The American Journal of Human Genetics, Volume *99*

Supplemental Data

Bi-allelic Mutations in *PKD1L1* Are Associated with Laterality Defects in Humans

Francesco Vetrini, Lisa C.A. D'Alessandro, Zeynep C. Akdemir, Alicia Braxton, Mahshid S. Azamian, Mohammad K. Eldomery, Kathryn Miller, Chelsea Kois, Virginia Sack, Natasha Shur, Asha Rijhsinghani, Jignesh Chandarana, Yan Ding, Judy Holtzman, Shalini N. Jhangiani, Donna M. Muzny, Richard A. Gibbs, Christine M. Eng, Neil A. Hanchard, Tamar Harel, Jill A. Rosenfeld, John W. Belmont, James R. Lupski, and Yaping Yang

Figure S1. Predicted consequences of the variant c.6473+2_6473+3delTG on splicing.

The analysis of the variant was performed by using five predicting tools (SpliceSiteFinder-like, MaxEndScan, NNSPLICE, GeneSplicer and Human Splicing Finder) performed with Alamut Visual of Interactive Biosoftware (http://www.interactive-biosoftware.com/doc/alamut-visual/2.0/splicing.html). In parenthesis are indicated the range of score on the right to the method used. Blue and green vertical bars indicate the predicted 5' and 3' acceptor sites, respectively, with score indicated for those varying between the wild-type and the variant. Sites with scores unchanged between the wild type and the mutated sequence are dimmed. The different predicting tools showed that the c.6473+2_6473+3delTG abolishes the canonical donor splice site.

A

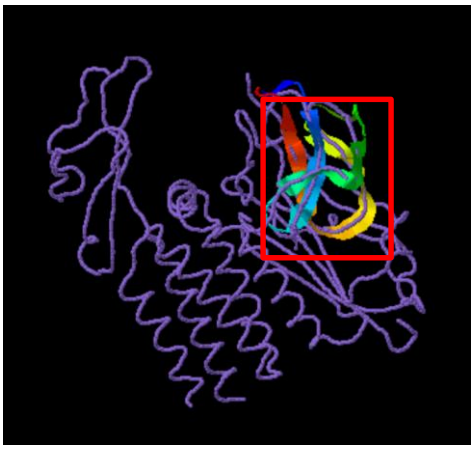
PKD1L1 1690/1722

Rank	PDB Hit	Iden1	Iden2	Cov	Norm. Z-score	
						20 40
						Sec.Str CCCSSSSSSSSSCCCCCCCCCCCCCCCCCCCCSSCCCCCCSSSSSSSCC
						Seq VHFQWIRCLFWDKREWKSERFSPQPGTSPEKVNCSYHRLAAFALLRRK
1	4dloA	0.23	0.25	0.98	1.42	NGTLNPYCVLWDLGTWSTQGCKTV-LTDASHTKCLCDRLSTFAILAQQ
2	4dloA	0.21	0.25	0.98	4.51	NGTLNPYCVLWDDGTWSTQGCKTVLT-DASHTKCLCDRLSTFAILAQQ
3	4dloA	0.23	0.25	0.98	1.31	IELLNPYCVLWDDGTWSTQGCKTVL-TDASHTKCLCDRLSTFAILAQQ
4	4dloA	0.23	0.25	0.98	1.44	NGTLNPYCVLWDLGTWSTQGCKTVL-TDASHTKCLCDRLSTFAILAQQ
5	4dloA	0.24	0.25	0.96	2.07	NGTLNPYCVLWDDGTWSTQGCKTVL-TDASHTKCLCDRLSTFAILAQ-
6	4dloA	0.26	0.25	0.98	1.89	NGTLNPYCVLWDDSEWSTQGCKTVL-TDASHTKCLCDRLSTFAILAQQ
7	4dloA	0.23	0.25	0.98	1.40	NGTLNPYCVLWDLGTWSTQGCKTV-LTDASHTKCLCDRLSTFAILAQQ
8	4dlqA	0.16	0.21	0.79	3.60	KNHFNANCSFWNYGYWSTQGCRLVES-NKTHTTCACSHL---------
9	4dlqA	0.16	0.21	0.79	1.07	FTVFNANCSFWNYGYWSTQGCRLV-ESNKTHTTCACSHL---------
10	4dlqA	0.16	0.21	0.79	1.56	FTVFNANCSFWNYGYWSTQGCRLV-ESNKTHTTCACSHL---------

B

Rank	PDB Hit	TM-score	RMSD	IDEN	Cov
1	4dloA	0.841	1.01	0.234	0.979
2	delqA2	0.643	1.29	0.158	0.792

C



D

Figure S2. ***In silico*** **modeling predictions based on highest ranked identified structural analogs by I-TASSER[1,2,3].**

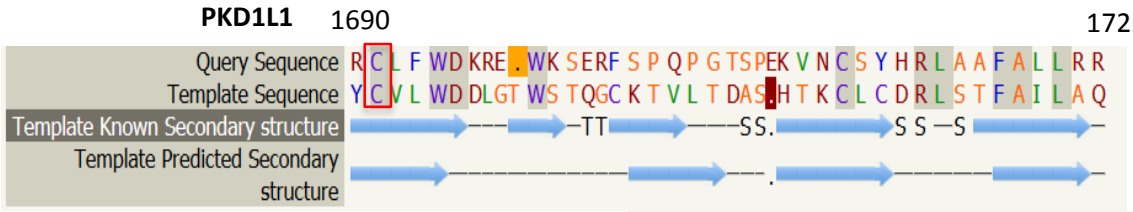
A) Alignments of PKD1L1 1690 to 1722 protein sequence containing the critical Cys1691 to sequence present in PDB library by I-TASSER which uses TM-align structural alignment program to match the first I-TASSER model to all structures in the PDB library (http://www.rcsb.org/pdb/home/home.do). Rank of templates represents the top ten threading templates used by I-TASSER and are obtained from several programs (1: MUSTER, 2: HHSEARCH2, 3: HHSEARCH I, 4: Neff-PPAS, 5: HHSEARCH, 6:dPPAS , 7: MUSTER, 8: HHSEARCH2, 9: HHSEARCH I , 10: HHSEARCH). GPS motif of GAIN and HormR domains of human brain angiogenesis inhibitor 3 (BAI3) (PDB accession code 4DLO, 1-7 predictions) and GAIN and HormR domains of rat CL1 (PDB accession code, 4DLQ, 8-10 predictions) were identified as the top two structural analogs. Residues in template which are identical to the residue in the query sequence are highlighted in color. PKD1L1 Cys1691 is boxed in red. S= strand; C=coil; Iden1=the percentage sequence identity of the templates in the threading aligned region with the query sequence; Iden2= percentage sequence identity of the whole template chains with query sequence; Cov=coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein. Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a normalized Z-score >1 mean a good alignment and vice versa.

B) Ranking of proteins is based on TM-score (measure of similarity between two protein structures with different tertiary structures) of the structural alignment between the query structure and known structures in the PDB library; RMSD represent root mean square deviation between residues that are structurally aligned by TM-align; IDEN is the percentage sequence identity in the structurally aligned region; Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein. TM-score has the value in (0,1), where 1 indicates a perfect match between two structures.

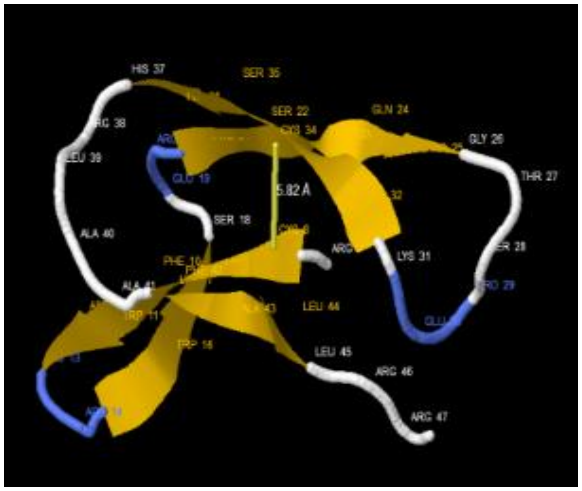
C) Query structure for residues 1690 to 1722 in PKD1L1 using GAIN and HormR domains of human BAI3 model is boxed in red.

D) Query structure for residues 1690 to 1722 in PKD1L1 using GAIN and HormR domains of rat CL1 is boxed in red.

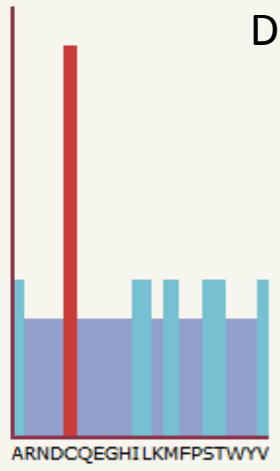
A


PKD1L1
1690
1722
Query Sequence
Template Sequence
Template Known Secondary structure
Template Predicted Secondary structure

B



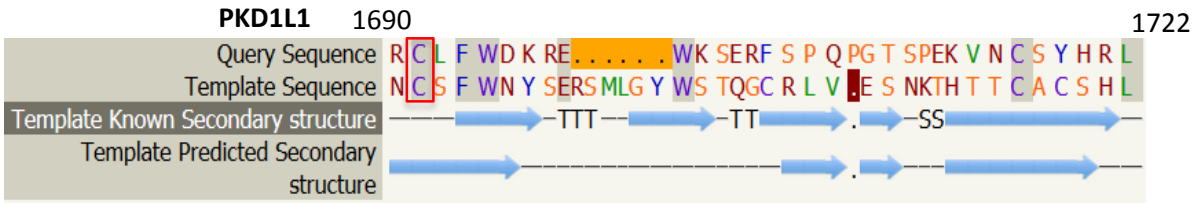
C


ARNDCQEGHILKMFPSTWYV

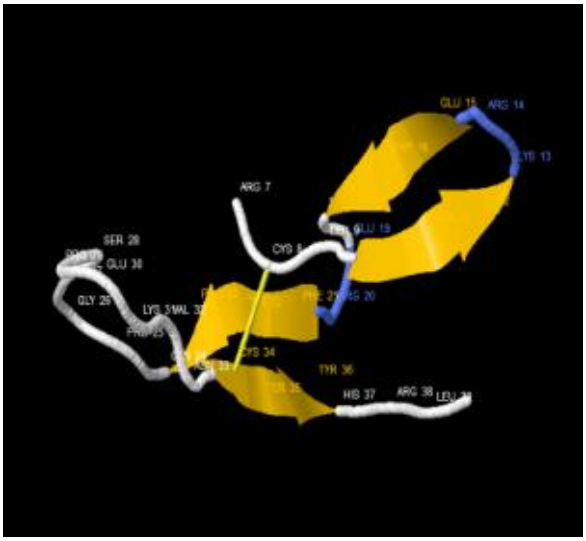
D


ARNDCQEGHILKMFPSTWYV

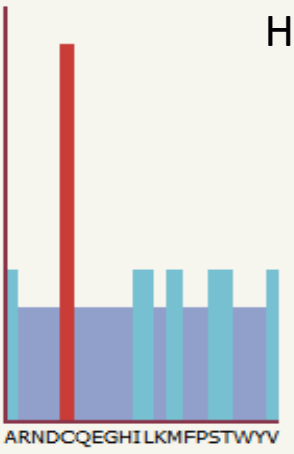
E


PKD1L1
1690
1722
Query Sequence
Template Sequence
Template Known Secondary structure
Template Predicted Secondary structure

F



G


ARNDCQEGHILKMFPSTWYV

H

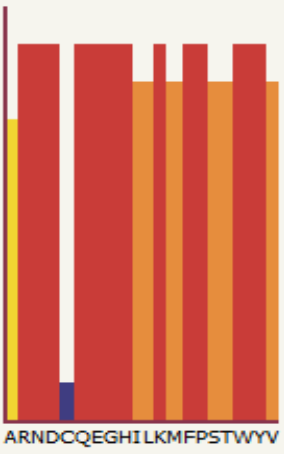

ARNDCQEGHILKMFPSTWYV

Figure S3. ***In silico*** **secondary and 3D structure predictions by Phyre2[4] based on highest ranked models and analysis of the Cys1691 mutation**.

Blue arrows= β-strands; T=hydrogen bonded turn; S=bend. Here the most relevant templates are shown based on the highest structural homology.

A), E) Alignments of PKD1L1 1690 to 1722 protein sequence containing the critical Cys1691 with crystal structures of the GPS motif present in GAIN and HormR domains of human brain angiogenesis inhibitor 3 (BAI3) and GAIN and HormR domains of rat CL1 respectively (PDB accession code 4DLO and 4DLQ respectively). B), F) 3D models showing the critical disulfide bridge between Cys8 (corresponding to the mutated Cys1691 in PKD1L1) and Cys34 (yellow bar).

C), G) Sequence profile graph representing residue preferences in the protein at the position corresponding at Cys1691 of PKD1L1. The twenty possible amino acids are labelled along the x-axis with their one-code letter. The value are calculated by scanning PKD1L1 sequence against a large sequence database using the iterative PSI-blast. The colored bars indicate the favourability of each residue type at the specific position 1691. Tall and red= favorable; short and blue= unfavorable.

D), H) Mutational analysis graph that represents the predicted effect of mutations at the position 1691 in PKD1L1 made by using SuSPect[5] method part of the Phyre2 software package. The twenty possible amino acids are labelled along the x-axis with their one-code letter code. The colored bars indicate the probability that a mutation to the corresponding residue will have a detrimental effect on the protein or the phenotype of the organism; tall and red=likely to affect function, short and blue=unlikely to affect function.

Table S1. WES Sequencing Data of the Probands

Subject	Illumina Platform	Unique Aligned (Mb) [a]	Total Pass filter (Mb) [b]	Avg % Align (PF) Read 1 [c]	Avg % Align (PF) Read 2 [d]	Avg % Error rate Read 1 [e]	Avg % Error rate Read 2 [f]	Unique-ness % [g]	Dupli-cate % [h]	Total Reads Aligned % [i]	Avg Coverage [j]	Reads hit target/ Buffer [k]	Bases 20+ coverage [l]
Subject 1 (629285)	HiSeq 2500	10,650	11,988	95	95	0.54	0.90	93	9.3	61	137	74	98.3
Mother of subject 1	HiSeq 2500	12,068	13,351	95	95	0.41	0.74	92	10.1	64	164	77	98.3
Father of subject 1	HiSeq 2500	11,431	12,431	95	95	0.53	0.92	93	9.2	60	144	73	98.5
Subject 3	HiSeq 2500	10,182.38	10,805	99.16	97.99	0.54	0.83	95.6	6.03	105,454,334	124	77.87	91.61

[a]Unique Aligned (Mbp): the total number of base-pairs in reads that align best to a single location in the reference genome

[b]Total Pass Filter (Mbp): the total number of base-pairs in reads that pass the Illumina quality filters

[c]Avg % Align (PF) Read 1: the average percentage of pass filter base-pairs in Read 1 that align best to a single location in the reference genome

[d]Avg % Align (PF) Read 2: the average percentage of pass filter base-pairs in Read 2 that align best to a single location in the reference genome

[e] Avg % Error rate Read 1: the calculated error rate of bases on Read 1, as determined by aligning to reference genome

[f]Avg % Error rate Read 2: the calculated error rate of bases on Read 2, as determined by aligning to reference genome

[g] Unique-ness %: Percentage of unique reads

[h]Duplicate %: fraction of reads that are identified as duplicate reads – reads whose alignment location is identical to other reads from the same library

[i]Total Reads Aligned: the number of reads that align to the reference genome

[j]Average Coverage: the total number of uniquely aligned bases to the reference genome divided by the size of the reference genome

[k]Reads hit target/buffer: the number of reads whose alignments overlap either a region targeted by the capture reagent, or the 100bp buffer (or both)

[l]Bases 20+ Coverage: the fraction of bases targeted by the capture reagent that are covered by 20 times or more uniquely aligned reads.

References

1.Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. Nat Methods 12, 7-8.

2.Yang, J., and Zhang, Y. (2015). Protein Structure and Function Prediction Using I-TASSER. Curr Protoc Bioinformatics 52, 5 8 1-15.

3. Yang, J., and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res 43, W174-181.

4. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10, 845-858.

5. Yates, C.M., Filippis, I., Kelley, L.A., and Sternberg, M.J. (2014). SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. J Mol Biol 426, 2692-2701.