Supplementary Material

Identification and Expression Analysis of the Complete Family of Zebrasish *pkd* Genes

Samantha J. England, Paul C. Campbell, Santanu Banerjee, Annika J. Swanson, and Katharine E. Lewis*

*Correspondence:* Professor K. E. Lewis. Email: kelewi02@syr.edu
1 Supplementary Figures and Tables

1.1 Supplementary Figures

Supplementary Figure 1: Alignment of polycystin-cation-channel domain from different vertebrate PKD1-like proteins.

Comparison of the amino acid sequences for the polycystin-cation-channel domains of human (Homo sapiens, hsa), mouse (Mus musculus, mmu), zebrafish (Danio rerio, dre), medaka (Oryzias latipes, ola), green spotted pufferfish (Tetraodon nigroviridis, tni), stickleback (Gasterosteus aculeatus, gac), spotted gar (Lepisosteus oculatus, loc) and elephant shark (Callorhinchus milii, cmi) PKD1-like proteins using Clustal Omega (version 1.2.3). This figure shows the region of the polycystin-cation-channel domain that is present in all of the proteins and was used to construct the phylogenetic tree in Fig. 3A. Different families of PKD1-like proteins are color-coded with the same color. This same color-coding is used in Figure 3. Numbers on either side of each sequence, indicate amino acid positions in the full-length sequences of each protein.
Supplementary Figure 2: Alignment of polycystin-cation-channel domain from different PKD2-like proteins.

Comparison of the amino acid sequences for the polycystin-cation-channel domains of human (Homo sapiens, hsa), mouse (Mus musculus, mmu), zebrafish (Danio rerio, dre), medaka (Oryzias latipes, ola), green spotted pufferfish (Tetraodon nigroviridis, tni), stickleback (Gasterosteus aculeatus, loc), spotted gar (Lepisosteus oculatus, loc), elephant shark (Callorhinchus milii, cmi) and Drosophila melanogaster (dme) PKD2-like proteins using Clustal Omega (version 1.2.3). This figure shows the region of the polycystin-cation-channel domain that is present in all of the proteins and was used to construct the phylogenetic tree in Fig. 3B. Different families of PKD2-like proteins are color-coded with the same color. This same color-coding is used in Figure 3B. Numbers on either side of each sequence, indicate amino acid positions in the full-length sequences of each protein.
Supplementary Figure 3: Spinal cord expression of zebrafish *pkd* genes

Lateral views of additional stages of *pkd* spinal cord expression not included in other figures. Rostral left, dorsal up. By 36 h, *pkd1b* is expressed in two ventral rows of cells in the spinal cord (A). This expression persists at 4 dpf (B). *pkd1l2a* and *pkd2l1* are co-expressed in KA cells (C-F) and occasional weak more dorsal cells (asterisks D & E). *pkd1, pkd1l1, pkd1l2b* and *pkd2* are not expressed in spinal cord (G-Z). Scale bar (A) = 50 µm.
### Supplementary Tables

| Primer Name        | Forward Primer Sequence | Reverse Primer Sequence | Annealing Temperature \(^\circ\text{C}\) | PCR Extension Time (Seconds) | Transcript Exons Mapped                   |
|-------------------|-------------------------|-------------------------|---------------------------------|-----------------------------|------------------------------------------|
| pkd1\_Map\_ Nested Set 1 | GTGCAGCACCAACTCTGAG   | CATGAAAAACACTTGAGTGCAC  | 62.0                           | 100                         | 1-595 bp of transcript\(\wedge\)        |
| pkd1\_Map\_ Nested Set 2 | CACCAACTCTGAGCCCAATCAC | GAGGACGAGAAGATGACCAGC  | 68.0                           | 100                         | 1-595 bp of transcript\(\wedge\)        |
| pkd1\_Map\_ Set 3  | AGGAGCTAGATCTGAGCAACAAC | CTCCTCTGCCATCCTTACTGAC  | 66.0                           | 60                          | 527-1143 bp of transcript\(\wedge\)     |
| pkd1\_Map\_ Set 4  | ACGTGTGTGTGTCCTCGAC   | GGATCCATTAACCTCCTAAGGGAAC TGTACTCTGGTATTTGTGC | 61.0 | 60 | 1575-2154 bp of transcript\(\wedge\) |
| pkd1\_Map\_ Set 5  | ATTCAGAAAGGGCTCAACTGAC | CAGAGATTGTGAGGATTGTGAC  | 69.0                           | 60                          | 2122-2920 bp of transcript\(\wedge\)    |
| pkd1\_Map\_ Set 6  | AAGGAGCAAGAAAGGAAGAAAG | CAAACGGGACTTCATGAGATCTG | 67.0                           | 60                          | 13121-14065 bp of transcript\(\wedge\)  |
| pkd1\_Map\_ Set 7  | CAATCCTCATGAAGTCCCCCTTGG | CTGACATACAATCGCCACAGAAC | 68.0                           | 60                          | 14102-15014 bp of transcript\(\wedge\)  |
| pkd1\_Map\_ Set 8  | TATCTGTGGGGCTGTCTGATG | GTGACTCTACTACAACACAACCTC | 66.0                           | 60                          | 15527-16413 bp of transcript\(\wedge\)  |
| pkd1\_Map\_ Set 9  | GTGCAGTGATCGACTAGTATGC | TTCTGAGGAAGCACAAGCTCTC | 65.0                           | 60                          | 17453-18305 bp of transcript\(\wedge\)  |
| pkd1\_Map\_ Set 10 | GTCTGTGAATGCGATGCACTTC | CTAAGAGCCGGCCACGACTCAACC | 69.0                           | 60                          | pkd1\_Map\_ Set 10                      |
| pkd1\_Map\_ Set 11 | CTCTGTTGTTGCTGTTTCTGATG | GTGACTCTACTACAACACAACCTC | 66.0                           | 60                          | pkd1\_Map\_ Set 11                      |
| pkd1\_Map\_ Set 12 | GCTGATCATAATGCTACGCTTC | CATCATCTGGTGTGTTCTGGAG | 69.0                           | 60                          | pkd1\_Map\_ Set 12                      |
| pkd1\_Map\_ Set 13 | TCTGAATGAGGCTGAGGCAG  | CTGACTTTCGCTGATGACAGGCAC | 67.0 | 100 | pkd1\_Map\_ Set 13                      |
| pkd1\_Map\_ Set 14 | GTCTGCTATGAGTTGCTGAGTGC | CTGACTTTCGCTGATGACAGGCAC | 67.0 | 100 | pkd1\_Map\_ Set 14                      |
### Supplementary Table 1. PCR Primers for mapping *pkd1, pkd1l2a* and *pkd1l2b* mRNA transcripts

PCR primers and conditions used to amplify and sequence overlapping fragments of zebrafish *pkd1, pkd1l2a* and *pkd1l2b* transcripts (genes for which annotations in Ensembl are incomplete). Column 1 lists names given to each primer set. Columns 2 and 3 list PCR primers used to map transcript region indicated in column 6. Columns 4 and 5 list annealing temperatures and extension times of PCR protocols used in each case. The rest of the protocol is provided in materials and methods. *These primers were used on a circularized cDNA product following inverse PCR to generate 5’ sequence. *These primers were used in a series of nested PCR reactions. Nested Set 1 primers were used on a circularized cDNA product following inverse PCR to generate 5’ sequence. Nested Set 2 primers were used in reactions seeded with the Nested Set 1 PCR product. See materials and methods for details. § These reverse primers contain sequence for T3 RNA Polymerase.

| Primer Set | PCR forward primer | PCR reverse primer | Annealing Temp | Extension Time |
|------------|--------------------|--------------------|----------------|---------------|
| *pkd1l2a* _Map_ Set 7 | CAAATGAAAGAGGACAGCAGCTC | AAAGAACATCAGAGCCCAGAGAG | 68.0 | 60 |
| *pkd1l2a* _Map_ Set 8 | TGTAACGCTCCCTATCCTTCTTC | AATTAACCCTCCTAAAGGGAGCGCCTA | 67.0 | 100 |
| *pkd1l2a* _Map_ Set 9 | CTTCAGCAGCTGTGATCAAC | GAGTACAAGAGCAAGAAGCTGC | 65.0 | 100 |
| *pkd1l2a* _Map_ Set 10 | GTCAAGGGCACATGAAAGAAGAG | AATTAACCCTCACTAAAGGGAGCACCCTC | 68.0 | 60 |
| *pkd1l2a* _Map_ Set 11 | TTATGTGCAGATTCCTCGGATGC | TTTCCACTCAGAAAAGCTGACA | 67.0 | 100 |
| *pkd1l2b* _Map_ Nested Set 1 | ATCCAGAATGTGAAACCACGC | CACGGTCACAAACTCATAGCAG | 68.0 | 100 |
| *pkd1l2b* _Map_ Nested Set 2 | GGACTGGTGAGTTGGACTTGTG | CAAAAGACTGCTGGTTCTCTGTG | 68.0 | 100 |
| *pkd1l2b* _Map_ Set 3 | ATCCAGAATGTGAAACCACGC | CACGGTCACAAACTCATAGCAG | 66.0 | 60 |
| *pkd1l2b* _Map_ Set 4 | TGAGCAAAAGATGAGGAGGAGGTC | AACCGTGTGAAGTGAGGATC | 70.0 | 100 |
| *pkd1l2b* _Map_ Set 5 | TCCATATGTTACCCAGTGTGCTTG | TGAGGAGGCTGCACTTCTCTC | 66.0 | 60 |
| *pkd1l2b* _Map_ Set 6 | TGGTGCTCTGCTGCTATGTTGAC | ACTTCTTCATCTCGCTGTCAAC | 68.0 | 60 |
| *pkd1l2b* _Map_ Set 7 | CTTCAACGGACTGTGAAATGTCGATC | CACGTCAATGACCCGGATTATG | 69.0 | 60 |
promoter at their 5’ end, since they were also used to generate in situ hybridization riboprobes (see Supplementary Table 3 for riboprobe primers). Since some of our mapped sequence data is not present in the current Ensembl chromosome 1 genomic sequence (GRCz10), we cannot map all exon boundaries for *pkd1* and so instead we show the base pair locations within the mapped mRNA transcript that were amplified and sequenced by these primers (see Figure 1 text for further information).

| Gene-Specific Primer | Primer Sequence         |
|----------------------|-------------------------|
| *pkd1*               | CGGCACCTCCTCATGACGAG    |
| *pkd1l2a*            | CTGTGGAGATGTTTGTGAAGGAG |
| *pkd1l2b*            | GTCCAAGATGCGATTGTACGATG |

**Supplementary Table 2. PCR Primers for Inverse PCR**

Sequences of gene-specific reverse primers used in first strand cDNA synthesis during inverse PCR for *pkd1*, *pkd1l2a* and *pkd1l2b* transcripts (see materials and methods for further details).
| Gene Name  | Forward Primer Sequence       | Reverse Primer Sequence       | PCR Product Size (bp) |
|------------|--------------------------------|-------------------------------|-----------------------|
| pkd1 – Set 1 | ACGTGTGTGTGTCTCTGGAC         | GGATCCATTAACCCCTCACTAAAGGGA | 580                   |
| pkd1 – Set 2 | CTCCAGCTACATATCATGGGAC       | AATTTAACCCCTCACTAAAGGGAATCTAGAACCAGCTCTCGTTAC | 1007                 |
| pkd1b       | GTTCTTTGGACGTGACATTTGTC      | AATTTAACCCCTCACTAAAGGGAATCATGGAAGGTCATGCAAAAGAGTCAC | 963                 |
| pkd1l1 – Set 1 | GCGTGTCCTCTGTCTTTACATG     | AATTTAACCCCTCACTAAAGGGAATCAGTCCAGTCCAACCAATCATCATG | 604                 |
| pkd1l1 – Set 2 | GTGATCTTCCATTTGTGCTCTGTAC  | AATTTAACCCCTCACTAAAGGGAATTCACAACATGCAAAAGAGGTG | 447                 |
| pkd1l2a – Set 1 | GTCAGGGCCAGTGAAAAGGAAG    | AATTTAACCCCTCACTAAAGGGAATCAGCCTACTTGGATATTTGAAC | 1136                |
| pkd1l2a – Set 2 | GCTGACTGCCCAAATATAC       | AATTTAACCCCTCACTAAAGGGAATCAGCCTACTTGGATATTTGAAC | 501                 |
| pkd1l2b     | CTGTGTCCTCTGTCTTTACATG      | AATTTAACCCCTCACTAAAGGGAATCAGCCTACTTGGATATTTGAAC | 982                 |
| pkd2        | CTACTTCAGGGAAGCTGAC         | AATTTAACCCCTCACTAAAGGGAATCAGCCTACTTGGATATTTGAAC | 878                 |
| pkd2l1      | CAGAGGCTGTGTTCCAAAGGTC      | AATTTAACCCCTCACTAAAGGGAATCAGCCTACTTGGATATTTGAAC | 553                 |

**Supplementary Table 3. PCR primers for creating in situ hybridization riboprobes**

Primer sequences used to generate riboprobes for *in situ* hybridization. Expected PCR product sizes (in base pairs) are indicated in column 4. * indicates primers published in Coxam *et al.*, (2014). All other riboprobe primers were designed during this study. T3 RNA polymerase promoter sequence at 5’ end of each reverse primer is bold and underlined. *pkd1* – Set 1, *pkd1l1* – Set 1 and *pkd1l2a* – Set 1 primers generated riboprobes that gave the strongest expression in our assays and so were used exclusively in this study. See materials and methods for further information.
| Species               | Percentage Amino Acid Identity to Full-Length Mouse PKD1L3 Protein |
|----------------------|---------------------------------------------------------------|
| Zebrafish            | aa 976-1120 (29%), 1148-1236 (37%), 1258-1316 (39%) and 1529-1601 (38%) |
| Green Spotted Pufferfish | aa 1004-1127 (34%), 1108-1208 (40%), 1209-1312 (38%) and 1539-1595 (46%) |
| Medaka               | aa 976-1044 (31%), 1037-1116 (40%), 1148-1209 (55%) and 1209-1312 (30%) |
| Stickleback          | aa 1037-1168 (35%), 1113-1209 (33%), 1206-1257 (38%), 1255-1312 (46%) and 1529-1594 (43%) |

**Supplementary Table 4. Percentage Amino Acid Identity Between Full-Length Mouse PKD1L3 Protein and Teleost Putative pkd1l3 Orthologs**

Results obtained from blasting full-length mouse PKD1L3 against teleost genomes. Regions of the mouse query sequence that have identity with sequences in respective teleost genomes are indicated by amino acid (aa) numbers and percentage amino acid sequence identity is indicated in parentheses. Numbers in bold correspond to regions of the mouse PKD1L3 query sequence that contain at least part of the GPS motif (amino acids 1020-1058) or PLAT/LH2 domain (amino acids 1131-1245).
Supplementary Material

| Species          | Current Ensembl Gene Name | Ensembl Gene ID       | Chromosome Location          | Supporting Evidence                                                                 |
|------------------|---------------------------|-----------------------|------------------------------|-------------------------------------------------------------------------------------|
| **pkd1**         |                           |                       |                              |                                                                                      |
| Spotted gar      | *pkd1*                    | ENSLOCG00000002998    | LG13(=): 4728181-4832181     | Shares synteny with teleost and mammalian *pkd1* genes (data not shown). Phylogeny (Fig. 3A). |
| Elephant shark   | *pkd1* (two *pkd1* genes in genome) | SINCAMG0000009812    | Scaffold_312(=): 34635-113581 | Shares synteny with teleost and mammalian *pkd1* genes (data not shown). Phylogeny (Fig. 3A). |
| **pkd1b**        |                           |                       |                              |                                                                                      |
| Spotted gar      | *pkd1b*                   | ENSLOCG00000010726    | LG10(=): 27484014-27547534   | Shares synteny with green spotted pufferfish and stickleback *pkd1b* genes (data not shown). Phylogeny (Fig. 3A). |
| Elephant shark   | *pkd1* (two *pkd1* genes in genome) | SINCAMG000001332     | Scaffold_10(=): 2273648-2326286 | Does not share synteny with teleost *pkd1b* genes (data not shown). Formerly called *pkd1* but phylogeny suggests that this gene is *pkd1b* (Fig. 3A). |
| **pkd1l1**       |                           |                       |                              |                                                                                      |
| Spotted gar      | Novel                     | ENSLOCG0000013313     | LG9(=): 50517659-50554155     | Shares synteny with zebrafish *pkd1l1* gene (data not shown). Shares homology with amino acids 205-642, 859-908, 1303-1400 and 1495-1607 of mouse PKD1L1 protein. Phylogeny (Fig. 3A). |
| Elephant shark   | *pkd1l1*                  | SINCAMG000004987      | Scaffold_55(=): 2766931-2840140 | Shares synteny with amniote *PKD1L1* genes (data not shown). Phylogeny (Fig. 3A). |
| **pkd1l2**       |                           |                       |                              |                                                                                      |
| Spotted gar      | Novel                     | ENSLOCG000003938      | LG23(=): 6943671-6977736      | Shares synteny with all teleost *pkd1l2a* genes (data not shown). Shares homology with amino acids 4-525, 1239-1667 and 1804-2410 of mouse PKD1L2 protein. Phylogeny (Fig. 3A). |
| Elephant shark   | *pkd1l2*                  | SINCAMG000003573      | Scaffold_12(=): 2515744-2542623 | Shares synteny with amniote *PKD1L2* genes (data not shown). Phylogeny (Fig. 3A). |
### pkdl3

|                | Species      | ENSLOC0G00000016654 | LG8(+) | Phylogeny (Fig. 3A). |
|----------------|--------------|---------------------|--------|----------------------|
| Spotted gar    | pkdl3        | ENSLOC0G000000143071 LG23(+) | 8258413-8268971 | Shares homology with amino acids 1037-1333 and 1525-1594 of mouse PKD1L3 protein (containing the GPS and PLAT/LH2 domains). Shares synteny with mammalian and teleost PKD1L3 genes (data not shown). |
| Elephant shark |              |                     |        |                      |

### pkdrez

|                | Species      | ENSLOC0G00000016654 | LG8(+): 40852661-40859002 | Phylogeny (Fig. 3A). |
|----------------|--------------|---------------------|--------------------------|----------------------|
| Spotted gar    | pkdrez       | ENSLOC0G00000016654 Scaffold_35(-): 5337490-5343965 | 5337490-5343965 | Shares synteny with mammalian PKDREJ genes (data not shown). Phylogeny (Fig. 3A). |
| Elephant shark |              |                     |                          |                      |

### pkd2

|                | Species      | ENSLOC0G00000013447 | LG4(-): 66726806-66740815 | Phylogeny (Fig. 3B). |
|----------------|--------------|---------------------|--------------------------|----------------------|
| Spotted gar    | pkd2         | ENSLOC0G00000013447 Scaffold_21(-): 1937754-1973648 | 1937754-1973648 | Shares synteny with green spotted pufferfish, stickleback and medaka pkd2 genes (data not shown). Phylogeny (Fig. 3B). |
| Elephant shark |              |                     |                          |                      |

### pkd2l1

|                | Species      | ENSLOC0G00000011334 | LG5(+): 28795999-28813061 | Phylogeny (Fig. 3B). |
|----------------|--------------|---------------------|--------------------------|----------------------|
| Spotted gar    | pkd2l1       | ENSLOC0G00000011334 Scaffold_3(+): 6986034-6999417 | 6986034-6999417 | Shares synteny with all teleost pkd2l1 genes examined (data not shown). Phylogeny (Fig. 3B). |
| Elephant shark |              |                     |                          |                      |

### pkd2l2

|                | Species      | ENSLOC0G00000011380 | LG6(-): 32405994-32414168 | Phylogeny (Fig. 3B). |
|----------------|--------------|---------------------|--------------------------|----------------------|
| Spotted gar    | pkd2l2       | ENSLOC0G00000011380 |                          | Shares synteny with all teleost pkd2l1 genes examined (data not shown). Phylogeny (Fig. 3B). |
| Elephant shark |              |                     |                          |                      |
**pkd2l2**

| Elephant shark | polycystic kidney disease 2 | SINCAMG0000005264 Scaffold_87(+) 2858280-2865700 | Shares no synteny with mammalian and teleost PKD2 genes. Does share synteny with mammalian PKD2L2 genes (data not shown). Phylogeny (Fig. 3B). |

**Supplementary Table 5. Characterization of pkd genes in the genomes of holostei and cartilaginous fish.**

Column one indicates the species. Column 2 lists gene name used in current Ensembl genome assembly (see materials and methods). Column 3 lists Ensembl gene ID. Column 4 shows position in current version of appropriate genome. Column 5 lists data that support the annotations shown here (see results for more info). *pkd* genes were first identified by performing textual searches of the appropriate genome assembly and their synteny and phylogeny examined (data not shown, Fig. 3A-B). We then searched for additional *pkd* genes that are present in either mammalian or teleost genomes, but had not been identified by our previous searches, by performing Tblastn analyses with the polycystin-cation-channel domain of the zebrafish Pkd protein, or the full-length sequence of the mouse PKD protein in question.
2 Supplementary References

Coxam, B., Sabine, A., Bower, N. I., Smith, K. A., Pichol-Thievend, C., Skoczylas, R., et al., (2014). Pkd1 regulates lymphatic vascular morphogenesis during development. *Cell Rep.* 7, 623-633. doi: 10.1016/j.celrep.2014.03.063