and either the half-map used for refinement (FSC$_{work}$) or the other half-map (no refinement; FSC$_{half2}$) were calculated. For comparison, the 'unbiased' model was also refined against the full map (FSC$_{full}$). The FSC curves were calculated by running REFMAC for zero cycles using the appropriate model and the REFMAC masked maps. The close agreement between

the FSC$_{work}$ and FSC$_{half2}$ curves (Supplementary Fig. 2e) indicates that the refinement procedure was not overfitting the atomic model.

Each chain of the final tetrameric model of PC2 encompasses the entire polypeptide chain between residues Arg213 and Asp702 apart from residues 298-303 which were poorly resolved in the map. This disordered loop region (EL2) connects α2 and the β1 strand in the TOP domain and contains two predicted N-glycosylation sites (Asn299/Asn305 – NAG1/2). In general, poor density was observed for the ends of the sidechains for Asp and Gly residues and these were not modelled unless sufficient density was observed. There is additional weak density in the unsharpened map at the C-terminus of S6 indicating some partial ordering of the remaining C-terminal residues (construct terminates at Asp723). However, the polypeptide chain could not be extended beyond Asp702 with any degree of confidence. The maps indicate that the S6B helix may continue for an additional two turns of α-helix to Lys710 (Supplementary Fig. 3h-k). Additionally, density is present aligned towards the pore axis between the extended S6B helices, consistent with a C-terminal two turn α-helix (Supplementary Fig. 3k). However it is not clear how these two elements should be connected and the sidechain density is insufficient to define the sequence register.

Although the protein used to prepare grids was enzymatically N-deglycosylated with α-L-Fucosidase and PNGaseF, residual oligosaccharides are visible in the unsharpened EM map (Supplementary Fig. 3f). Single N-acetyl glucosamine (NAG) units have been modelled attached to Asn328 (NAG3) and Asn362 (NAG4), while a dual NAG-NAG has been modelled on Asn375 (NAG5). The remaining two reported glycosylation sites, Asn299 and Asn305, are located in a poorly ordered loop region. Asn299 has not been modelled and density in the vicinity of Asn305 is not sufficient to model the glycosylation status at this site.

Crystallisation, X-ray crystallography and labelling of exposed cysteines with mercury. Protein was concentrated to 16-17 mg/ml. Crystals were grown at 20 ºC from sitting drops (250 nl + 250 nl) set up in 96-well format using a Mosquito crystallisation robot (TTP Labtech) and protein:reservoir ratios of 1:1. The reservoir solution contained 28 % (v/v) polyethylene glycol (PEG) 400, 1 mM methyl-β-cyclodextrin, 0.1 M sodium citrate pH 5.5, 0.15 M lithium sulphate and 0.05 M sodium chloride. Crystals were dehydrated for 80 min over a reservoir containing mother liquor supplemented with PEG 400 to a final concentration of 50% and cryo-cooled by plunging into liquid nitrogen. To validate the sequence assignment in the traced EM maps (see Supplementary Fig. 3), heavy atom labelling of accessible cysteines with mercury was performed using a final concentration of 5 mM

ethylmercurisalicylate (EMTS). 0.25 µl of mother liquor (20 mM HEPES, 7.5, 250 mM NaCl, 28 % (v/v) polyethylene glycol (PEG) 400, 0.1 M sodium citrate pH 5.5, 0.15 M lithium sulphate) containing 5mM EMTS and 0.035% UDM was added to (75 nl + 75 nl) drops containing crystals. The crystals were dehydrated for 4 h and cryocooled as described above.

Crystallographic data were collected at 100 K on either I04-1 (native) or the I03 (EMTS soak) beamlines (Diamond Light Source Ltd) with a fine phi slicing strategy and processed with XDS[67] and AIMLESS[68]. Native crystals diffracted anisotropically with diffraction spots to 4 Å in the best and around 4.6 Å in the worst directions. The crystals have unit cell dimensions of $a=b=128.87$ Å, $c=132.73$ Å, $\alpha\beta\gamma = 90°$, they belong to space group $P42_12$ and contain one monomer per asymmetric unit. Data were collected to 5.5 Å from a single EMTS-soaked crystal. The native dataset was initially phased by molecular replacement using PHASER[69] with the TRPV1 pore S5-S6 region as a search model (PDB: 3J5P)[30] before the EM structure was available, then subsequently phased with the coordinates of the EM model. Initial phase estimates were then used with the anomalous differences from the EMTS-soaked dataset to calculate an anomalous difference Fourier map to locate the bound Hg atoms. A total of five Hg atoms were localised per monomer in the TM helices (Supplementary Fig. 3g).

Preparation of PC2 reconstituted into proteoliposomes for activity assays

Liposomes were prepared with soybean lipids (Sigma Aldrich) to a final concentration of 20 mg/ml in buffer (20 mM HEPES pH 7.5, 200 mM NaCl). Lipids were first dried under argon and vacuum and rehydrated in buffer using a sonication bath for 1 h. 35 mM CHAPS was added and further sonicated until the solution became clear. Protein solution from size exclusion chromatography was added to the lipid solution, mixed well and dialysed in a Spectra-Por® Float-A-Lyzer® G2, MWCO 100kDa (Sigma) against dialysis buffer (20 mM HEPES pH 7.5, 200 mM NaCl) for 3 days with 6 changes of buffer at 4 ℃. The proteoliposomes were extruded 21 times through a 100 nm membrane, flash frozen in liquid nitrogen and stored at -80˚C. Control liposomes were prepared as described above. To mimic the addition of protein solution, the same volume of SEC buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 0.01 % DDM) was added before dialysis.

Electrophysiological data

Recombinant PC2$^{P185-D723}$ truncation construct was reconstituted into planar phosphatidylethanolamine lipid bilayers as previously described [70]. The *trans* chamber was

held at ground and the *cis* chamber was held at potentials relative to ground. Recording solutions contained 210 mM KCl, 20 mM HEPES, and 10 μM free Ca^{2+}, pH 7.2. The free $[Ca^{2+}]$ and pH of solutions were measured using a Ca^{2+} electrode (Orion 93-20) and Ross-type pH electrode (Orion 81-55) as described previously [70]. Single-channel events were recorded with a BC-525D amplifier (Warner Instruments, Hamden, USA), sampled at 20 KHz and low pass filtered at 800 Hz. All analysis was performed in pClamp10.6 (Molecular Devices, Foster City, CA). Po was measured over 3 min of continuous recording. Current amplitude measurements of sub-conducting states were obtained from amplitude histograms but full open and closed levels were determined using manually controlled cursors because transitions to these levels were infrequent in comparison.

No currents were recorded when control 'empty' liposomes were fused with the bilayer (7/7 independent experiments; Supplementary Fig. 4g). Representative single-channel fluctuations of the PC2$^{P185-D723}$ construct in symmetrical (210 mM KCl) ionic conditions and with an 810 mM KCl *cis:* 210 mM KCl trans gradient are illustrated in Supplementary Fig. 4. The full open channel conductance was 224 ± 8 pS (n = 5 independent experiments, mean ± s.e.m.) in symmetrical 210 mM KCl and increased to 443 ± 11 pS (n = 6 independent experiments, mean ± s.e.m.) in KCl gradient (810 mM KCl *cis*: 210 mM KCl *trans*) (Supplementary Fig. 4d). The reversal potential shifted to -30 ± 2 mV (mean ± s.e.m., n=6 independent experiments) in gradient, close to the reversal potential for K^+ (-33 mV) demonstrating that PC2$^{P185-D723}$ is selective for cations. The increase in PC2$^{P185-D723}$ single-channel conductance after applying the high *cis*: low *trans* ionic gradient was also reported to be a feature of the full length PC2 channel [20,71].

Single-channel analysis demonstrated that PC2$^{P185-D723}$ gates in a voltage-dependent manner, transitioning into sublevels approximately 90% of the time (see amplitude histograms to the right of traces). The channels often transitioned more frequently into one particular sub-conductance level for several seconds but then would spontaneously shift gating behavior so that another sub-level became the preferred gating state (as shown by ** in the expanded trace in Supplementary Fig. 4b). A frequently observed sub-conductance level is shown in Supplementary Fig. 4a, which is approximately one third of the full open current level. The voltage sensitivity of PC2$^{P185-D723}$ was not affected by applying the KCl gradient (Supplementary Fig. 4e, f). The channels dwelt in the full open state for longer at -40 mV than at +40 mV. At positive potentials, gating in the lower conductance sub-levels were favored (Supplementary Fig. 4b) whereas at negative potentials (Supplementary Fig. 4c), predominant sub-level transitions were to states of higher conductance, nearer to the full open state level.

Sub-conductance gating behavior has been previously reported for full length PC2 channels[20,27,71] indicating that this is an intrinsic gating feature of PC2. However, the distinctive rapid flickery nature of gating of the truncated protein, of which the inter-sub-conductance state gating constitutes 90% of all transitions, may have been caused by the deletion of regulatory domains. Flickery sub-state gating was reported to occur in the human TRPA1 channel after deletion of the first 688 residues[72]. In future experiments it will be important to examine how specific domains, such as the EF-hand or residues predicted to be glycosylated or phosphorylated influence the gating behavior and voltage-dependence of the channel. The $PC2^{P185-D723}$ construct also retains other ion selectivity and permeation properties reported for the full length protein, including block of K^+ current by Ca^{2+} [20,28] (Supplementary Fig. 4h,j) and block by amiloride[27,28] (Supplementary Fig. 4i,j). These results confirm that the truncated $PC2^{P185-D723}$ protein retains important functional features of the full length protein.

Statistical analyses

All values are the mean ± s.e.m. Student's t-test (two-sided) was used to assess differences between groups. $P<0.05$ was considered to be statistically significant.

Evaluation of ADPKD-associated missense variants.

A list of point variants was obtained from the PKD Mutation Database (pkdb.mayo.edu) and mapped onto the structure of PC2. Variants that are present in the NCBI SNP [48] and ExAC databases [49]were excluded from the comparison. Missense variants were classified based on their predicted structural effects (1 – minor, 5 – severe) of the amino acid substitution using a series of criteria including the nature of the mutation, its location in the structure and sequence conservation considerations (Supplementary Table 1). For comparison, Supplementary Table 1 also shows the variant scores for the Mayo clinic database (pkdb.mayo.edu). The scores are generally well-aligned. In the case of PC1, mutations were mapped onto a structural model of its channel-like domain that was generated for residues Arg3666-Glu4123 using MODELLER[73] based on the PC2 EM model and the alignment shown in Supplementary Fig. 1. The description of severity of changes relates only to the overall structure of the protein, it is not intended as a prediction of changes in protein activity and does not relate to the severity of ADPKD.

Preparation of figures.

Figures were prepared using PyMOL[74] and UCSF Chimera[75]. Pore radii (Fig. 4e, f) were calculated using the program HOLE[38] integrated within COOT[76]. Electrostatic surface potentials (Fig. 4a,b) were calculated using the APBS plugin within PyMOL and the PDB2PQR server[77]. Hydrogens and missing sidechain atoms were added automatically to the refined EM model using ICM-pro (Molsoft LLC) prior to electrostatic surface calculations. All electrostatic surface potentials were visualized in UCSF-Chimera and colored between -10 (red) and +10 (blue) kT/e⁻. Conservation analysis was performed using CONSURF[36] using an alignment of 177 PC2 sequences.

Data Availability Statement

The cryo-EM density map is available from the Electron Microscopy Data Bank (EMDB) under the accession code EMD-8200. In addition the RELION half maps and *B*-factor sharpened/filtered map used for model building are available as additional files in the EMD-8200 data bundle. The atomic coordinates of the PC2 structure is available from the Protein Data Bank (PDB) under the accession and code 5K47. Both the map and the structure have been available since the 24th of August, 2016. Requests for materials should be addressed to E.P.C. (liz.carpenter@sgc.ox.ac.uk) and J.T.H. (juha@strubi.ox.ac.uk).

Methods-only References

55   Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Meth* **10**, 584-590 (2013).

56   Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* **142**, 334-347 (2003).

57   Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J Struct Biol* **189**, 114-122 (2015).

58   Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. *J Struct Biol* **157**, 3-18 (2007).

59   Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333**, 721-745 (2003).

60   Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24-35 (2013).

61   Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-65 (2014).

62   Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486-501 (2010).

63   DiMaio, F. *et al.* Atomic-accuracy models from 4.5-A cryo-electron microscopy data with density-guided iterative local refinement. *Nat Methods* **12**, 361-365 (2015).

64   Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr D Biol Crystallogr* **71**, 136-153 (2015).

65   Nicholls, R. A., Long, F. & Murshudov, G. N. Low-resolution refinement tools in REFMAC5. *Acta Crystallogr D Biol Crystallogr* **68**, 404-417 (2012).

66   Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485-1489 (2014).

67   Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125-132 (2010).

68   Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72-82 (2006).

69   Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213-221 (2010).

70   Sitsapesan, R., Montgomery, R. A., MacLeod, K. T. & Williams, A. J. Sheep cardiac sarcoplasmic reticulum calcium-release channels: modification of conductance and gating by temperature. *J Physiol* **434**, 469-488 (1991).

71   Cantero Mdel, R. & Cantiello, H. F. Effect of lithium on the electrical properties of polycystin-2 (TRPP2). *Eur Biophys J* **40**, 1029-1042 (2011).

72   Moparthi, L. *et al.* Human TRPA1 is intrinsically cold- and chemosensitive with and without its N-terminal ankyrin repeat domain. *Proc Natl Acad Sci U S A* **111**, 16901-16906 (2014).

73   Eswar, N., Eramian, D., Webb, B., Shen, M. Y. & Sali, A. Protein structure modeling with MODELLER. *Methods Mol Biol* **426**, 145-159 (2008).

74   Schrodinger, LLC. *The JyMOL Molecular Graphics Development Component, Version 1.0* (2010).

75   Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).

76   Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).

77   Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**, W665-667 (2004).

Supplementary Figures
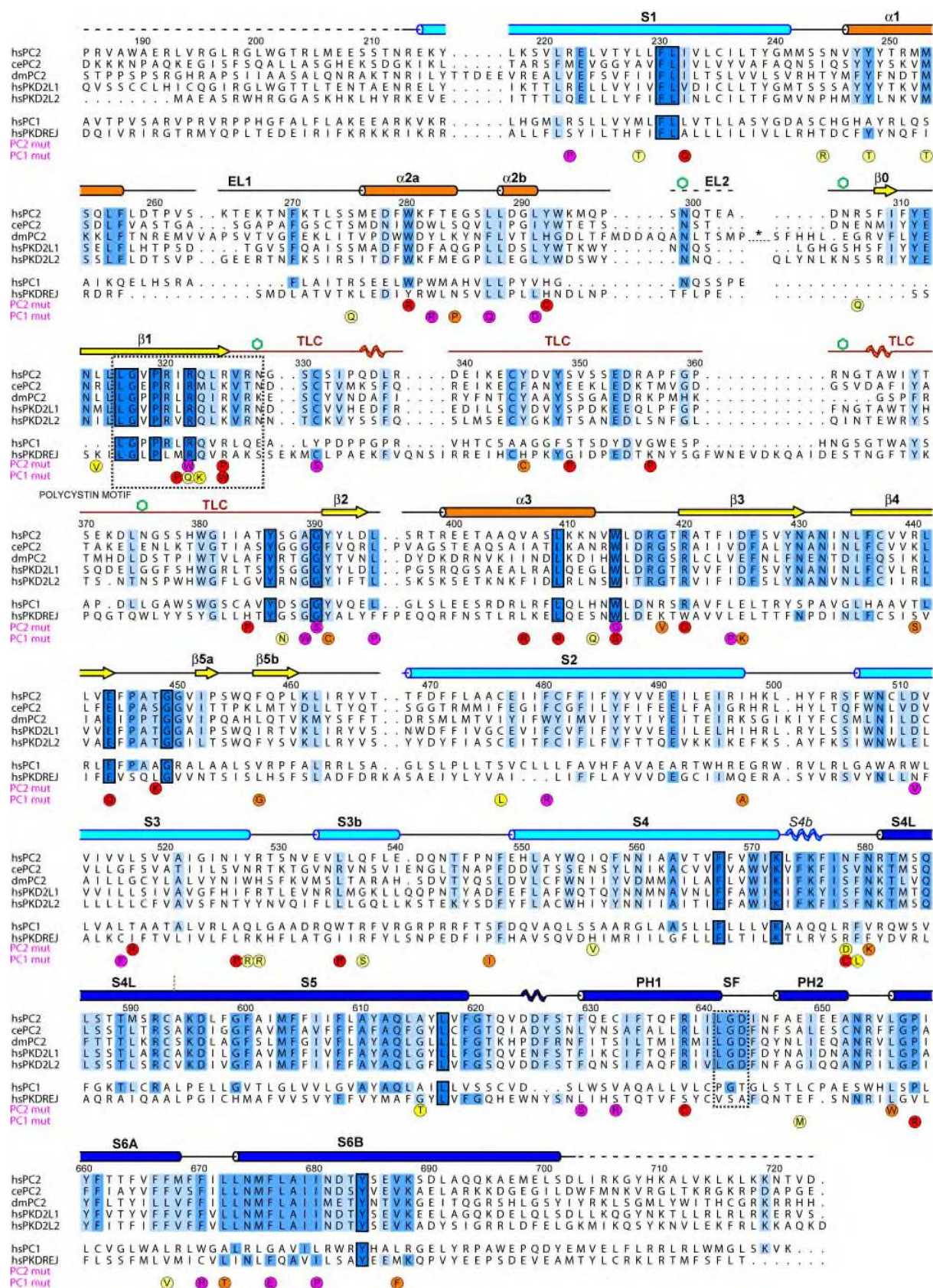
Supplementary Figure 1: Sequence alignment for PC2 and related orthologs.

The sequence of human PC2$^{P185-D723}$ (hsPC2; TRPP1) is shown along with the corresponding PC2 sequences from *C. elegans* (ce) and *D. melanogaster* (dm). The alignment also includes sequences for human PKD2L1 (TRPP2; aa 65-603), PKD2L2 (TRPP3; aa 1-534), PKD1 (aa 3635-4147) and PKDREJ (aa 1677-2208). Residues conserved in all sequences are boxed and colored dark blue. Secondary structural elements derived from the PC2 EM model (see Fig.2a and b) are shown above the alignment and colored by domain. Dashed regions represent residues that were not resolved in the EM model. Glycosylation sites in the human PC2 protein are highlighted by green hexagons. A selection of residues that are mutated in ADPKD patients in PC1 or PC2 are shown as colored circles under the alignment with the amino acid change indicated. The color of the circles indicates the predicted impact of the variant on the structure of the protein (colored according to classification in Supplementary Table 1; yellow: minimal predicted impact on the structure; magenta: serious predicted impact on the structure). Uniprot accession codes: hsPC2 (Q13563), cePC2 (Q9U1S7), dmPC2 (Q9VK95), hsPKD2L1 (Q9P0L9), hsPKD2L2 (Q9NZM6), hsPC1 (P98161) and hsPKDREJ (Q9NTG1). The asterisk (*) indicates the location of a large 90 amino acid insertion in the α2b-β0 loop in the fruit fly sequence (dmPC2, residues 328-418) that has been omitted for clarity
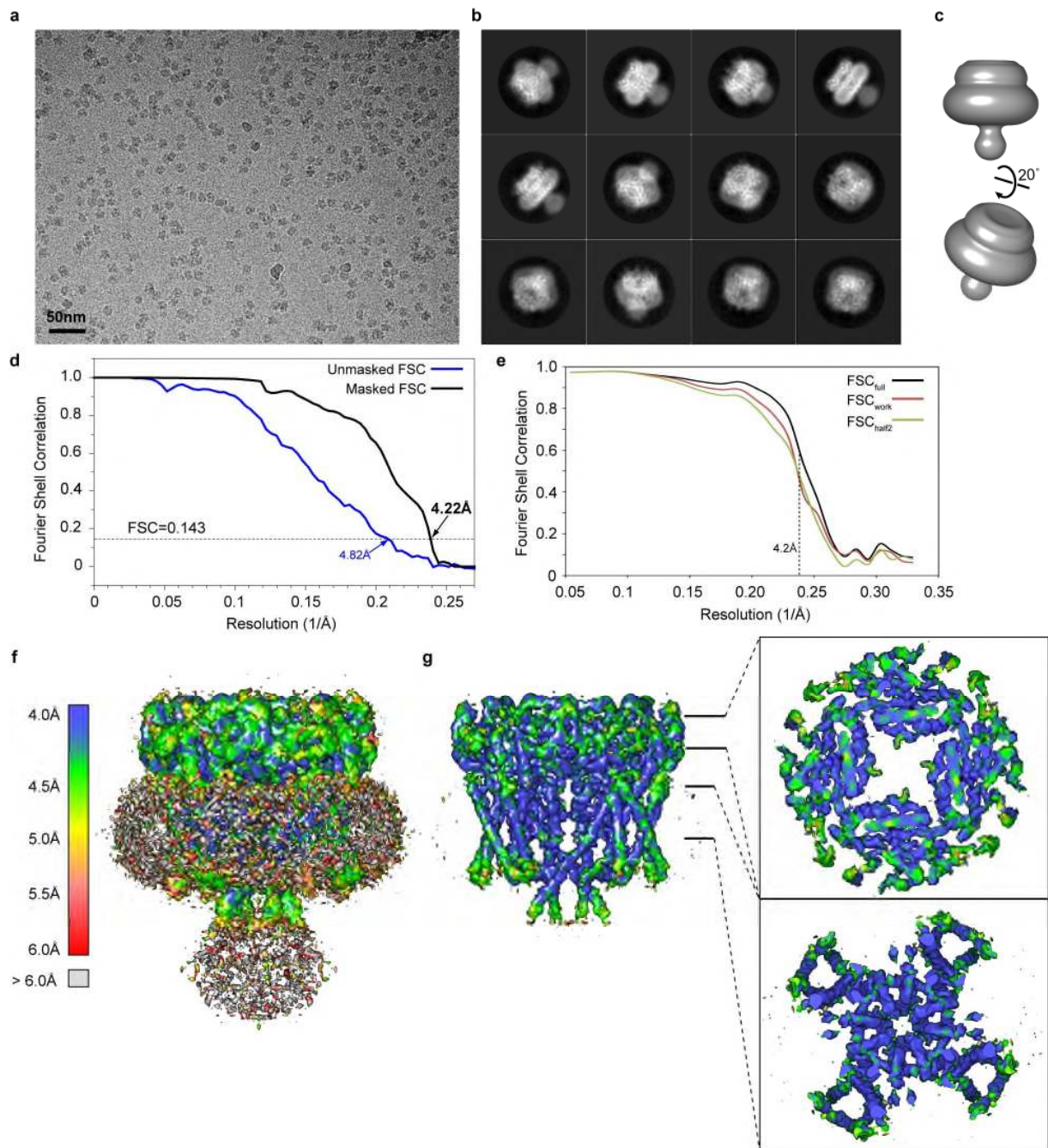
Supplementary Figure 2 Single particle cryo-EM of human PC2.
a) Representative raw particles from an original micrograph movie stack after initial frame alignment and summation. (b) Reference-free 2D class averages of detergent solubilised PC2 particles. (c) Initial 'spinning top' 3D model generated from a side-on 2D class average used for initial rounds of 3D classification. (d) Resolution estimation based on the gold standard FSC curve from RELION. The FSC curves for both unmasked and masked models calculated using RELION's automatic post-refinement masking procedure are shown. The overall resolution is calculated to be 4.22 Å based on FSC 0.143. (e) FSC curve of the final 4.2 Å REFMAC-EM refined model versus the map it was refined against (FSC$_{full}$, black line). For the purposes of validation, the (semi-randomised) model was refined against a reconstruction generated from half of the particles and then this refined model was compared to an independent reconstruction generated using the remaining particles. The FSC curves are shown for the unbiased model refined against the first of two independent (half) maps (to which it was refined

40

against; FSC$_{work}$; red line) or the same refined model versus the second independent half map (to which is was not refined; FSC$_{half2}$, green line). (f-g) Local resolution variation throughout the map estimated by RESMAP. Local resolution variation is color-coded from 4-6 Å. (f) Map contoured at low level to indicate the disordered detergent micelle ring covering the TM regions and the partially ordered C-terminal region. (g) Map contoured at higher level showing the PC2 tetramer viewed in the membrane plane (*left*) and two perpendicular slices through the density at the level of the TOP domain and around the pore. The central helices surrounding the pore and the core elements of the TOP domain exhibit resolutions around 4 Å. The cytosolic N- and C-termini as well as the S2-S3 and S4-S4L loop regions and the extremities of the TOP domain are less well ordered.
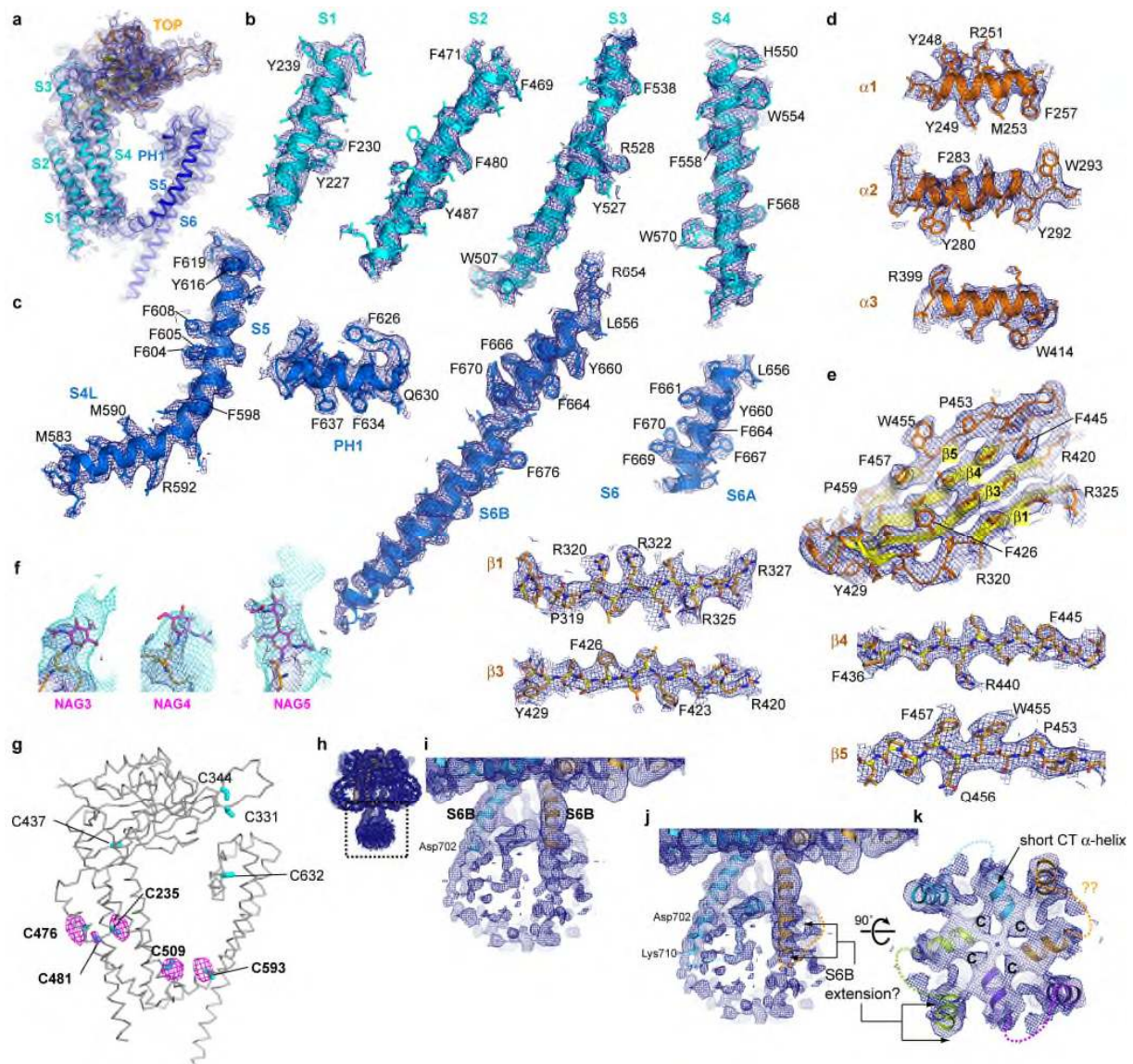
Supplementary Figure 3. Quality of EM density map.

EM density maps are shown for (a) the entire monomer, (b) 4 TM helices (S1-S4; cyan) of the VSLD, (c) helices (S4L, S5, PH, S6; blue) in the pore domain, (d) external helices (α1-3; orange) in the TOP domain, (e) central β sheet of TOP domain along with density for individual strands, (f) three glycosylation sites in the TLC motif of the TOP domain (NAG3-5). The EM density map shown corresponds to the final RELION reconstruction filtered at 4 Å and sharpened with a B-factor of −150 Å². This sharpened map was used also used as input to REFMAC-EM for model refinement in reciprocal space. The map (dark blue mesh) is contoured at 3.5σ. For the glycosylation sites shown in (f), the unsharpened EM map, contoured at 3.5σ is also shown (cyan mesh). Bulky amino acids that aided sequence assignment and register are highlighted. (g) 8.5 Å Anomalous difference Fourier map calculated from the crystallographic data collected from the EMTS-soaked crystals contoured at 3σ. The map (magenta mesh) is overlaid on the ribbon trace of a PC2 monomer with the positions of the nine cysteine residues highlighted. Hg peaks corresponding to all cysteines in the TM helices are observed. (h) Overall view of PC2 tetramer showing the unsharpened 4.2 Å EM density map. Dotted region indicates characteristic C-terminal 'bulb' at end of S6B helices. (i) Final refined model with overlaid unsharpened map contoured at low (4.3σ) level. (j) Possible extension of S6B helix has been

modelled along with an additional short C-terminal helix. Density for the helical portions is resolved but the connecting regions are not. (k) Density between S6, pointing towards the pore axis, is compatible with a short helix.
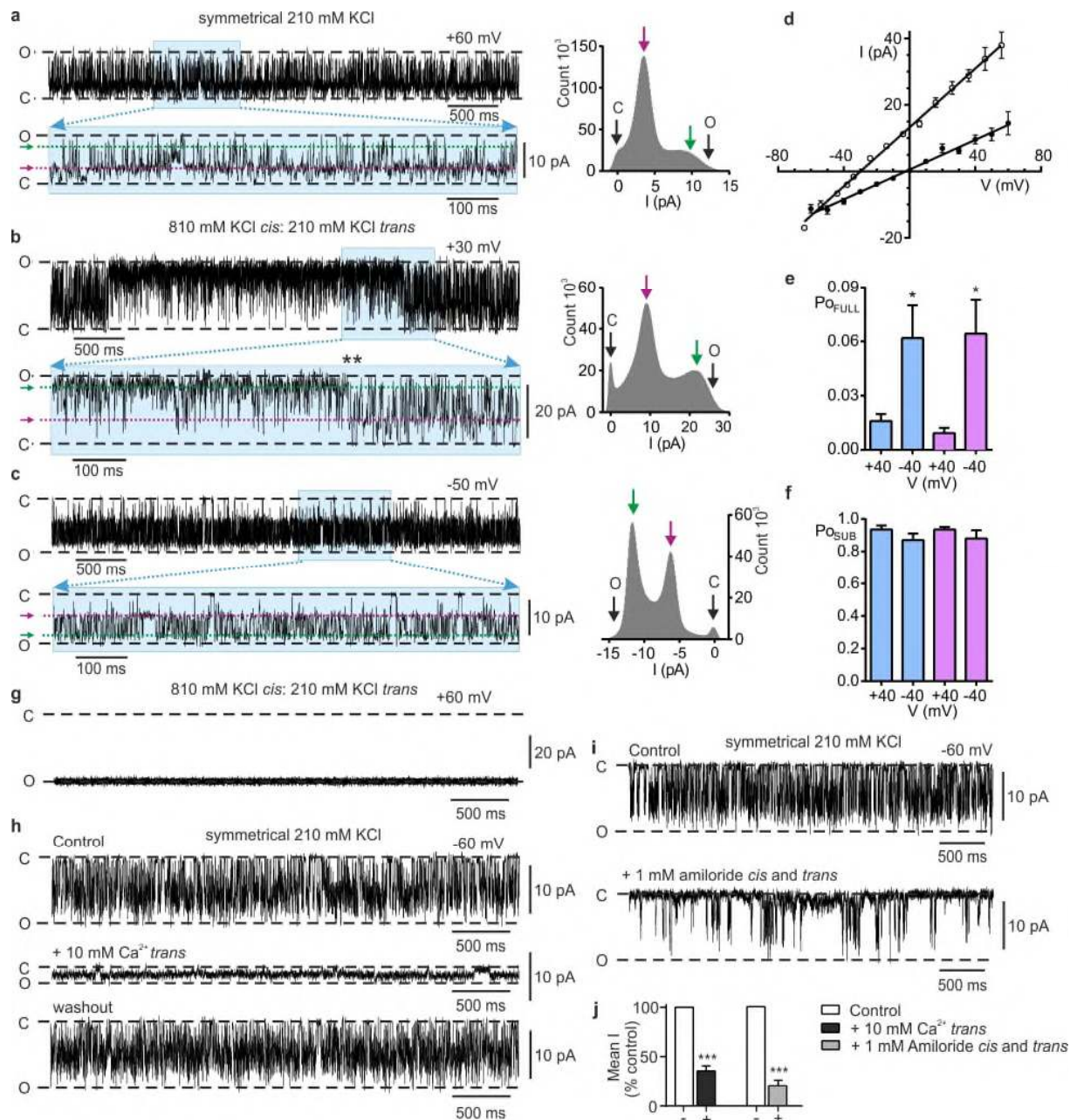
**Supplementary Figure 4. Single-channel properties of PC2$^{P185-D723}$.**

Representative single-channel currents of PC2$^{P185-D723}$ in (a) symmetrical 210 mM KCl at +60 mV and (b) in asymmetrical 810 mM (*cis*) and 210 mM (*trans*) KCl solutions at +30 mV and (c) -50 mV are shown. An expanded portion of each trace (*highlighted in blue*) illustrates the rapid gating between multiple sub-conductance levels and the sudden shifting of gating into different preferred sub-conductance states (shown by **). In each trace, the most frequently visited sub-conductance levels are highlighted with green and purple dotted lines and with arrows. The fully open (O) and closed (C) channel levels are indicated with black dashed lines. The all-points histograms shown on the right were obtained from 2 min of consecutive recording and the colored arrows correspond to the predominant sub-conductance levels identified in those traces. (d) Linear regression of the current-voltage relationships for PC2$^{P185-D723}$ in the same symmetrical KCl (closed circles, n=5 independent experiments) and gradient KCl (open circles, n=6 independent experiments) ionic conditions as in (a-c). Each data point represents the mean ± s.e.m. (e-f) Bar charts compare the probability of PC2^{P185-}

^{D723} gating to the full open state (Po$_{FULL}$) (e) and to sub-conductance open states (Po$_{SUB}$) (f) in symmetrical 210 mM KCl solutions (blue) and in KCl gradient (purple). Po$_{FULL}$ is significantly higher at negative holding potential irrespective of whether the channels were in symmetrical (n=5 independent experiments,*P <0.05) or gradient conditions (n=6 independent experiments,*P<0.05). Under all the conditions examined, PC2$^{P185-D723}$ channel exhibits high levels of activity preferentially dwelling in sub-conductance states (note the different scales on the y-axis in (e) and (f). Data shown are the mean values ± s.e.m. (g) Example of bilayer recording where control 'empty' liposomes were used. (h) Effect of Ca^{2+}: Typical control single-channel recordings in 210 mM KCl at -60 mV are shown (top trace). A reduction in current amplitude is observed after adding 10 mM *trans* Ca^{2+} (middle trace) which is reversed after washout of the Ca^{2+} (bottom trace). (i) Effect of amiloride: A representative control recording in symmetrical 210 mM KCl at -60 mV is shown in the top trace and the effects of adding symmetrical 1 mM amiloride to both *cis* and *trans* chambers is shown in the bottom trace. The majority of recordings were of multiple channels and therefore (j) shows bar charts illustrating the reduction in mean current (mean I) that was observed after addition of 10 mM *trans* Ca^{2+} (black bar; n=6 independent experiments;*** $P < 0.001$) or 1 mM symmetrical amiloride (grey bar; n=5 independent experiments;*** p < 0.001) to channels gating in symmetrical 210 mM KCl at +40 mV (controls, white bars). Data shown are the mean values ± s.e.m. (Student's t-test, *P<0.05; *** $P < 0.001$).
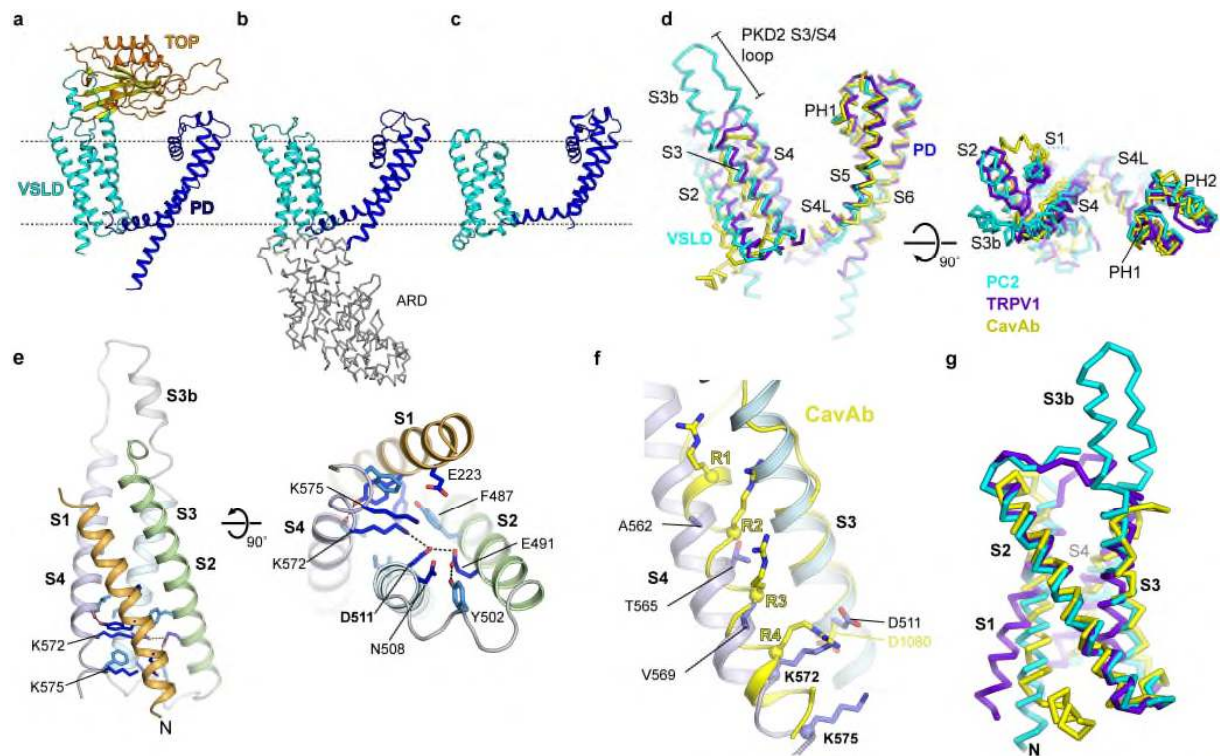
Supplementary Figure 5. PC2 fold, topology and VSLD structure.

Comparison of the monomer structure for (a) PC2; (b) TRPV1 (PDB: 3J5P) and (c) VGIC CavAb (PDB: 4MVO) viewed in the plane of the membrane (dotted line). Domains are colored according to domain (as Fig. 2a). The N-terminal Ankyrin repeat domain (ARD) of TRPV1 is colored grey. (d) Superposition of monomer structures of PC2 (cyan), TRPV1 (purple) and CavAb (yellow). The structures were superposed with the SSM matching routine in COOT using the S5-S6 pore domain of PC2 tetramer as a reference (to avoid bias from the differently positioned VSLDs). All 3 structures show similar relative orientations of the pore and VSDs. (e) Two perpendicular views of the VSLD (S1-S4). A cluster of conserved residues at the cytosolic end of the four-helix bundle are show in stick form. The conserved lysine residues on S4 (Lys572, Lys575) are in close proximity to a series of buried negatively charged residues. The sidechains of residues that could not be resolved in the EM maps (due to radiation damage/flexibility) have been modelled for illustrative purposes. (f) Comparison of the location of positively charged residues on PC2's S4 VSLD helix (light blue) with the S4 voltage sensor helix of NavAb / CavAb (yellow, PDB: 4MVO) which contains four arginines (R1-R4). (g) Superposition of the individual VSLD for PC2, TRPV1 and Ca$_V$Ab (colored as in (d)).
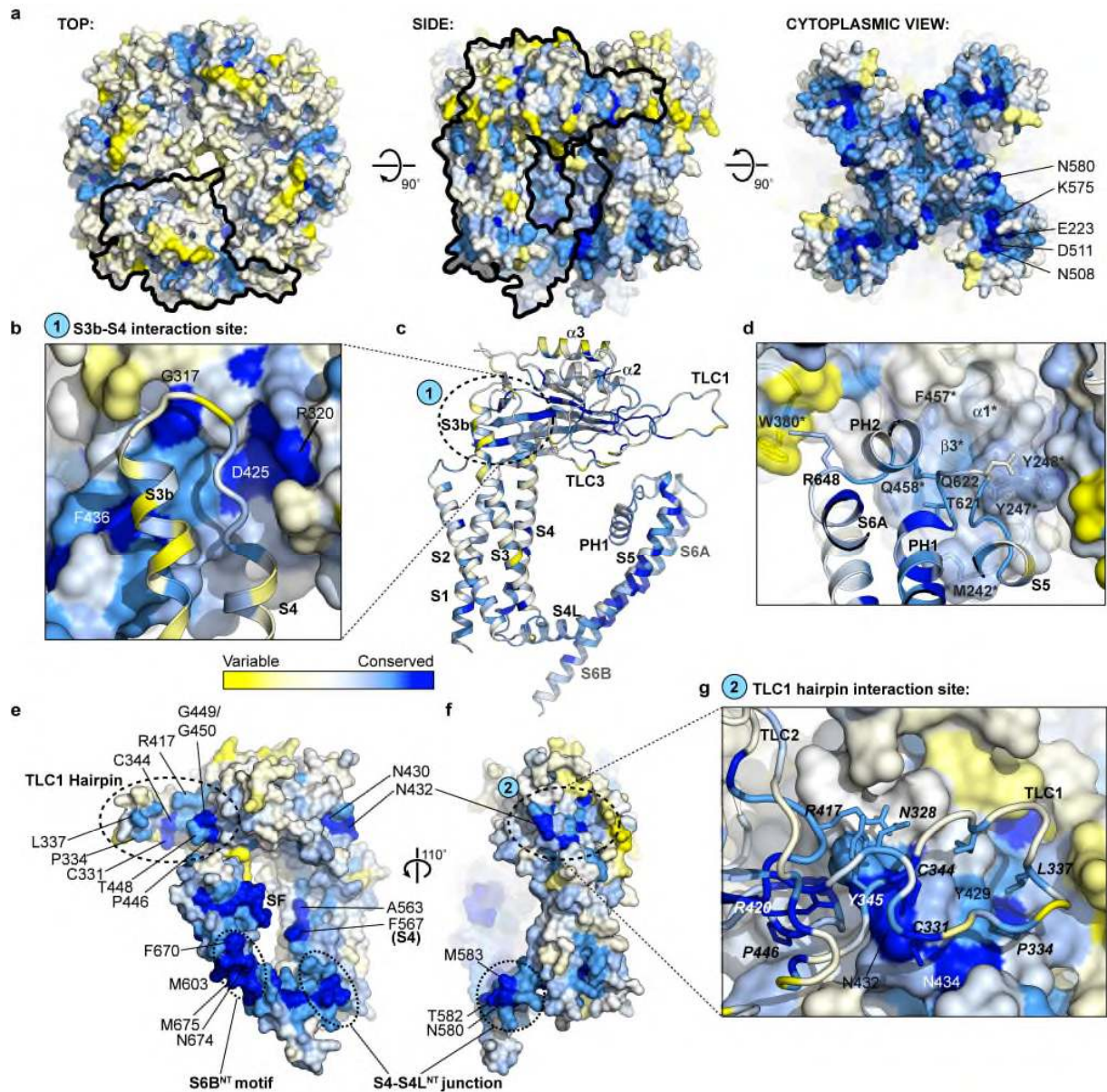
Supplementary Figure 6. Sequence conservation of residues in the interface between the TOP domains, the TLC and the S3-S4 extension.

(a) Three views of the PC2 tetramer (looking from top (external face), side (membrane plane) and bottom (cytoplasmic face)). The molecular surface is shown colored by sequence conservation (variable: yellow, highly conserved: dark blue) based on CONSURF analysis of an alignment of 177 PC2 homolog sequences. The external surface displays minimal sequence conservation apart from a small patch on each TLC extension involved in interdomain interaction. In contrast, the cytoplasmic face displays a high degree of conservation around the S6 bundle crossing and within the exposed face of the VSLD four-helix bundle. (b) Detail of the S3b-S4 interface. The VSLD extension is show in ribbon form and colored by sequence conservation. Residues contributing to the conserved interaction site on the TOP β-sheet are labelled. (c) Schematic cartoon ribbon for PC2 monomer colored by conservation viewed in the same orientation as the side view above in (a). The S3b-S4 interaction site with the outer face of the TOP β-sheet is highly conserved. The S3b-S4 interface shown in panel (b) is highlighted. (d) Interaction of the TOP domain with the pore turret region. The

pore turret region (ribbons) interacts with the α1 helix and β5 strand of the adjacent TOP domain (molecular surface; regions from the adjacent protomer are indicated by asterisks). Close contacts are made between the S5-PH loop including a direct hydrogen bond between Gln[622] and the backbone carbonyl of Phe[457] located in β5b. Residues in the turret region and the molecular surface from the adjacent TOP domain are colored by sequence conservation. A significant conserved patch maps to the TOP domain region that interacts with the pore turret. (e) Internal face view of PC2 monomer (180° rotation compared to panel (c)) highlighting the conserved nature of the TLC1 hairpin extension (dotted oval). (f) Perpendicular view corresponding to the approximate orientation of a neighbouring monomer in the tetramer. The surface that the TLC1 hairpin (shown in (e)) interacts with on the other end of the neighbouring TOP domain is indicated with dotted oval and labelled '2'. (g) Detail of the TLC1 hairpin extension interaction site.
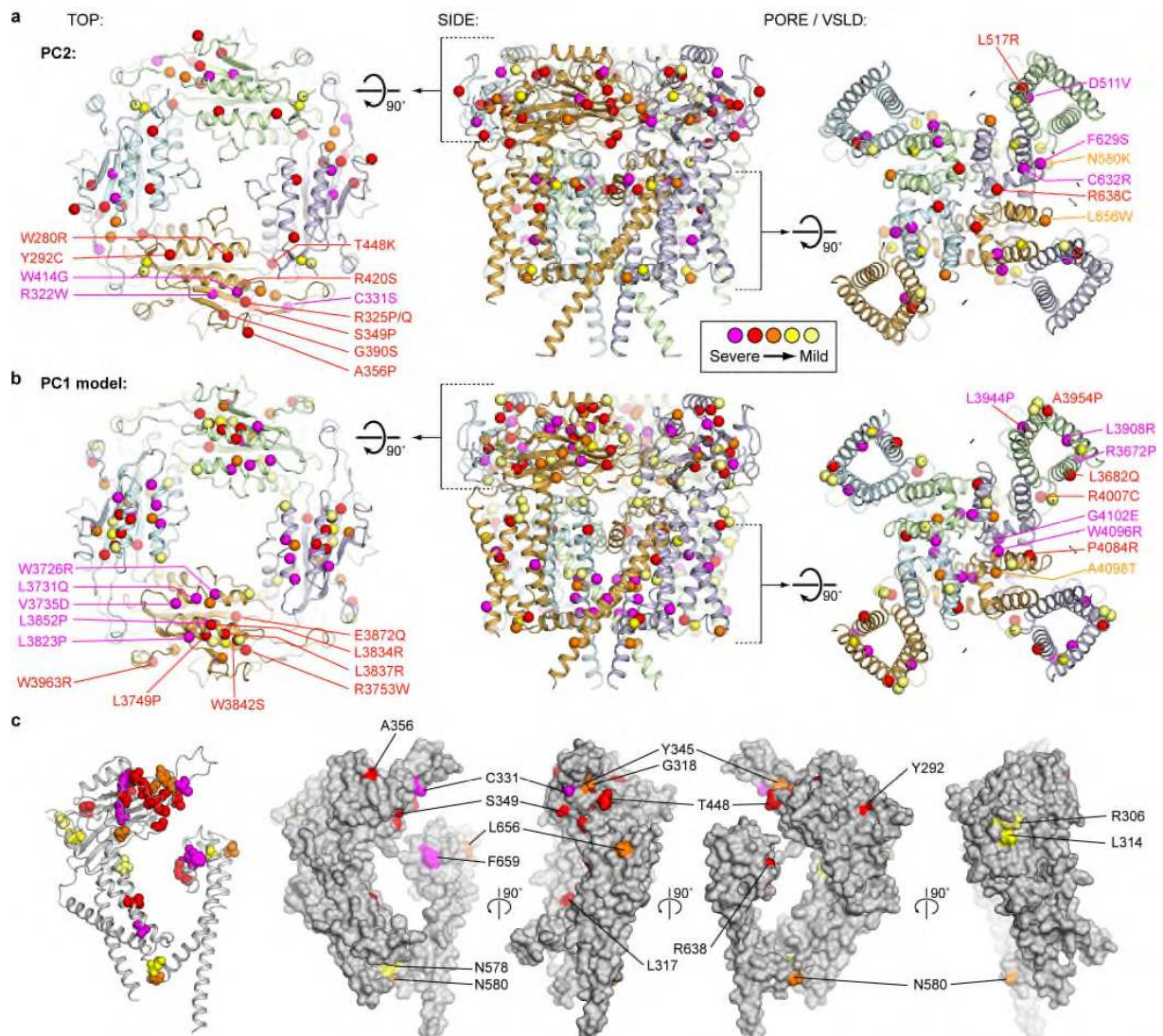
Supplementary Figure 7. Location of ADPKD associated missense variants in PC2 and PC1.

(a) PC2 tetramer shown looking onto TOP domain, from the side and a pore domain/VSLD sliced view. Positions of missense variant sites are highlighted by spheres colored according to the predicted impact of a mutation on the channel structure as indicated in Supplementary Table 1 (low – yellow; high – magenta). Variant sites cluster in the TOP domain and around the top of the pore. Fewer variants lie in the VSLD. (b) ADPKD-associated missense variants of PC1 mapped to a model of PC1 channel-like domain derived from the PC2 structure. It is unclear whether PC1 forms a homotetramer or a heterotetramer with PC2 or other polycystins. As with PC2, missense variants in PC1 cluster in the TOP domain. (c) PC2 variant sites mapped on to either a PC2 monomer cartoon representation (left panel) or molecular surface representation. Only a small percentage of the missense variant sites are exposed on the protein surface with the rest buried. The most closely packed cluster maps to the TOP domain TLC extension arm that participates in TOP-TOP interaction (Gly318/Cys331/Tyr345/Thr448).

Supplementary Table 1 | Summary of ADPKD variants in PC2$^{P185-D723}$ and their likely effects on the stability, structure or function of the channel domain.

VARIANT	DOMAIN	CONS (%)[1]	STRUCTURAL EFFECT	STRUCTURAL IMPACT SCORE (1-5)[2]	PKD MUTATION DATABASE VARIANT SCORE[3]
W280R	TOP	96 (0)	Located in α2a and packs closely against α3 C-terminal end	4	13
Y292C	TOP	85 (0)	Located at C-terminal end of α2b and packs against α3	4	11
R306Q	TOP	23 (11)	Located at TOP-TOP interface. Interacts with neighbouring TLC1 extension (Glu^{340})	1	-5
L314V	TOP	63 (3)	Located in β1 strand (polycystin motif) at interface between TOP β-sheet and S3b-S4 extension	2	-6
R322W	TOP	99 (0.5)	Located in β1 strand (polycystin motif). Buried in TOP hydrophobic core	5	36
R325P	TOP	45 (0)	Located in β1 strand (polycystin motif). Packs against TLC2 and α3-β3 loop	4	11
R325Q	TOP	45 (0)	Located in β1 strand (polycystin motif). Packs against TLC2 and α3-β3 loop	4	11
C331S	TOP	100 (0)	Strictly conserved. Located in TLC1 hairpin. May form a disulphide bond with Cys^{344} to stabilise TLC1 hairpin arm (involved in TOP-TOP interaction)	5	13
Y345C	TOP	83 (0)	In TLC1 extension and packs against β3 strand N-terminus	3	11
S349P	TOP	79 (0)	Located at TLC1/2 junction	4	6
A356P	TOP	53 (0)	Located in TLC2 hairpin	4	6
A384P	TOP	62 (0)	Located in TLC3 hairpin that lines lateral fenestration between TOP and PD	4	2
G390S	TOP	100 (0)	Strictly conserved. Located at N-terminus of β2 strand and packs against TLC extension	5	10
W414G	TOP	99 (0)	Anchors C-terminal end of α3 into hydrophobic core of TOP domain	5	18
G418V	TOP	87 (0)	Located in α3-β3 loop close to TOP-TOP interface	3	10
R420G	TOP	98 (0)	Located at N-terminus of β3 strand at TOP-TOP interface	4	16
R440S	TOP	66 (3)	Located on the β4 strand at interface between TOP β-sheet and S3b-S4 extension	3	9
T448K	TOP	88 (0)	Located in the β4-β5 loop at TOP-TOP interface. Backbone carbonyl hydrogen bonds to neighbouring TOP domain β3-β4 loop	4	5
D511V	VSLD	92 (0)	Located in the S3 helix. Participates in hydrogen-bonded network at cytosolic end of VSLD four-helix bundle. Interacts with Lys^{572} on S4	5	34
L517R	VSLD	77 (3)	Located halfway up S3. Sidechain projects into lipid bilayer and also packs against S4	4	7
I556V	VSLD	31 (8)	Located at N-terminal end of S4 at the membrane interface	1	-8
N578D	VSLD	47 (0.6)	Located at C-terminal end of S4	2	-5
N580K	PD	92 (0)	Located in S4-S4L loop	3	9
A615T	PD	53 (0.6)	Located at C-terminal end of S5. Buried and interacts with pore helix	2	1
F629S	PD	81 (0)	Located at N-terminal end of PH1. Sidechain projects into lipid bilayer	5	8
C632R	PD	57 (0)	Located in PH1. Sidechain is buried between pore helix and S5	5	9
R638C	PD	93 (0)	Located in PH1. Sidechain adjacent to Asp^{643} close to pore entrance and projects into upper vestibule	4	11
L656W	PD	85 (0)	Located at N-terminal end of S6. Sidechain packs against S5 and S4$^{\#}$. Listed in SNP database but retained for this analysis	3	-2

[1] Degree of Conservation at missense mutation site. Percentage occurrence of amino acid at its position in an alignment of 177 PC2 sequences (% occurrence of mutant amino acid at same position; Reflects likelihood that mutant amino acid can be accommodated at this position)

[2] Predicted impact of a mutation on the structure: this score is based on role in maintaining structure, sequence conservation, nature of mutation (whether change could be accommodated / (non-) conservative nature of mutation. 1=low impact, 5=high impact. Coloring used in Fig. 6, Supplementary Figs. 1 and 7.

[3] Calculated variant score from PKDB http://pkdb.pkdcure.org – Highly likely pathogenic>=11; Likely pathogenic between 5-10; Indeterminate variants between 4 and -4; Neutral < -5

SUPPLEMENTARY NOTE

Extended discussion of PC2 disease-associated variants

Sequencing of the *PKD1* and *PKD2* genes from patients with ADPKD has identified diverse changes throughout both genes (there are 278 variants in PC2 and 2323 in PC1 in the ADPKD database at the Mayo Clinic: http://pkdb.mayo.edu/) and generally individual changes have been identified in one or a small number of families. It is likely that many more deleterious mutations will be identified as genes are sequenced from more patients with ADPKD in the future, so the examples given here are likely to be a small subset of all possible disease causing ADPKD mutations. Many of these mutations cause frameshifts, insertions, deletions or truncations, suggesting that loss of polycystin function is the cause of ADPKD. There are 67 germline missense mutations reported in PC2, 41 (59%) of which lie within the channel and TOP domains. We will focus on 27 variants that are likely to be associated with disease. The other twelve PC2 variants may be polymorphisms, not associated with disease, as they are found in the single nucleotide polymorphism databases NCBI SNP[48] and ExAC[49] databases). These databases include sequences for more than 60,000 individuals and the clinical status of patients is not reported in the database. ADPKD has a prevalence of approximately 1:800, so approximately 75 of the individuals would be expected to have the disease, but many of these individuals would have changes in PC1, or PC2 truncations, deletions or insertions, so it is statistically unlikely that many of the missense variants found in the ExAC database would be associated with ADPKD disease. For our discussion of variants, we included only those that occur in the ADPKD database, but not in the ExAC or SNPDB databases. In addition to the missense changes, there are three insertions and five deletions that do not cause frameshifts and three of these lie within the channel and TOP domains. Similarly, for PC1 there are 327 mutations and variants in the channel-like domain, including 53 single residue substitutions that are found in patients, but not in the general population.

Many of the TOP domain missense variants involve residues that are buried between the structural elements that form the core of the domain and are therefore likely to disrupt the structural integrity of this domain (Fig. 6b; Supplementary Fig. 7c). They include a 7 residue insertion after Gln456 in the middle of the β5 strand; changes to sidechains packed between the β-sheet and the three α-helices (Trp280Arg, Tyr292Cys and Trp414Gly); mutations on the lower surface of the β-sheet (Arg322Trp and Arg440Ser); and mutations that lie on the TLC β-hairpin that packs against the β-sheet (Gly390Ser and Ser378Del). Many of these mutations

involve substantial changes in the size and charge of the sidechains, thus disrupting the TOP domain fold. Arg322 is part of the highly conserved "polycystin motif". The guanidinium moieties of the Arg322 and Arg440 sidechains are 4 Å apart and they will interact with the adjacent acidic sidechains of Asp425 and Glu559. They link together β1 and β4, and interact in addition with Phe423 on β3. In addition Arg440 is adjacent to His379 and the lipid binding site and could stabilize the lipid head groups (see Wilkes et al., for details of these interactions). Trp280, Tyr292, Leu415 and the aliphatic part of the sidechain of Lys410 form a hydrophobic cluster. The Trp280Arg mutation would disrupt the hydrophobic cluster and would displace Lys410, which stabilizes the interface between the α2 and α3 helices, through salt bridges with Glu277 and Glu385.

The Trp414Gly mutation is intriguing as it shows differing effects on channel activity and localisation, depending on the experimental setup. The Trp414Gly mutation, when combined with a Phe604Pro gain-of-function (GOF) mutation, it is stably expressed and trafficked to the plasma membrane of oocytes, but has reduced channel activity[28]. However without the GOF mutation, the Trp414Gly mutation has been reported to prevent trafficking to the cilium in pig kidney epithelial cells, while maintaining channel activity close to WT in the ER[50]. Therefore this mutation can affect either channel activity or trafficking in different experimental systems. Trp414 is packed within a cluster of hydrophobic residues (Phe279, Ile321, Leu393, Leu409, Leu415) that anchors the C-terminal end of α3 to the central β-sheet. So loss of this sidechain would destabilize this hydrophobic cluster and the interactions that keep α3 in place.

Missense variants at the interface between the TOP domain and S4 are also likely to be deleterious. Ser378 lies near the interface between the TOP domain and S4, so deletion would also disrupt the TOP domain-S4 interface.

The interface between adjacent TOP domains is clearly important as it is conserved and has thirteen ADPKD missense variants (Fig. 6c, Supplementary Figs. 6, 7). They include seven changes that map to one side of the TOP-TOP interface (Arg325Pro or Arg325Gln, Tyr345Cys, Gly418Val, Arg420Gly and Trp414Gly, Thr448Lys), two that map to the opposite side, below the TLC arm (Arg306Gln, Leu314Val) and four that map to the TLC extension (Cys331Ser, Ser349Pro, Ala356Pro and Ala384Pro) (Figs. 6b, c). Arg306 is at a critical position; it interacts with both the end of the TLC1 extension of the adjacent TOP domain and also the top of the S3-S4 extension, which is linked to the VSLD of the same chain. In addition it is adjacent to the Asn305 residue, which has variable glycosylation and acts as a switch between two states[35]. Interestingly, the Arg420Gly mutation, at the centre of

the interface, causes loss of channel activity (in the GOF site-directed mutation background[28]), whereas the Arg325Pro variant, which is nearer the surface of the protein, does not lead to loss of activity, and may instead be involved in trafficking. Tyr345 seems to be a key residue; it lies at the intersection of the three TLC loops and the TOP-TOP interface. Gly418 packs against Tyr345, such that any mutation of the glycine would displace the tyrosine sidechain. In addition Gly418 is packed between two arginines (Arg325 and Arg420), which are also mutated in some patients with ADPKD. The adjacent β4-β5 loop is the site of the Thr448Lys variant, so each of the loops at this end of the β-sheet has at least one ADPKD associated missense variant. The TLC extension harbors four variant sites (Cys331Ser, Ser349Pro, Ala356Pro and Ala384Pro) (Fig. 6b). Cys331 is one of a pair of cysteines that lie near each other at the start and end of the first TLC β-hairpin in the TOP domain. The EM maps suggest that in this closed conformation there is no disulphide bond at this site (see Online Methods for further discussion). The open conformations[35] suggest that the disulphide bond is found only in one of the two states. The other mutations that lie on the TLC domain are Ser349Pro, Ala356Pro and Ala384Pro; they are in the sections that cover the end and the outer surface of the β-sheet.

There are four ADPKD-associated missense variants on the VSLD and six that lie in the pore domain (Supplementary Table 1). Asp511, the site of the well-studied Asp511Val channel-inactivating mutation[17], lies on S3 at the cytoplasmic face of the four-helix bundle, with its sidechain facing into the space between the S1-S4 helices, positioned to form a hydrogen bonding network involving Lys572 and Lys575, the two potential voltage sensing residues, and Tyr487, Glu491 and Asn508 (Fig. 6d). Further up S3 the Leu517Arg mutation would insert a destabilising charged residue into the centre of the membrane.

Six potentially pathogenic missense variants and one deletion variant lie directly within the pore domain: there are potentially pathogenic variants at the junction of S4 and S4L (Asn580Lys); half way along S5 (where Phe605 is deleted); next to the PH1 (Ala615Thr), on the PH1 itself (Phe629Ser, Cys632Arg, Arg638Cys) or at the N-terminal end of S6 (Leu656Trp). The ΔPhe605 mutation on S5 would cause a rotation of the S5 helix, moving the sidechains and therefore completely disrupting the structure. Phe605 is adjacent to the GOF Phe604Pro site directed mutation site which is described in the main text. There are also a series of mutation sites on the PH1, involving significant changes in the size of the sidechain (Phe629Ser, Leu656Trp) and in some cases also the charge (Cys632Arg, Arg638Cys). These large changes in the vicinity of the pore are likely to disrupt the pore architecture and thus affect ion flux (Fig. 6e).