MATERIALS AND METHODS

Study patients

We performed WES on a total of 179 unrelated patients with bilateral kidney cysts and CKD. This comprised 174 patients drawn from the CRISP (Consortium for Radiologic Imaging Study of PKD) (S1) and HALT-PKD (HALT Progression of Polycystic Kidney Disease) (S2) NIH-sponsored studies in ADPKD and five patients with TBM disease followed up at the nephrology clinics at 12 de Octubre University Hospital, Madrid, Spain (four patients) and Yale University School of Medicine (one patient). Of the 174 patients from the CRISP and HALT-PKD studies, 161 were known to carry an underlying *PKD1* or *PKD2* mutation and constituted the 'PKD mutation-defined ADPKD' group; 13 patients also clinically and radiologically classified as ADPKD in these studies were of milder severity based on their height-adjusted total kidney volume (height-TKV) (S3) but were not known to carry a *PKD1* or *PKD2* mutation and constituted the NMD-ADPKD group (S4).

De-identified patient DNA and clinical information was made available to us by the NIDDK Central Repository and the 12 de Octubre University Hospital in accordance with standard protocols for transfer of human biologic samples. The study was approved by the Yale University Institutional Review Board. Information on the previously performed genetic analysis of the *PKD1* and *PKD2* genes was available for the CRISP and HALT patients through the NIDDK Central Repository.

Exome sequencing and data analysis

Genomic DNA (2-5 ug) was subjected to WES at the Yale Center for Genome Analysis using standard protocols and the workflow previously described (S5). Briefly, whole exome capture was done using the xGen exome capture reagent from Integrated DNA Technologies (IDT®) followed by sequencing of prepared genomic libraries on a NovaSeq6000 Illumina platform. Sequencing data was analyzed using the Genome Analysis Toolkit (GATK) for variant calling and variants annotated using ANNOVAR for detailed variant characterization (S6, S7). False-positive variant calls were excluded by using genotype quality filters and by visual analysis of chromosomal positions. Variant prioritization included rare variants with ethnicity-specific general population minor allele frequency (MAF) cutoff of 0.01% for heterozygous variants and 0.1% for recessive variants in the genome aggregation database (gnomad.broadinstitute.org/) meeting pathogenicity criteria by mutation type and their predicted impact on protein function. Loss of function variants including truncating, frameshift insertion/deletions and splice site variants were classified as 'pathogenic'. Mutation effects on splicing and cryptic splice sites were predicted using the Berkeley Drosophila Genome Project browser and the Human Splicing Finder (S8, S9). Missense variants deemed deleterious on bioinformatics score predictions of MetaSVM were classified as likely pathogenic (S10, S11). Identity by descent was ruled out for the entire 179 patient cohort using a kinship analysis script (S12). The ethnic background for all the patients was defined using a Principal component analysis comparing these to individuals of defined ancestry from the HapMap345 (S13).

For the NMD-ADPKD and TBM patients with kidney cysts we examined all prioritized rare variants including those in genes known to associate with 'kidney cysts', 'CKD', 'FSGS' or 'monogenic ciliopathies' that often cause cystic kidneys. These gene panels were prepared using the OMIM database (https://omim.org) and a manual literature search. The PKD mutation-defined ADPKD patients were investigated for the burden of *COL4A1, COL4A4, COL4A5* mutations using the same variant characterization and pathogenicity criteria (MAF<0.01%, loss of function and deleterious missense variants as detailed above) as for the NMD-ADPKD group. We also calculated the total burden of rare damaging mutations using the same criteria in *COL4A1,*

COL4A4, COL4A5 in the European (non-Finnish) individuals in the gnomAD (gnomad.broadinstitute.org) database (v2.1.1). Allele frequency was calculated as allele counts of variants meeting our MAF and deleteriousness criteria divided by allele number of total alleles tested. Case versus control p-value comparisons were performed using the Fisher's Exact test. Since all these deleterious alleles were present in heterozygous state, for P-value calculations we compared deleterious alleles/normal alleles of the total alleles tested.

Splice assay

We used an in vitro minigene splice assay to investigate the effect of a substitution variant in exon 28 of *COL4A4* predicted by *in silico* analysis to alter normal splicing (S14, S15). A 2.8Kb fragment of *COL4A4* containing the entire genomic sequence spanning from the middle of exon 27 to the middle of exon 29 including introns with or without the mutation in exon 28 was amplified by PCR from patient genomic DNA (patient has one normal and one wild-type allele) and cloned in to an EGFP expression plasmid to create a GFP-*COL4A4* fragment fusion transcript. This minigene plasmid was a gift from Tom Misteli (Addgene plasmid # 20292) (S16). The wild-type or mutant *COL4A4* minigene construct was transfected into HEK cells. mRNA was isolated 24 hours after transfection and reverse transcribed into cDNA using standard protocol. RT-PCR using cDNA from cells transfected with wild-type or mutation minigene plasmid was performed using an EGFP specific forward primer and a *COL4A4* exon 29 reverse primer to assess for presence or absence of the expected GFP-exon27-exon28-exon29 spliced sequence, a larger unspliced product, or alternatively spliced products. The wild-type and mutated amplicons were assessed for their size/migration on an agarose gel and sequenced to demonstrate the altered splicing pattern of the mutated minigene.

SUPPLEMENTARY REFERENCES

S1. Grantham JJ, Torres VE, Chapman AB, Guay-Woodford LM, Bae KT, King BF, Jr., et al. Volume progression in polycystic kidney disease. N Engl J Med. 2006;354(20):2122-30.

S2. Schrier RW, Abebe KZ, Perrone RD, Torres VE, Braun WE, Steinman TI, et al. Blood pressure in early autosomal dominant polycystic kidney disease. N Engl J Med. 2014;371(24):2255-66.

S3. Irazabal MV, Rangel LJ, Bergstralh EJ, Osborn SL, Harmon AJ, Sundsbak JL, et al. Imaging classification of autosomal dominant polycystic kidney disease: a simple model for selecting patients for clinical trials. J Am Soc Nephrol. 2015;26(1):160-72.

S4. Rossetti S, Consugar MB, Chapman AB, Torres VE, Guay-Woodford LM, Grantham JJ, et al. Comprehensive molecular diagnostics in autosomal dominant polycystic kidney disease. J Am Soc Nephrol. 2007;18(7):2143-60.

S5. Gulati A, Somlo S. Whole exome sequencing: a state-of-the-art approach for defining (and exploring!) genetic landscapes in pediatric nephrology. Pediatr Nephrol. 2018;33(5):745-61.

S6. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.

S7. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc. 2015;10(10):1556-66.

S8. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. J Comput Biol. 1997;4(3):311-23.

S9. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37(9):e67.

S10. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet. 2015;24(8):2125-37.

S11. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11(4):361-2.

S12. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.

S13. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904-9.

S14. Cooper TA. Use of minigene systems to dissect alternative splicing elements. Methods. 2005;37(4):331-40.

S15. Besse W, Dong K, Choi J, Punia S, Fedeles SV, Choi M, et al. Isolated polycystic liver disease genes define effectors of polycystin-1 function. J Clin Invest. 2017;127(5):1772-85.

S16. Scaffidi P, Misteli T. Reversal of the cellular phenotype in the premature aging disease Hutchinson-Gilford progeria syndrome. Nat Med. 2005;11(4):440-5.

S17. Badenas C, Praga M, Tazon B, Heidet L, Arrondel C, Armengol A, et al. Mutations in theCOL4A4 and COL4A3 genes cause familial benign hematuria. J Am Soc Nephrol. 2002;13(5):1248-54.

S18. Weng YC, Sonni A, Labelle-Dumais C, de Leau M, Kauffman WB, Jeanne M, et al. COL4A1 mutations in patients with sporadic late-onset intracerebral hemorrhage. Ann Neurol. 2012;71(4):470-7.