Supplementary References

S1. Chapman AB, Guay-Woodford LM, Grantham JJ, et al. Renal structure in early autosomal-dominant polycystic kidney disease (ADPKD): The Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) cohort. *Kidney Int.* 2003;64(3):1035-1045.

S2. Torres VE, Abebe KZ, Chapman AB, et al. Angiotensin blockade in late autosomal dominant polycystic kidney disease. *N Engl J Med.* 2014;371(24):2267-2276.

S3. Fromer M, Moran JL, Chambert K, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *American journal of human genetics.* 2012;91(4):597-607.

S4. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv.* 2019:531210.

S5. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet.* 2002;3(5):370-379.

S6. Hubley R, Finn RD, Clements J, et al. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* 2016;44(D1):D81-89.

S7. Longa L, Scolari F, Brusco A, et al. A large TSC2 and PKD1 gene deletion is associated with renal and extrarenal signs of autosomal dominant polycystic kidney disease. *Nephrol Dial Transplant.* 1997;12(9):1900-1907.

S8. Ariyurek Y, Lantinga-van Leeuwen I, Spruit L, Ravine D, Breuning MH, Peters DJ. Large deletions in the polycystic kidney disease 1 (PKD1) gene. *Hum Mutat.* 2004;23(1):99.

S9. Obeidova L, Elisakova V, Stekrova J, et al. Novel mutations of PKD genes in the Czech population with autosomal dominant polycystic kidney disease. *BMC Med Genet.* 2014;15:41.

S10. Xu D, Ma Y, Gu X, et al. Novel Mutations in the PKD1 and PKD2 Genes of Chinese Patients with Autosomal Dominant Polycystic Kidney Disease. *Kidney Blood Press Res.* 2018;43(2):297-309.

S11. Miyatake S, Koshimizu E, Fujita A, et al. Detecting copy-number variations in whole-exome sequencing data using the eXome Hidden Markov Model: an 'exome-first' approach. *J Hum Genet.* 2015;60(4):175-182.

S12. Pounraja VK, Jayakar G, Jensen M, Kelkar N, Girirajan S. A machine-learning approach for accurate detection of copy number variants from exome sequencing. *Genome research.* 2019;29(7):1134-1143.

Methods

Cohort enrollment and informed consent

Patient enrollment was performed with informed consent as part of our Yale School of Medicine Internal Review Board approved study. De-identified cases were also included when sent from collaborators whose studies allowed for sample sharing. The 128 cases evaluated for this study had a PCLD or ADPKD-NMD phenotype.

Whole exome sequencing and XHMM application

Whole exome sequencing was performed on cases as well an available controls from our institution using Illumina HiSeq 2000 at the Yale Center for Genome Analysis (YCGA). The SeqCap EZ Exome v2 capture reagent (NimbleGen/Roche) was used for the sequencing of n= 73 cases and n=500 controls. The xGen® Exome Research Panel (IDT) reagent was used for the remaining n=55 cases and n=339 controls. Seventy-five base pair reads were mapped to the hg19 reference genome using the Burrows Wheeler algorithm (BWA) and Genome Analysis Toolkit (GATK)[18] to generate a *.bam format file.
XHMM analysis was run on the *.bam format files separately for the samples captured by the v2 reagent from those captured with the xGen®xGen reagent. XHMM considers the number of reads ("read depth") at a given position in each case's *.bam file in comparison to other samples which underwent the same exome capture library prep and sequencing modality to report a Z-score. Principal Component Analysis (PCA) was used to eliminate effects of experimental variation prior to normalization. XHMM calls a CNV when the sample's Z-score remains significantly and consistently positive or negative at contiguous locations sampled, suggesting a duplication or deletion respectively. CNVs were filtered to consider only those with a minor allele frequency (MAF) of <0.1% in gnomAD [19] and <10% in the cohort for this autosomal dominant rare disease analysis. CNV calls were then manually reviewed for quality and genomic locations were compared with the coding sequence of the established ADPKD and PCLD disease genes.

Validation of Heterozygous Deletions with genomic qPCR

Primer pairs (Supplemental Table 1) were designed using NCBI Primer Design Tool at two central points in the suggested deletion region and on either side where XHMM suggested read depth had returned to the level of the remaining cohort samples. Quantitative PCR was carried

out with technical triplicates for each reaction using SYBRgreen supermix (Bio-Rad) with genomic DNA as the template. For each investigated region, the relative allele count for the implicated case and three controls was evaluated. Each sample's average cycle threshold for the experimental amplification was normalized to its respective amplification of an intronic region of *GAPDH* on the same plate. Case fold difference was compared to the mean fold difference of the three control samples to determine the relative allele count.

When the central location primer pairs indicated approximately 50% relative allele count, this confirmed the deletion. Additional primer pairs were designed to evaluate relative allele count at locations moving outward on either side at approximately 5 kilobase (kb) intervals until the case had similar allele count to controls, indicating the location is outside of the deletion region. We suggest that such validation be carried out using at least two unique primer pairs, and with attention to common intronic variants which affect primer binding.

PCR amplification and sequencing of the deletion allele

A forward primer upstream and a reverse primer downstream of the region of reduced allele count were designed using the NCBI Primer Design Tool. Phusion polymerase (NEB) was used to amplify the deletion allele between these locations. PCR amplicons were purified following agarose gel electrophoresis using a gel purification kit (Qiagen) and sent for Sanger sequencing. Protocol modifications to attempt to amplify challenging PCR amplicons included the following: alternative thermocycler conditions, alternative primer sites, alternative buffers and polymerases, and additives such as DMSO or betaine.

Supplemental Table 1: Genomic DNA qPCR primers

GANAB

location	Forward Primer	hg19 Position	Reverse Primer	hg19 Position
upstream	ACACTGTCAGGGAAGAAGCG	62430783	GATTCCCCTCACCCACCTCT	62430675
upstream	GTGTTCAGATGCAGTAAGGAAAGG	62426091	ATTGAGCATGCCAAGCCGT	62426011
upstream	TTGCTTCCTTGTAGTGCTCAGG	62417485	GCCCAGAATGGGGTCTACG	62417411
5' UTR/ exon 1	CCCTGCGATAATTTGGAGTGC	62414233	GCCACGTATCCTAGTTCCCG	62414127
exon 3/ intron 3	GTCCTGATTCCCTCACGGTC	62406874	TTGGCCACAATTTCCCCCTT	62406751
intron 18	AGGGACTGGGCTTCATCTCT	62396188	GCTCCCCTTCACTCTTGACC	62396098
intron 20/ exon 21	ATGGTGGAAGCCAAAGGAGG	62394489	GTCTGGGGACCATGATGCTT	62394360
3'UTR	GTTCCCAGCTGTGACACGTTTTG	62392358	TCTAGGCTTCTAGGGTCTGTGG	62392242

SEC63

location	Forward Primer	hg19 Position	Reverse Primer	hg19 Position
upstream/exon 1	CGGCAGTGTCCAAGCTACG	108279319	TGTTCCCACTGTCATCGTACTG	108279174
intron 1	TGAAAGTGGGATACTGTAGAGGC	108257241	AGAGAACACTTTTGAAAATACCACT	108257161
intron 1	TGACTTAGTGGTCAAGGGGAAC	108253253	TGGCCTAAGACAACGCAGC	108253140
intron 1/exon 2	CCAGTTAGCTGGTTTTACCTTCAC	108250788	AACGATACCACATACACCTTCCA	108250662
intron 2	GGCTATGGGGTGTACTGGTT	108250011	AACTCTCTCACCTCCTGTCAGT	108249914
intron 2	GAACACATTGGAGGGAGCAGA	108246794	ACAGCCACACAATGCAACAT	108246665
intron 3	ATGCAAAGTTAAAACTGGGAGGAT	108245147	AGACCCCCAACCTAGATTTTAC	108245041
exon 4/intron 4	ATCCAGATAAAGGAGGTGATGAGG	108243058	TGGATCCTGATACTTACGCAGC	108242984