Supplementary Information

A discarded synonymous variant in *NPHP3* explains nephronophthisis and congenital hepatic fibrosis in several families

Appendix: Genomics England Research Consortium

Ambrose, J. C.[1] ; Arumugam, P.[1] ; Bleda, M.[1] ; Boardman-Pretty, F.[1,2] ; Boustred, C. R.[1] ; Brittain, H.[1] ; Caulfield, M. J.[1,2] ; Chan, G. C.[1] ; Fowler, T.[1] ; Giess A.[1] ; Hamblin, A.[1] ; Henderson, S.[1,2] ; Hubbard, T. J. P.[1] ; Jackson, R.[1] ; Jones, L. J.[1,2] ; Kasperaviciute, D. [1,2] ; Kayikci, M. [1] ; Kousathanas, A. [1] ; Lahnstein, L. [1] ; Leigh, S. E. A.[1] ; Leong, I. U. S.[1] ; Lopez, F. J.[1] ; Maleady-Crowe, F.[1] ; Moutsianas, L.[1,2] ; Mueller, M.[1,2] ; Murugaesu, N.[1] ; Need, A. C.[1,2] ; O'Donovan P.[1] ; Odhams, C. A.[1] ; Patch, C.[1,2] ; Perez-Gil, D.[1] ; Pereira, M. B.[1] ; Pullinger, J.[1] ; Rahim, T.[1] ; Rendon, A.[1] ; Rogers, T.[1] ; Savage, K.[1] ; Sawant, K.[1] ; Scott, R. H.[1] ; Siddiq, A.[1] ; Sieghart, A.[1] ; Smith, S. C.[1] ; Sosinsky, A.[1,2] ; Stuckey, A.[1] ; Tanguy M.[1] ; Thomas, E. R. A.[1,2] ; Thompson, S. R.[1] ; Tucci, A.[1,2] ; Walsh, E.[1] ; Welland, M. J.[1] ; Williams, E.[1] ; Witkowska, K.[1,2] ; Wood, S. M.[1,2].

[1] Genomics England, London, UK

[2] William Harvey Research Institute, Queen Mary University of London, London, EC1M 6BQ, UK

Methods

Ethical approvals, patient inclusion and clinical evaluation

This study was approved by the North East - Newcastle & North Tyneside 1 Research Ethics Committee (18/NE/350) and the Genomics England 100,000 Genomes Project was approved by the Health Research Authority Research Ethics Committee East of England – Cambridge South (REC Ref 14/EE/1112). Ethical approval by the Scientific Research Committee at Royal Hospital (Oman) was obtained and the study was approved by Research Advisory Council at King Faisal Specialist Hospital and Research Centre (KFSHRC, RAC# 2160 022), Riyadh, Saudi Arabia. Approval for human subject's research was obtained from the Boston Children's Hospital Institutional Review Boards. Clinical data and blood samples were obtained from individuals with echogenic kidneys and CKD using established questionnaire (http://www.renalgenes.org). All patients had clinical features strongly suggestive of an inherited ciliopathy with liver and renal phenotypes. Detailed phenotyping was undertaken by the recruiting physicians and authors of this study. Written and informed consent was obtained from all patients, guardians and family members involved in this study.

DNA isolation, library preparation and Whole Exome Sequencing

gDNA was isolated from whole blood from the Omani index family as previously reported (Al Alawi et al., 2020) and WES was performed via Novogene Co., Ltd (China) using SureSelect Human All Exon V6 Enrichment Kit (Agilent Technologies, CA, USA) and Illumina HiSeq platform (Illumina, San Diego, CA, USA). Analysis of raw data (FASTQ format) were performed including sequence reads mapping to the human reference genome hg19 using BWA (Li & Durbin, 2009), removal of PCR duplicates using Picard (http://picard.sourceforge.net), alignment refinement using GATK, coverage analysis and SNP and *indel* calling using GATK's Haplotype Caller (McKenna et al., 2010).

For the index patient OM-1, total reads were 77,972,988, mapped reads 75,169,466, % mapped: 99.9% and 20x coverage for 95.6%.

Variants and CNVs detection and annotation

SNP and INDEL VCF files were investigated using Qiagen Clinical Insight tool for variants filtration and annotation. Initial standard filtering approaches were as follows: homozygous variants (assumption based on parental consanguinity), AF<1% in any subpopulation in reference populations (gnomAD, ExAC, NHLBI ESP exomes, 1000 Genomes Project), ACMG guideline classification as likely pathogenic or pathogenic, frameshift, in-frame indel, or start/stop codon change, a missense unless predicted tolerated by SIFT or PolyPhen-2 and a splice site loss up to 7 bases into intron. Subsequently, filter settings were adjusted to look for any homozygous, rare (AF<1% in any subpopulation in reference populations) variants among disease-associated candidate genes. To detect regions of homozygosity, WES genotype data were used to create homozygosity maps using the online homozygosity mapper tool (http://www.homozygositymapper.org/). *NPHP3* variants were annotated using transcript RefSeq NM_153240.5.

Variants validation by Sanger sequencing

Sanger sequencing was utilized to confirm variants and their segregation. PCR amplification was performed using *Taq* PCR master mix (Qiagen) kit, as per the manufacturer's instructions using *NPHP3* specific intronic primers (5-GGTGGAATTCAGATAGTGAGTGG-3 and 5-CTCTGGCTTGCCATTGATCT-3). Sanger sequencing was outsourced to EuroFins GATC Biotech (Germany).

RNA preparation and RT-PCR

Total RNA from blood was isolated using RNeasy mini kit (Qiagen) according to the manufacturer's instructions and quantified using a NanoDrop 2000 spectrophotometer. 1µg RNA was reverse-transcribed using an oligo-dT primer and SuperScript III Reverse Transcriptase (Thermo Fisher Scientific). The resulting cDNA was diluted 1 in 10 and used in PCR reactions using GoTaq® DNA Polymerase (Promega) and *NPHP3* gene and transcript-specific primer pairs (Forward primer: 5-CAGCCAGCAGGGAAGTAAACA-3 and reverse primer: 5-GCCAAACTTCTTCCACTGCA-3), to identify splice products of *NPHP3.* Product electrophoresis was performed on a 2% agarose gel. Sanger sequencing of RT-PCR products was outsourced to EuroFins GATC Biotech (Germany).

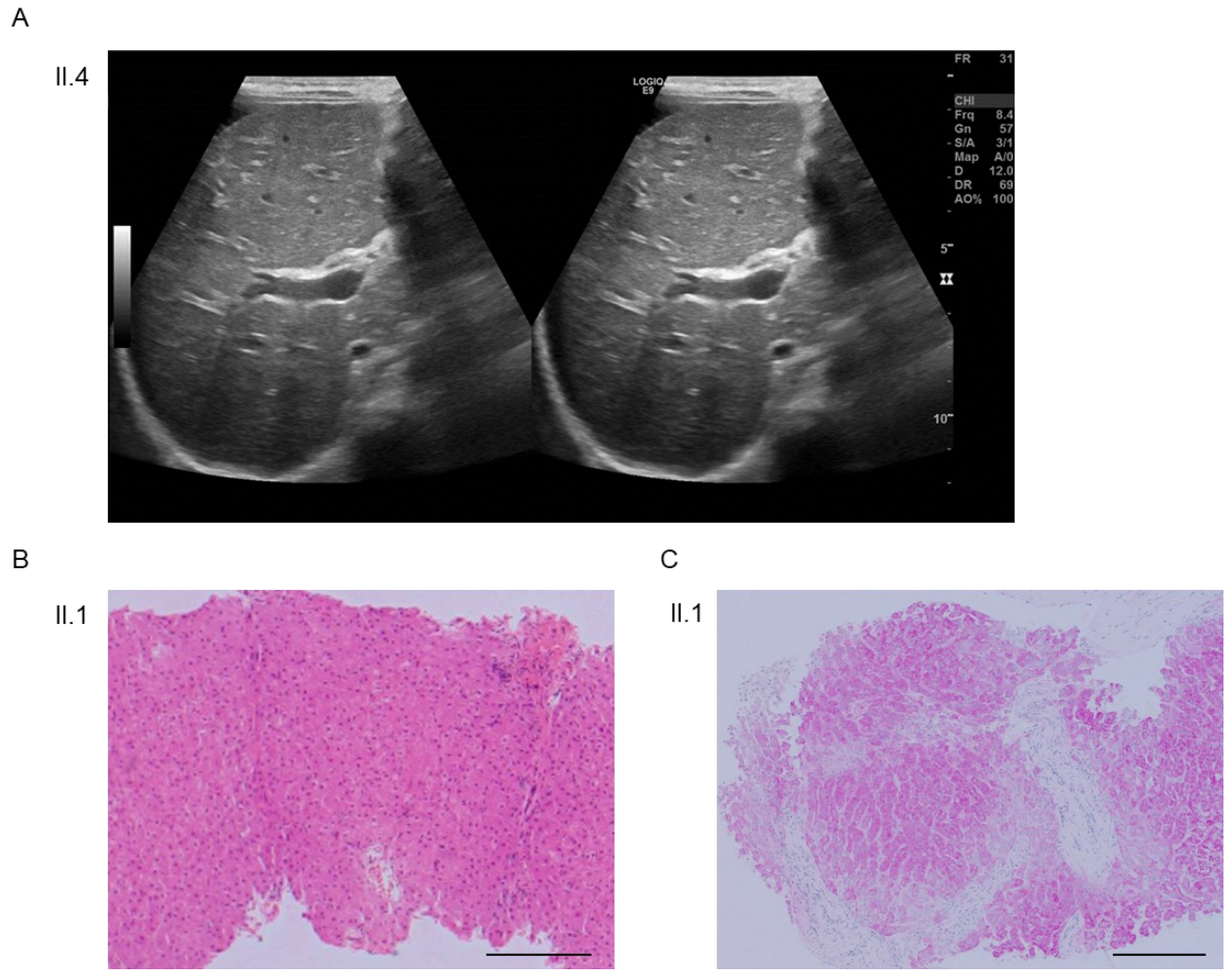
Genomics England 100,000 Genomes Project

All participants in the 100,000 Genomes Project have provided written consent to provide access to their anonymised clinical and genomic data for research purposes. The 100,000 Genomes research and clinical project model and its informed consent process has been approved by the National Research Ethics Service Research Ethics Committee for East of England – Cambridge South Research Ethics Committee. Whole genome sequencing was performed by Genomics England via the 100,000 Genomes Project using the Illumina TruSeq DNA PCR-Free sample preparation kit (Illumina, Inc.) and an Illumina HiSeq 2500 sequencer, generating a mean depth of 45x (range from 34x to 72x) and greater than 15x for at least 95% of the reference human genome. WGS reads were aligned to the Genome Reference Consortium human genome build 37 (GRCh37) using Isaac Genome Alignment Software (version 01.14; Illumina, Inc.). Sequence data was analysed using bcftools scripts designed to search vcf.gz files and individual BAM files were viewed using IGV.

UK Biobank

UK Biobank is a large and detailed prospective study with over 500,000 participants aged 40–69 years when recruited in 2006–2010 and globally accessible to approved researchers who are undertaking health-related research that is in the public interest (Sudlow et al., 2015). The study has collected and continues to collect extensive phenotypic and genotypic detail about its participants, including genome-wide genotyping and longitudinal follow-up for a wide range of health-related outcomes. Genome-wide genotyping was performed on all UK Biobank participants using the Applied Biosystems UK Biobank Axiom Array. Approximately 850,000 variants were directly measured and over 90 million variants were imputed using the Haplotype Reference Consortium and UK10K + 1000 Genomes reference panels, as extensively described elsewhere (Bycroft et al., 2018). Furthermore, exome data on ~200,000 individuals have been made available (Van Hout et al., 2020). Variants in *NPHP3* in the UK biobank exomes were extracted from the population level exome OQFE variants in pVCF format (field 23156) using bcftools. Variants were examined in R using the VariantAnnotation Bioconductor library (Obenchain et al., 2014). Only individuals included in the currently consented individuals were screened. Ethics approval for the UK Biobank study was obtained from the North West Centre for Research Ethics Committee (11/NW/0382).

Supp. Figure S1. Liver ultrasonography and histopathology findings in Omani patients

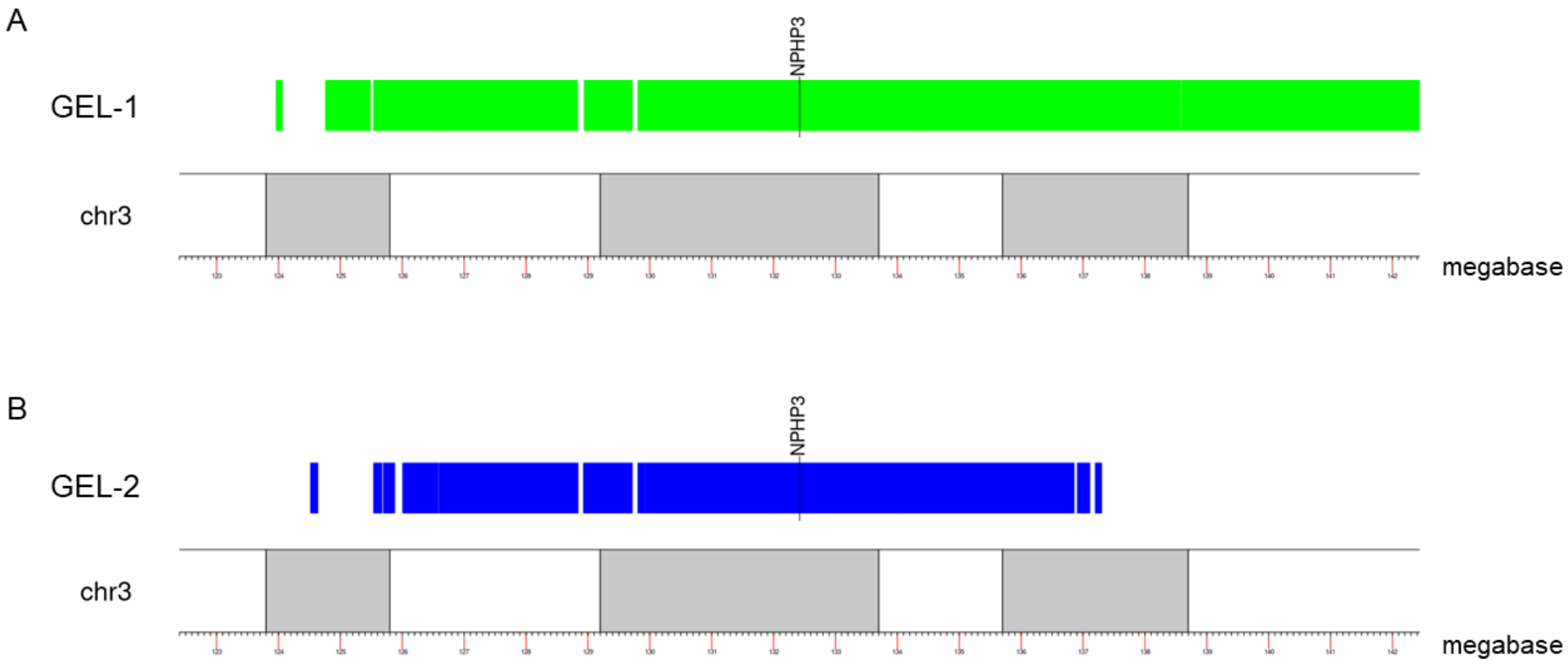


A. Liver ultrasonography from the patient OM-2 (II.4) showing coarse and increased liver echogenicity, with increased periportal echogenicity. There are no focal liver lesions. B. Haematoxylin & Eosin stain on liver biopsy from patient OM-1 (II.1) showing vague nodular pattern. Bar: 200µm. C. Periodic Acid Schiff stain on liver biopsy from patient OM-1 (II.1) showing nodular pattern with ductular proliferation resembling ductal plate malformation. Bar: 200µm

Supp. Figure S2. Genome-wide shared homozygosity in index family and location of disease-associated candidate genes

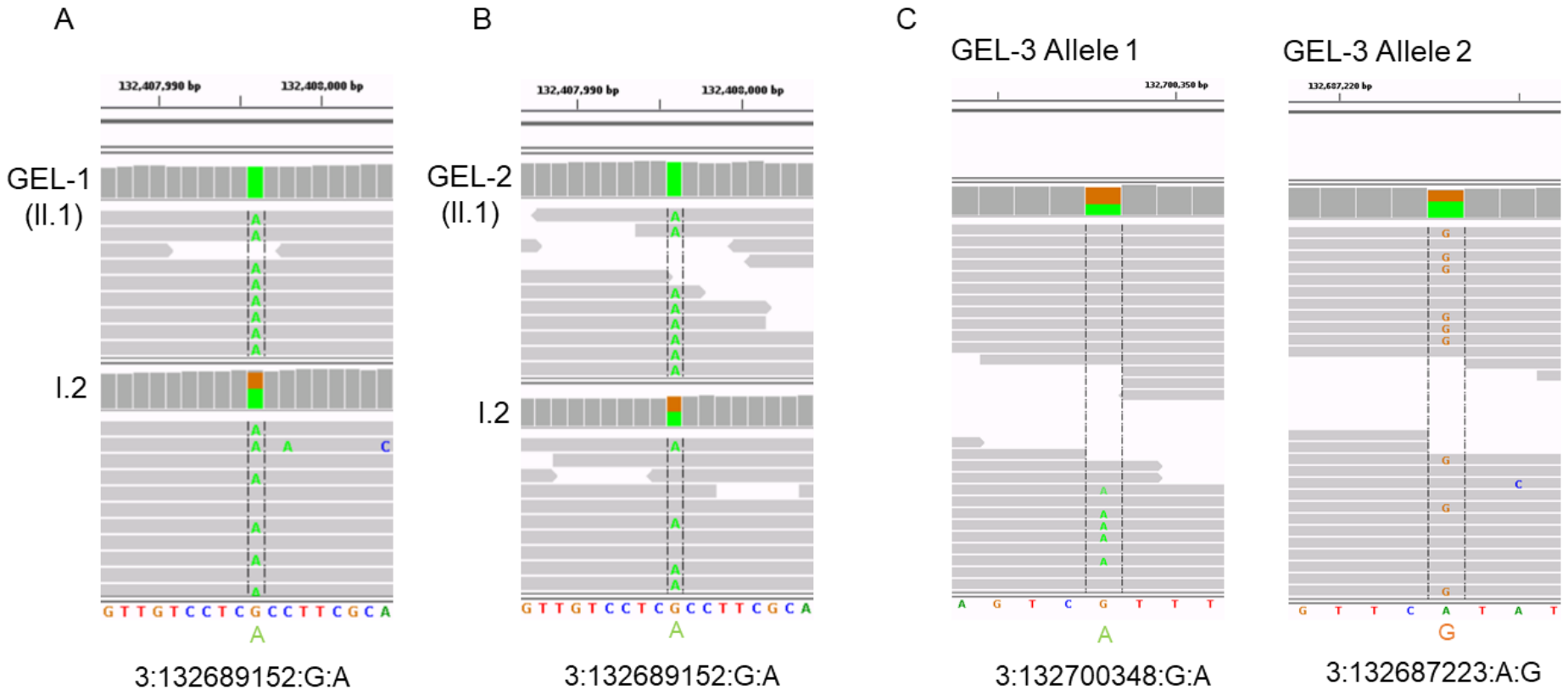
Genome-wide homozygosity map overlay for patients OM-1 (red) and OM-2 (cyan). Candidate genes associated with autosomal recessive cystic kidney disease and congenital hepatic fibrosis (Table S1) are indicated. Note that only *NPHP3* on chromosome 3 lies within a region of shared homozygosity.

Supp. Figure S3. Homozygosity maps



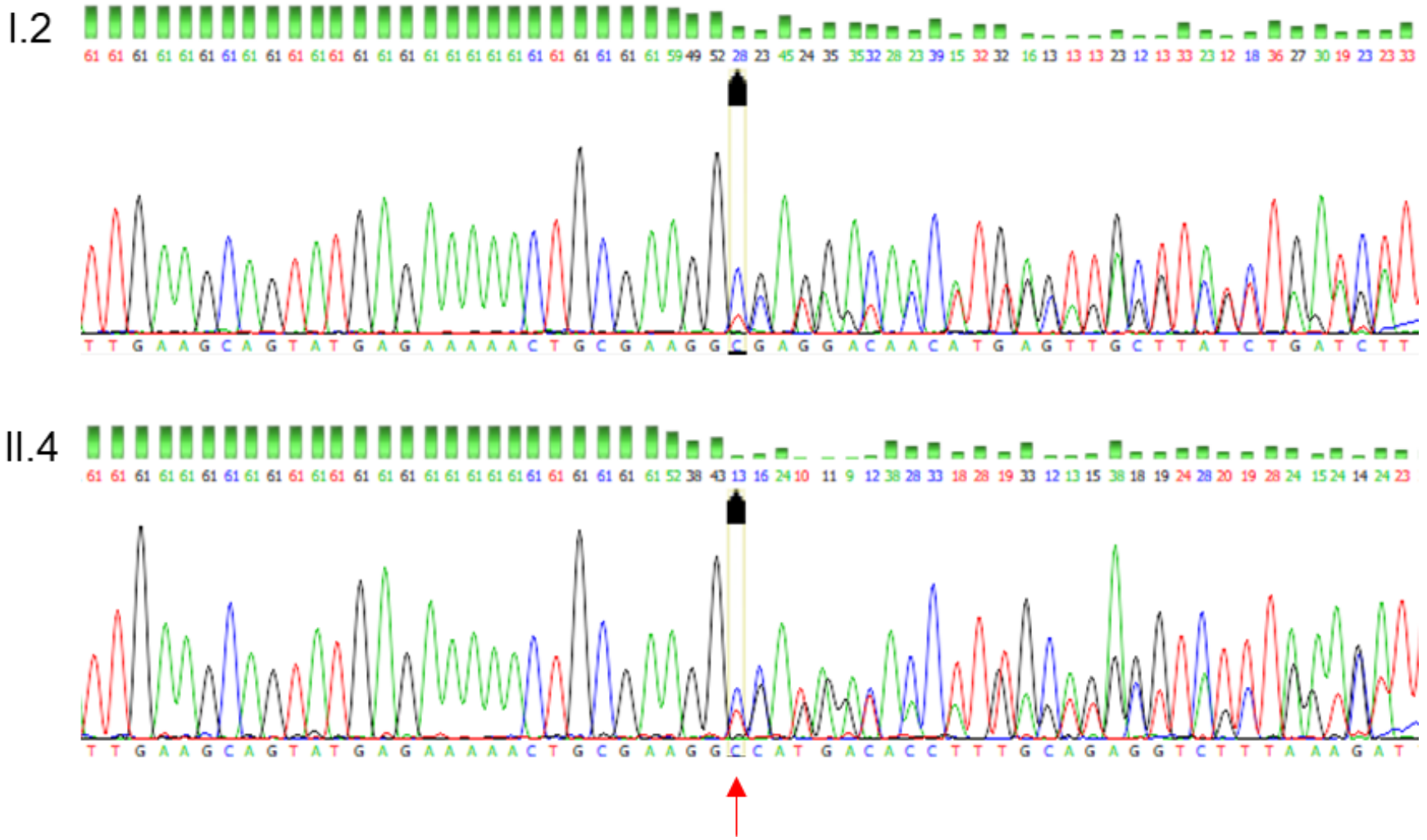
Homozygosity map zoom-in on chromosome 3 showing regions of homozygosity centred on *NPHP3* for patient GEL-1 (A) and GEL-2 (B).

Supp. Figure S4. NGS read alignments from the 100,000 Genomes Project



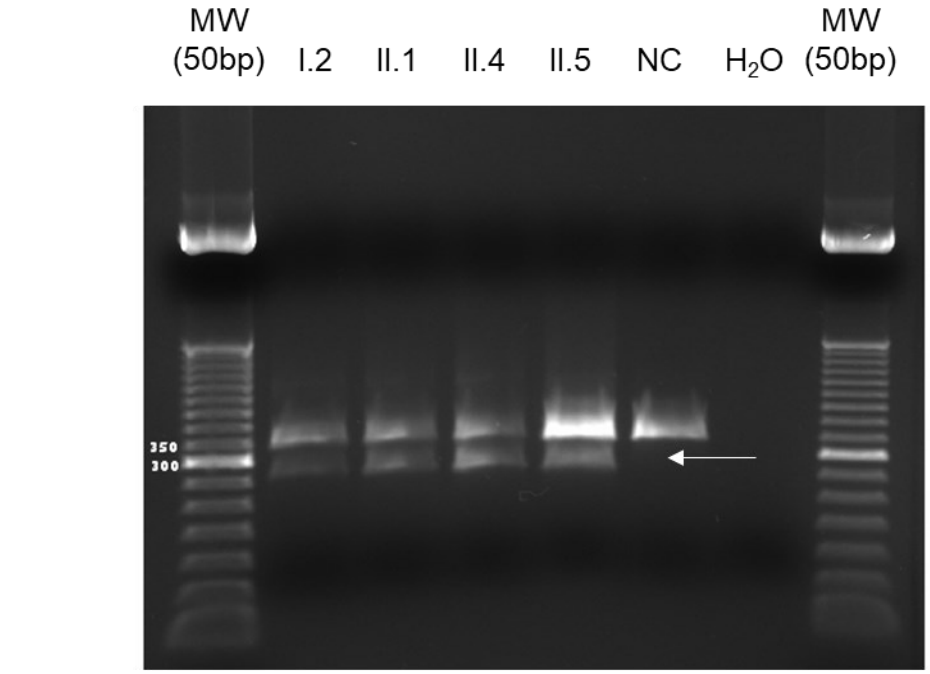
NGS read alignments using BAM files from families recruited in Genomics England 100,000 Genomes Project and assessed using Integrative Genomics Viewer (Robinson et al., 2011). Shown are patient GEL-1 and mother (A), patient GEL-2 and mother (B) and patient GEL-3 with the 2 identified *NPHP3* variants.

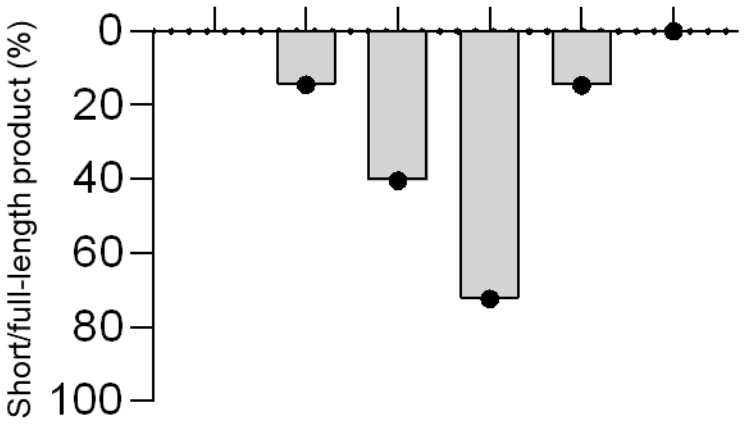
Supp. Figure S5. RT-PCR Sanger chromatograms in heterozygous and homozygous individuals

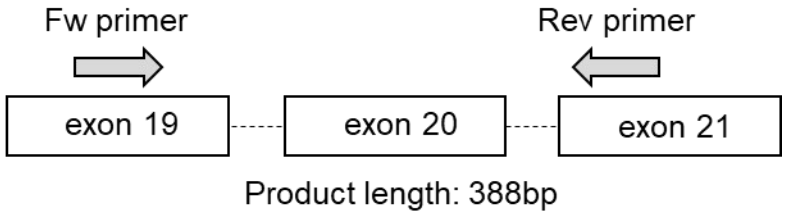


Sanger sequence extract from RT-PCR (forward primer exon 19, reverse primer exon 21) performed on whole blood RNA from heterozygous mother I.2 and homozygous patient OM-2 (II.4). Sequence reveals presence of alternative splicing event introduced by c.2805C>T (arrow) identical to sequence shown in Figure 2 and present in both individuals.

Supp. Figure S6. RT-PCR on blood RNA

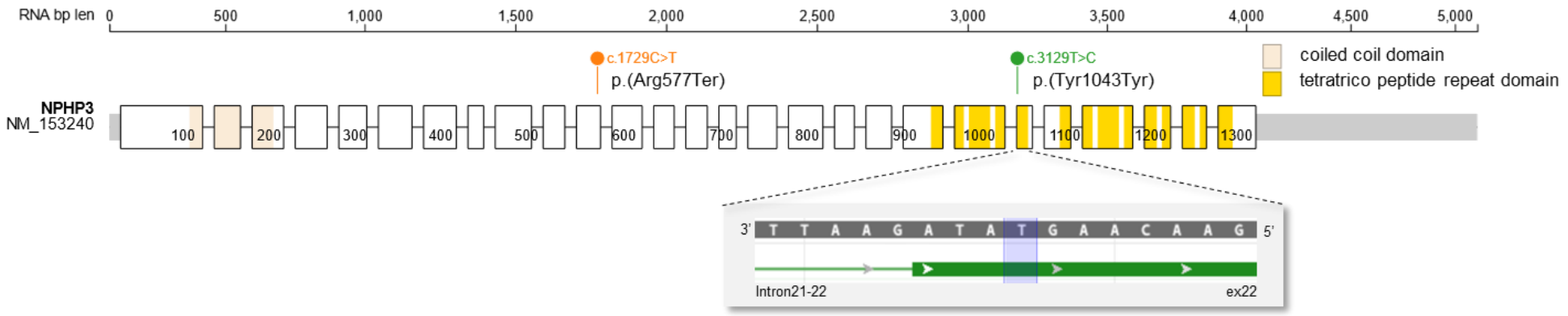






RT-PCR using RNA isolated from whole blood on unrelated control (NC), heterozygous mother (I.2), homozygous patient OM-1 (II.1), homozygous sister OM-2 (II.4) and unaffected (presumed heterozygous) sister (II.5). *NPHP3* primer pair (arrows) designed as illustrated on the right with expected product length. Note shortened transcript (white arrow) detected in heterozygous and homozygous carriers of the c.2805C>T variant. The observed shift is compatible with the predicted loss of 80 nucleotides. Note also the increased relative abundance of shortened transcript to full length transcript observed in homozygotes (II.1 & II.4) compared to (presumed) heterozygotes (I.2 & II.5), as shown below the blot. Quantification using ImageJ (Schneider et al., 2012).

Supp. Figure S7. *NPHP3* variants detected in patient GEL-3



NPHP3 RNA (RefSeq NM_153240.5) and exon structure with UTR in grey and annotated with domains (coiled coil and tetratrico peptide repeat domain). Variants c.1729C>T predicted to lead to a premature stop codon and variant c.3129T>C predicted to lead to a synonymous change are identified. The position of c.3129T>C near exon-intron boundary is highlighted below.

Supp. Table S1: Candidate genes associated with autosomal recessive cystic kidney disease and congenital hepatic fibrosis

Gene name	OMIM phenotype	MIM	Genomic coordinates (GRCh38)
AHI1	Joubert syndrome 3	608629	6:135,283,531-135,497,770
B9D1	?Meckel syndrome 9 Joubert syndrome 27	614209 617120	17:19,334,694-19,377,912
B9D2	?Meckel syndrome 10 Joubert syndrome 34	614175 614175	19:41,354,416-41,364,533
CC2D2A	COACH syndrome 2 Joubert syndrome 9 Meckel syndrome 6	619111 612285 612284	4:15,468,659-15,601,556
CEP164	Nephronophthisis 15	614845	11:117,316,345-117,413,265
CEP290	Joubert syndrome 5 Leber congenital amaurosis Meckel syndrome 4 Senior-Loken syndrome 6	610188 611755 611134 610189	12:88,049,012-88,142,215
DZIP1L	Polycystic kidney disease ¥	617610	3:138,061,989-138,115,861
INPP5E	Joubert syndrome 1	213300	9:136,428,618-136,439,860
MKS1	Bardet-Biedl syndrome 13 Joubert syndrome 28 Meckel syndrome 1	615990 617121 249000	17:58,205,435-58,219,604
NPHP3	Meckel syndrome 7 Nephronophthisis 3 Renal-hepatic-pancreatic dysplasia 1	267010 604387 208540	3:132,680,608-132,722,408
PKHD1	Polycystic kidney disease 4, with or without hepatic disease	263200	6:51,614,684-52,087,624
RPGRIP1L	Joubert syndrome 7 Meckel syndrome 5	611560 611561	16:53,598,152-53,703,858
TCTN2	?Meckel syndrome 8	613885	12:123,671,107-123,708,404
TMEM67	COACH syndrome 1 Joubert syndrome 6 Meckel syndrome 3 Nephronophthisis 11	216360 610688 607361 613550	8:93,754,843-93,832,652
TMEM216	Joubert syndrome 2 Meckel syndrome 2	608091 603194	11:61,391,981-61,398,846
TTC21B	Nephronophthisis 12 Short-rib thoracic dysplasia 4 with or without polydactyly	613820 613819	2:165,873,361-165,953,780
WDR19	Nephronophthisis 13 Senior-Loken syndrome 8	614377 616307	4:39,182,472-39,285,809

¥ Ductal plate malformation was detected in a *Dzip1l* loss-of-function mouse model and 1 patient described with hepatosplenomegaly (Lu et al., 2017)

Supp. Table S2: Clinical features and *NPHP3* alleles in patient GEL-3

Patient ID	Gender, age	Clinical features	Alleles	Gene	Nucleotide change	Predicted amino acid change	ACMG Classification	GnomAD	HSF impact prediction
GEL-3	F, 34	ESKD in childhood, hypertension, AR inheritance	Allele 1	*NPHP3*	c.1729C>T	p.(Arg577Ter)	pathogenic	3/0/282570	NA
			Allele 2	*NPHP3*	c.3129T>C	p.(Tyr1043Tyr)	uncertain signif.	1/0/237158	Alteration of auxiliary sequences: Significant alteration of ESE / ESS motifs ratio

Transcript used: RefSeq NM_153240.5

Abbreviations: AR autosomal recessive; HSF, Human Splicing Finder; ESE, exonic splicing enhancer; ESKD, end stage kidney disease; ESS, exonic splicing silencer; GnomAD, Genome Aggregation Database

HSF impact prediction was generated in 06/2021 through https://hsf.genomnis.com (Desmet et al., 2009) using search term ENST00000337331.10:c.3129T>C.

Supplementary References

Al Alawi, I., Molinari, E., Al Salmi, I., Al Rahbi, F., Al Mawali, A., & Sayer, J. A. (2020). Clinical and genetic characteristics of autosomal recessive polycystic kidney disease in Oman. *BMC Nephrology*, *21*(1), 347. https://doi.org/10.1186/s12882-020-02013-2

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., & Béroud, C. (2009). Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*, *37*(9), e67. https://doi.org/10.1093/nar/gkp215

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lu, H., Galeano, M. C. R., Ott, E., Kaeslin, G., Kausalya, P. J., Kramer, C., Ortiz-Brüchle, N., Hilger, N., Metzis, V., Hiersche, M., Tay, S. Y., Tunningley, R., Vij, S., Courtney, A. D., Whittle, B., Wühl, E., Vester, U., Hartleben, B., Neuber, S., … Bergmann, C. (2017). Mutations in DZIP1L, which encodes a ciliary-transition-zone protein, cause autosomal recessive polycystic kidney disease. *Nature Genetics*, *49*(7), 1025–1034. https://doi.org/10.1038/ng.3871

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010).

The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Obenchain, V., Lawrence, M., Carey, V., Gogarten, S., Shannon, P., & Morgan, M. (2014). VariantAnnotation: A Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics (Oxford, England)*, *30*(14), 2076–2078. https://doi.org/10.1093/bioinformatics/btu168

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Van Hout, C. V., Tachmazidou, I., Backman, J. D., Hoffman, J. D., Liu, D., Pandey, A. K., Gonzaga-Jauregui, C., Khalid, S., Ye, B., Banerjee, N., Li, A. H., O'Dushlaine, C., Marcketta, A., Staples, J., Schurmann, C., Hawes, A., Maxwell, E., Barnard, L., Lopez, A., … Baras, A. (2020). Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature*, *586*(7831), 749–756. https://doi.org/10.1038/s41586-020-2853-0