Cystin gene mutations cause autosomal recessive polycystic kidney disease associated with altered *Myc* expression

Chaozhe Yang[1], Amber K. O'Connor[1], Robert A. Kesterson[2], Jacob A. Watts[1, 2], Amar J. Majmundar[3], Daniela A. Braun[3], Monkol Lek[4], Kristen M. Laricchia[4], Hanan M. Fathy[5], Shirlee Shril[3], Friedhelm Hildebrandt[3], and Lisa M. Guay-Woodford[1, 2]

[1]Center for Translational Research, Children's National Research Institute, Washington, DC 20010

[2]Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294

[3]Department of Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115

[4]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA

[5]Pediatric Nephrology Unit, Alexandria University, Alexandria, Egypt

To whom correspondence should be addressed:

Lisa M. Guay-Woodford, MD

Center for Translational Research

111 Michigan Ave NW

Washington, DC 20010

Tel: 202-476-6439

Fax: 202-476-4610

LGuaywoo@childrensnational.org

Running headline: Cystin defects and *Myc* expression in ARPKD

1.

Abstract

Mutation of the *Cys1* gene underlies the renal cystic disease in the *Cys1$^{cpk/cpk}$* (*cpk)* mouse that phenocopies human autosomal recessive polycystic kidney disease (ARPKD). Cystin, the protein product of *Cys1*, is expressed in the primary apical cilia of renal ductal epithelial cells. In previous studies, we showed that cystin regulates *Myc* expression via interaction with the tumor suppressor, necdin. Here, we demonstrate rescue of the *cpk* renal phenotype by kidney-specific expression of a cystin-GFP fusion protein encoded by a transgene integrated into the *Rosa26* locus. In addition, we show that expression of the cystin-GFP fusion protein in collecting duct cells down-regulates expression of *Myc* in *cpk* kidneys. Finally, we report the first human patient with an ARPKD phenotype due to homozygosity for a predicted deleterious splicing defect in *CYS1*. These findings suggest that mutations in the *Cys1* mouse and *CYS1* human orthologues cause an ARPKD phenotype that is driven by overexpression of the *Myc* proto-oncogene.

Key words:

ARPKD, *Cys1*/*CYS1*, *Rosa26* locus, cystin, *Ksp-cre*, *Myc*

Translational Statement:

The cystin-deficient *cpk* mouse is a model for the study of autosomal recessive polycystic kidney disease (ARPKD). We show that the *cpk* mouse phenotype is associated with altered *Myc* expression. To date, the clinical relevance of cystin deficiency to human disease was unclear, due to the absence of ARPKD cases associated with *CYS1* mutations. We report the first case of ARPKD linked to a *CYS1* mutation disrupting normal splicing. These findings confirm the relevance of cystin deficiency to human ARPKD, implicate *Myc* in disease initiation or progression, and validate the *cpk* mouse as a translationally relevant disease model.

Introduction

Autosomal recessive polycystic kidney disease (ARPKD; MIM 263200) affects 1:26,500 live births.[1] Cohort studies indicate that 75-80% of patients with typical ARPKD have mutations in the Polycystic Kidney and Hepatic Disease 1 (*PKHD1*) gene.[2-5] Mutations in the *DZIP1L* gene account for less than 0.1% of affected patients,[6] while mutations in other hepato-renal fibrocystic disease genes, eg. *HNF1B*, *PKD1*, *NPHP2*, *NPHP3*, and *NPHP13,* can phenocopy ARPKD.[7] For reasons yet to be explained, mice with targeted disruption of *Pkhd1* exhibit little or no kidney disease.[8-15] In the absence of a *Pkhd1* mutant mouse model that accurately recapitulates the human disease phenotype, the *cpk* mouse carrying a spontaneous truncating mutation in *Cys1* has been the most widely studied mouse model of ARPKD.[16,17] Cystin, the *Cys1* gene product, is a 145-amino acid cilia-associated protein that is expressed in mouse embryonic kidney and liver ductal epithelium.[18] Disruption of cystin function results in elevated *Myc* expression in collecting duct epithelial cells[19-22] and increased cell proliferation.[19,23] In previous work, we have demonstrated that in renal collecting duct epithelia, cystin physically interacts with necdin in a regulatory complex that modulates *Myc* expression.[24]

Cystin deficiency-associated disruption of ciliary signaling and/or overexpression of *Myc* is associated with aberrant SMAD3 phosphorylation,[25] overexpression of *Fos* and *Kras* proto-oncogenes,[19-21] elevated levels of growth factors,[26] aberrant localization and abundance of the epidermal growth factor receptor (EGFR) on the apical surface of collecting duct cells[27], abnormal levels of basement membrane components,[28-30] and epithelial cell adhesion molecules.[31,32] Until now, the relevance of these effects of cystin deficiency for human disease was unclear in the absence of ARPKD patients with mutations in human *CYS*. Here we present the first case of human ARPKD due to homozygosity for a *CYS1* mutation, in this case predicted to cause defective splicing. We also show that complementation of defective *Cys1* in the kidneys of *Cys1^cpk/cpk^* (*cpk*) mice rescues both *Myc* overexpression and the collecting duct cyst phenotype. These studies suggest that up-regulation of *Myc* expression *in vivo* may play a central role in the

3.

pathogenesis of mouse recessive polycystic kidney disease, with important implications for human ARPKD.

Results

Phenotypic rescue of cpk mice by kidney-specific expression of a cystin-GFP fusion protein

We generated a conditional expression *Cys1* transgenic *Cys1^cpk/cpk^ (cpk)* mouse line carrying a *Cys1-GFP* transgene knock-in at the *Rosa26* locus. In these mice, *Cys1-GFP* transgene expression is precluded by the presence of a loxP-flanked termination sequence consisting of a PGK-Neo cassette (Figure1A, T^{OFF} allele). The *Cys1*-GFP transgene is expressed by the ROSA26 promoter only after *Cre*-mediated deletion of the loxP-flanked PGK-Neo cassette (Figure1A, T^{ON} allele). We crossed *Rosa26-Cys1-GFP* mice with *Cys1^cpk/+^* mice to generate *Cys1^cpk/+^;Rosa26-Cys1-GFP* mice, which were then crossed with *Ksp-Cre* transgenic mice[33] to generate *Cys1^cpk/+^;Rosa26-Cys1-GFP;Ksp-Cre* progeny. In these mice, *Cre* expression, controlled by the *Ksp*-cadherin regulatory elements, occurs exclusively in the developing distal renal tubular epithelium and the genitourinary tract[34] resulting in high level expression of the cystin-GFP fusion protein in the collecting ducts and loops of Henle and low or no expression in the proximal tubules. Finally, the rescue experiments were carried out by crossing *Cys1^cpk/+^;Rosa26-Cys1-GFP;Ksp-Cre* mice with *Cys1^cpk/+^* mice. Genotype-confirmed *Cys1^cpk/cpk^;Rosa26-Cys1-GFP;Ksp-Cr*e experimental "rescue" (R) mice were compared to their *Cys1^+/+^;Rosa26-Cys1-GFP;Ksp-Cre* control (C) littermates (Figure 1B). While *cpk* mice are characteristically smaller than wild-type littermates and die by 21 days of age,[35] no differences in body size or survival were observed between the rescue (R) mice and their littermate controls (Figure 1C, left panel). Kidney sizes at postnatal days 14 and 21 were not significantly different in R and wild-type (WT) mice (Figure 1C, right panel), while age-matched *Cys1^cpk/cpk^* (*cpk*) mice

4.

exhibited the characteristic cystic kidney phenotype. These results indicate that the gross renal phenotype of the *cpk* mouse was rescued by kidney-specific expression of cystin-GFP.

Expression of cystin-GFP fusion protein in the kidneys of rescued cpk mice

We examined the expression of the cystin-GFP fusion protein in the kidneys of R mice. Endogenous cystin was detectable in the kidneys of both WT and C mice and absent from the kidneys of R mice (Figure 1D). The cystin-GFP fusion protein of ~50kDa was detected in R and C mice (Figure 1D, lanes 3 and 4). These results demonstrate that cystin-GFP expression was associated with Cre-mediated excision of the PGK-Neo cassette.

Dual immunofluorescence staining with antibodies against GFP and aquaporin-2 (AQP2) was used to examine cystin-GFP expression in nephron segments of kidneys from R and C mice (Figure 2). AQP2 is expressed primarily on apical cell membranes of collecting duct cells.[36,37] The cystin-GFP fusion protein was detected in AQP2-positive collecting ducts of R mice (Figure 2C and I) and C mice (Figure 2B and H), while cystin-GFP was absent in the *Rosa26-Cys1-GFP* mice that do not carry *Ksp-Cre* transgene (Figure 2A and G). Co-localization of AQP2 and cystin-GFP demonstrated cystin-GFP fusion protein expression in the collecting duct cells.

Histological evaluation of cystogenesis in rescued cpk mice

The gross evaluation of kidneys from R mice suggested that the renal histology would be normal. Histological evaluation showed that, while the majority of the nephrons in R kidneys appeared to have normal dimensions, occasional cystic structures were present (Figure 3C and F). Using DBA and LTA lectins, markers of distal and proximal tubules, respectively,[38,39] we observed that the cystic structures stained with LTA (Figure 3I). These findings suggest that while expression of the cystin-GFP in collecting ducts markedly attenuated the *cpk* renal phenotype, sporadic cyst formation did occur in proximal tubular segments of these kidneys (Figure 3I).

Expression of Myc in rescued cpk mice

Myc overexpression in *cpk* kidneys is well-documented.[19-22] In previous work we demonstrated that cystin physically interacts with the DNA-binding protein necdin in a regulatory complex that binds to the *Myc* P1 promoter.[24] Necdin enhances *Myc* promoter activity and cystin antagonizes this effect. In a previous report, we proposed that *Myc* up-regulation in *cpk* kidneys results directly from disruption of the cystin-necdin interaction. In the current study, we examined the relative abundance of C-MYC protein in the kidneys of 14-day old R mice as compared to *cpk* mice. Quantitative immunoblotting revealed comparable levels of C-MYC in WT and R kidneys that were markedly lower than in kidneys of *cpk* mice (Figure 4). These results demonstrate that transgene rescue of the cystic kidney phenotype in *cpk* mice is associated with down-regulation of C-MYC protein expression, suggesting a central role for *Myc* overexpression in the renal cystogenesis in this mouse model.

A case of human ARPKD associated with homozygous CYS mutation

While the *cpk* mouse phenotype recapitulates important clinical features of ARPKD, to date no human cases of the disease have been linked to mutation of the *CYS1* gene. Recently, a 5-year-old male patient (subject B783), the offspring of consanguineous parents, was evaluated for monogenic cystic renal disease. The patient presented with polyuria and polydipsia and ultrasound examination revealed multiple medullary and cortical cysts consistent with polycystic kidney disease. Renal function was mildly reduced (creatinine 0.6 mg/dL) for a child of the subject's age. DNA samples obtained from from the proband and both parents were analyzed by trio whole exome sequencing (TRIO-WES). Homozygosity mapping of exome variant data for B783 indicated 108 Mbp of homozygosity by descent (Figure 5A), suggesting that the parents are approximately fourth degree relatives. Based on this mapping, it was hypothesized that a biallelic gene mutation residing within a homozygous peak region caused the subject's renal disease. To identify the most probable disease-causing mutation, we used the following criteria for exome

variant filtering[40-43]: (1) exclusion of all variants that did not change the amino-acid sequence or affected canonical splice sites (defined as ± 6 nucleotides surrounding the exon-intron boundary), (2) exclusion of variants reported in the homozygous state or with a minor allele frequency greater than 0.1% in a control cohort (ExAC and gnomAD genome databases), (3) inclusion of homozygous bi-allelic variants with appropriate parental segregation consistent with our above hypothesis, and (4) assessment of variants for deleteriousness based on *in silico* prediction of their impact on protein structure and/or splice site function. This approach, however, failed to yield a strong candidate mutation. Therefore, trio whole genome sequencing (TRIO-WGS) was performed in order to extend coverage to non-coding genome regions.

Employing the same filtering criteria described for TRIO-WES led to the identification of a mutation lying in a region of homozygosity of descent (Figure 5A) representing a splice-site mutation (c.318+5G>A) in the exon 1 donor site of *CYS1* (Figure 5B,C). The following criteria strongly suggest a deleterious mutation: (i) the splice variant is extremely rare, being absent from the gnomAD database (gnomAD, version 2.1.1), and (ii) four independent *in silico* prediction tools indicate a deleterious effect of the mutation on splicing (Table 1). Importantly, neither TRIO-WES nor TRIO-WGS analysis revealed causative mutations in >100 cystic kidney disease genes, including *PKHD1* and *DZIP1L*. Allele-specific PCR analysis confirmed homozygosity and heterozygosity of the mutation in the proband and both parents, respectively (Supplementary Figure S1).

In light of the above genetic findings, a further evaluation of the clinical history of subject B783 revealed mild congenital liver fibrosis, which is also consistent with the established phenotype of the *cpk* mouse.[44-46] Our findings strongly support the first identification of a causative mutation in *CYS1* in a human patient with ARPKD.

Discussion

In the current study, we demonstrate that expression of the *Cys1* transgene in renal

collecting ducts of *cpk* mice rescues the cystic phenotype and down-regulates expression of *Myc in vivo*. We also report the first ARPKD patient with a homozygous *CYS1* mutation, a c.818+5G>A variant predicted to disrupt splicing. This variant affects the +5-position of the canonical donor splice site, AG/GURAGU. Pathogenic variants affecting +5 position of the donor splice site have been reported in other genes[47-49] and detailed analysis of G-to-A sequence changes at the +5 position have revealed disrupted base pairing between donor splice site of a pre-mRNA and the U1snRNP of the spliceosome leading to decreased efficiency of the splice site recognition and exon skipping.[50]

The identification of a pathogenic *CYS1* variant in a patient with an ARPKD-like phenotype confirms the importance of the *CYS1* gene product for normal function of human collecting duct cells. While it is surprising that *CYS1* deficiency causing ARPKD has been observed in only one family, it is important to note that this gene is GC-rich, particularly its first exon (Figure S3). Such GC-rich regions can be difficult to amplify and sequence using Sanger methodology[51] and can be missed in next-generation sequencing because low sequence complexity prevents efficient capture prior to library construction.[52]

Development of collecting duct cysts in humans and mice with cystin deficiency suggests a shared pathobiology and possibly similar molecular mechanisms underlying cyst formation. In the *cpk* mouse model, renal cysts initially develop at embryonic day 15.5 (E15.5) and are restricted to proximal tubules.[53,54] As *cpk* mice develop and disease progresses, cysts predominantly affect the distal collecting duct region.[53] Specific expression of cystin-GFP in developing ureteric bud-derived collecting ducts rescued the renal cystic phenotype. Transgenic *cpk* mice expressing the fusion protein only in AQP2-positive collecting duct cells exhibited survival rates and kidney sizes similar to wild type mice. Interestingly, while collecting duct cysts were absent in rescued *cpk* mice, these animals did express proximal tubular cysts, suggesting that the initial phase of proximal tubular cystogenesis was not rescued. Proximal tubule cysts have also been observed in human ARPKD fetal specimens between 14 and 26 weeks of

8.

gestation, but not in the kidneys of fetuses older than 34 weeks of gestation.[39] These observations suggest a gradual shift of cyst formation sites from proximal tubules to collecting ducts during the early fetal development in both human and mouse ARPKD.

Cystin is a cilium-associated protein that localizes to the basal bodies and the ciliary axoneme.[17,18,55] Treatment of *cpk* mice with paclitaxel, which promotes microtubule assembly, prevents renal cyst formation, suggesting that cystin may stabilize microtubule assembly within the ciliary axoneme.[56] The primary cilia of the collecting duct epithelium function as transmitters of mechano- and chemosensory stimuli to signaling pathways that regulate multiple key cellular processes including differentiation, proliferation, apoptosis, tissue homeostasis and cell polarity.[57] We have previously demonstrated that cystin, with two functional nuclear localization signals, can be released from the ciliary membrane through a myristoyl-electrostatic switch and translocate to the nucleus where it forms a regulatory complex with necdin to modulate *Myc* expression.[24]

The *Myc* proto-oncogene plays a critical role in normal kidney development[58] and several lines of evidence suggest a central role for dysregulated *Myc* expression in the pathophysiology of polycystic kidney disease. First, overexpression of *Myc* in the kidneys of SBM transgenic mice causes polycystic kidney disease.[59] Furthermore, renal cystic disease remitted in SBM mice that underwent spontaneous reversion to normal kidney *Myc* expression.[60] Second, treatment of *cpk* mice with antisense *Myc* oligonucleotides mitigated the cystic phenotype.[22] Third, *Myc* is overexpressed in mouse models of autosomal dominant polycystic kidney disease (ADPKD)[61,62] and *Myc* expression appears to be tightly regulated by PC1, the product of the *Pkd1* gene.[63] Fourth, pharmacological inhibition of glucogen synthase kinase 3 beta (GSK3beta), which accelerates cyst formation in *cpk* mice, leads to decreased *My*c expression and amelioration of the cystic phenotype.[64] Similarly, *Myc* is down-regulated in $Cys1^{cpk/cpk};Smad3^{+/-}$ mice and such double mutants have a milder phenotype than *cpk* mice.[25] Our findings that complementation of mutant *cpk* with cystin-GFP rescues the cystic phenotype and restores normal *Myc* expression confirms that cystin acts *in vivo* as a negative regulator of *Myc*.

In summary, we demonstrate that cystin deficiency causes an ARPKD-like phenotype in mice that can be rescued by targeted renal expression of a cystin-GFP fusion protein, most probably by downregulating *Myc* expression in collecting duct cells. Our identification of the first case of human ARPKD associated with a *CYS1* mutation confirms the relevance of the *cpk* mouse as an ARPKD model yielding important insights into molecular mechanisms underlying disease pathobiology.

Methods

Animal study approvals

All mouse experiments were approved by the Institutional Animal Care and Use Committees of Children's National Research Institute and the University of Alabama at Birmingham (UAB). Knock-in transgenic mice were generated at the University of Alabama at Birmingham (UAB) Transgenic & Genetically Engineered Models Core facility. *Ksp-Cre* mice were obtained from Jackson Laboratory (Bar Harbor, ME). Mouse colonies were maintained in the animal facility at Children's National Research Institute.

Antibodies and lectins

Anti-GAPDH antibody was purchased from Cell Signaling Technologies (# 2118). Anti-AQP2 antibody was purchased from Santa Cruz Biotechnologies (# SC9882). Alexa Fluor 488 conjugated anti-GFP antibody was obtained from Life Technologies (# A21311). Polyclonal rabbit anti-cystin antibody (70053) was generated in our lab and described previously.[18] Rabbit monoclonal anti-c-Myc antibody was purchased from Abcam (# ab32072). Goat anti-rabbit HRP conjugated secondary antibody was purchased from American Qualex Solution Products (# A102PS). Donkey anti-Goat IgG Alexa Fluor 555 was obtained from Life Technologies (# A21432). Lectins LTA-FITC (# W0909) and DBA-Rhodamine (# Y0828) were obtained from Vector Laboratories.

Vector cloning

Cys1-GFP cDNA was amplified from previously described pEGFP-N1.[18] Gateway PCR primers were used to add flanking attB sites. A one-tube Gateway reaction was performed using pDonr221 and pRosa26 Dest.[65] The reaction product was used to transform competent STBL3 cells (Life Technologies, # C7373-03) that were plated on Ampicillin and Kanamycin to select for destination and entry clones, respectively. Destination clones were screened by restriction enzyme digest prior to sequencing. The pRosa26 Dest *Cys1-GFP* targeting vector was linearized with *Kpn1* and electroporated into ES cells. G418 resistant colonies were screened by long-range PCR as described[66] and positive clones were used to produce chimeric founder mice.

PCR genotyping

PCR conditions are described in Supplemental Table 1.

Immunoblotting

Kidney tissue was collected, homogenized, and processed for immunoblotting as previously described.[24] For cystin and control western blots, immuno-reactive protein bands were visualized using SuperSignal West Dura chemiluminescent substrate (Thermo Fisher Scientific, # 34076) and exposed to film. For C-MYC and control western blots, images were obtained with ChemiDoc Imaging System (Bio-Rad laboratory, Inc.) Densitometry was analyzed using Image Lab (Bio-Rad laboratory, Inc., Version 6.0).

Kidney histology

Tissue samples were collected and fixed in 10% formalin (Fisher Scientific, # 23-245-684) for 2 days, then stored in 70% ethanol. The samples were dehydrated, paraffin embedded, cut into 5 µm sections and stained with hematoxylin and eosin (H&E) and slide-mounted for examination by the UAB Comparative Pathology Laboratory.

Immunofluorescence analysis

Tissue samples were collected from 6 week-old mice and processed, using published methods.[33] Immunofluorescence detection and image acquisition were performed using an Olympus FLUOVIEW FV1000 confocal laser scanning microscope configured with both an Argon Laser

11.

(488 nm) and a Laser diode (405 nm, 440 nm, and 559 nm). Images were analyzed using Olympus FV10-ASW 3.0 Viewer software.

Lectin staining

Five μm sections of fixed paraffin embedded kidney tissues (prepared as described for histopathology) were stained with Rhodamine labeled DBA (Vector Laboratory, # RL-1032) and Fluorescein labeled LTA (Vector Laboratory, # FL-1321).[67] Immunofluorescence detection, image acquisition, and analysis were performed as described above.

Human study approval

Approval for human subjects research was obtained from Institutional Review Boards of the University of Michigan, Boston Children's Hospital, and local IRB equivalents.

Human research subjects

Blood samples and pedigrees were obtained following informed consent from individuals with cystic kidney disease or their legal guardians. The diagnosis of cystic kidney disease was based on published clinical criteria. Clinical data were obtained using a standardized questionnaire (http://www.renalgenes.org).

Whole genome sequencing and mutation calling

TRIO-WES and data processing were performed by the Genomics Platform at the Broad Institute of Harvard and MIT (Broad Institute, Cambridge, MA). Exome sequencing (>250 ng of DNA, at >2 ng/μl) was performed using Illumina exome capture (38 Mb target). Single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) were jointly called across all samples using the Genome Analysis Toolkit (GATK) HaplotypeCaller. Default filters were applied to SNP and indel calls using the GATK Variant Quality Score Recalibration approach. Lastly, variants were annotated using the Variant Effect Predictor. For additional information, refer to the Supporting Information Section S1 in the exome aggregation consortium (ExAC) study.[68] The variant call set was uploaded on to Seqr (https://seqr.broadinstitute.org) and analysis of the entire WES output was performed. TRIO-WGS and data processing were performed by the Genomics

Platform at the Broad Institute of MIT and Harvard. PCR-free preparation of sample DNA (350 ng input at >2 ng/ul) was accomplished using Illumina HiSeq X Ten v2 chemistry. Libraries were sequenced to a mean target coverage of >30x. Genome sequencing data was processed through a pipeline based on Picard, using base quality score recalibration and local realignment at known indels. The BWA aligner was used for mapping reads to the human genome build 38. Single Nucleotide Variants (SNVs) and insertions/deletions (indels) were jointly called across all samples using Genome Analysis Toolkit (GATK) HaplotypeCaller package version 3.4. Default filters were applied to SNV and indel calls using the GATK Variant Quality Score Recalibration (VQSR) approach. Annotation was performed using Variant Effect Predictor (VEP). Lastly, the variant call set was uploaded to seqr for collaborative analysis between the CMG and investigator.

Mutation calling was performed in line with proposed guidelines,[41] and the following criteria were employed as previously described[42,43]. The variants included were rare in the population with mean allele frequency <0.1% and with 0 homozygotes in the adult reference genome databases ExAC and gnomAD. Additionally, variants were non-synonymous and/or located within splice-sites. Based on an autosomal homozygous recessive hypothesis, homozygous variants were evaluated. Subsequently, variant severity was classified based on prediction of protein impact (truncating frameshift or nonsense mutations, essential or extended splice-site mutations, and missense mutations). Splice-site mutations were assessed by *in silico* tools MaxEnt, NNSPLICE, HSF, and CADD splice-site mutation prediction scores.[69-72] Missense mutations were assessed based on SIFT, MutationTaster and PolyPhen 2.0 conservation prediction scores[73-75] and evolutionary conservation based on manually derived multiple sequence alignments.

Homozygosity mapping (HM)

Homozygosity mapping was calculated based on whole exome sequencing data. In brief, aligned BAM files were processed using Picard and SAMtools4 as described previously.[76] Single nucleotide variant calling was performed using Genome Analysis Tool Kit (GATK).[77] The resulting VCF files were used to generate homozygosity mapping data and visual outputs using the

program Homozygosity Mapper.[78]

Web resources

UCSC Genome Browser, genome.ucsc.edu

Ensembl Genome Browser, www.ensembl.org

gnomAD browser 2.0.3., gnomad.broadinstitute.org

Polyphen2, genetics.bwh.harvard.edu/pph2

Sorting Intolerant From Tolerant (SIFT), sift.jcvi.org

MutationTaster, www.mutationtaster.org

Combined Annotation Dependent Depletion, cadd.gs.washington.edu

NNSPLICE splice-site mutation prediction, www.fruitfly.org/seq_tools/splice.html

MaxEnt splice prediction, hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html

Human Splice Finder, www.umd.be/HSF/

Disclosures

F.H. is a co-founder of Goldfinch Biopharma Inc.

Acknowledgements

15.

References

1. Alzarka, B., et al., *Design and Implementation of the Hepatorenal Fibrocystic Disease Core Center Clinical Database: A Centralized Resource for Characterizing Autosomal Recessive Polycystic Kidney Disease and Other Hepatorenal Fibrocystic Diseases.* Front Pediatr, 2017. **5**: p. 80.

2. Szabo, T., et al., *Comprehensive genetic testing in children with a clinical diagnosis of ARPKD identifies phenocopies.* Pediatr Nephrol, 2018. **33**(10): p. 1713-1721.

3. Melchionda, S., et al., *Expanding the mutation spectrum in 130 probands with ARPKD: identification of 62 novel PKHD1 mutations by sanger sequencing and MLPA analysis.* J Hum Genet, 2016. **61**(9): p. 811-21.

4. Adeva, M., et al., *Clinical and molecular characterization defines a broadened spectrum of autosomal recessive polycystic kidney disease (ARPKD).* Medicine (Baltimore), 2006. **85**(1): p. 1-21.

5. Bergmann, C., et al., *Clinical consequences of PKHD1 mutations in 164 patients with autosomal-recessive polycystic kidney disease (ARPKD).* Kidney international, 2005. **67**: p. 829-848.

6. Lu, H., et al., *Mutations in DZIP1L, which encodes a ciliary-transition-zone protein, cause autosomal recessive polycystic kidney disease.* Nat Genet, 2017. **49**(7): p. 1025-1034.

7. Bergmann, C., *Early and Severe Polycystic Kidney Disease and Related Ciliopathies: An Emerging Field of Interest.* Nephron, 2019. **141**(1): p. 50-60.

8. Moser, M., et al., *A mouse model for cystic biliary dysgenesis in autosomal recessive polycystic kidney disease (ARPKD).* Hepatology, 2005. **41**(5): p. 1113-21.

9. Garcia-Gonzalez, M.A., et al., *Genetic interaction studies link autosomal dominant and recessive polycystic kidney disease in a common pathway.* Hum Mol Genet, 2007. **16**(16): p. 1940-50.

10. Woollard, J.R., et al., *A mouse model of autosomal recessive polycystic kidney disease with biliary duct and proximal tubule dilatation.* Kidney Int, 2007. **72**(3): p. 328-36.

11. Gallagher, A.R., et al., *Biliary and pancreatic dysgenesis in mice harboring a mutation in Pkhd1.* Am J Pathol, 2008. **172**(2): p. 417-29.

12. Kim, I., et al., *Fibrocystin/polyductin modulates renal tubular formation by regulating polycystin-2 expression and function.* J Am Soc Nephrol, 2008. **19**(3): p. 455-68.

13. Williams, S.S., et al., *Kidney cysts, pancreatic cysts, and biliary disease in a mouse model of autosomal recessive polycystic kidney disease.* Pediatr Nephrol, 2008. **23**(5): p. 733-41.

14. Bakeberg, J.L., et al., *Epitope-tagged Pkhd1 tracks the processing, secretion, and localization of fibrocystin.* J Am Soc Nephrol, 2011. **22**(12): p. 2266-77.

15. Outeda, P., et al., *A novel model of autosomal recessive polycystic kidney questions the role of the fibrocystin C-terminus in disease mechanism.* Kidney Int, 2017. **92**(5): p. 1130-1144.

16. Nagao, S., et al., *Animal models for human polycystic kidney disease.* Exp Anim, 2012. **61**(5): p. 477-88.

17. Hou, X., et al., *Cystin, a novel cilia-associated protein, is disrupted in the cpk mouse model of polycystic kidney disease.* J Clin Invest, 2002. **109**(4): p. 533-40.

18. Tao, B., et al., *Cystin localizes to primary cilia via membrane microdomains and a targeting motif.* J Am Soc Nephrol, 2009. **20**(12): p. 2570-80.

19. Cowley, B.D., Jr., et al., *Elevated c-myc protooncogene expression in autosomal recessive polycystic kidney disease.* Proc Natl Acad Sci U S A, 1987. **84**(23): p. 8394-8.

20. Cowley, B.D., Jr., et al., *Elevated proto-oncogene expression in polycystic kidneys of the C57BL/6J (cpk) mouse.* J Am Soc Nephrol, 1991. **1**(8): p. 1048-53.

21. Harding, M.A., et al., *Localization of overexpressed c-myc mRNA in polycystic kidneys of the cpk mouse.* Kidney Int, 1992. **41**(2): p. 317-25.

22.   Ricker, J.L., et al., *c-myc antisense oligonucleotide treatment ameliorates murine ARPKD.* Kidney Int, 2002. **61**(1 Suppl): p. S125-31.

23.   Gattone, V.H., 2nd, et al., *Autosomal recessive polycystic kidney disease in a murine model. A gross and microscopic description.* Lab Invest, 1988. **59**(2): p. 231-8.

24.   Wu, M., et al., *The Ciliary Protein Cystin Forms a Regulatory Complex with Necdin to ModulateMyc Expression.* PLoS One, 2013. **8**(12): p. e83062.

25.   Hama, T., et al., *Aberrant Smad3 phosphoisoforms in cyst-lining epithelial cells in the cpk mouse, a model of autosomal recessive polycystic kidney disease.* Am J Physiol Renal Physiol, 2017. **313**(6): p. F1223-F1231.

26.   Nakamura, T., et al., *Growth factor gene expression in kidney of murine polycystic kidney disease.* J Am Soc Nephrol, 1993. **3**(7): p. 1378-86.

27.   Orellana, S.A., et al., *Epidermal growth factor receptor expression is abnormal in murine polycystic kidney.* Kidney Int, 1995. **47**(2): p. 490-9.

28.   Ebihara, I., et al., *Altered mRNA expression of basement membrane components in a murine model of polycystic kidney disease.* Lab Invest, 1988. **58**(3): p. 262-9.

29.   Taub, M., et al., *Altered basement membrane protein biosynthesis by primary cultures of cpk/cpk mouse kidney.* Kidney Int, 1990. **37**(4): p. 1090-7.

30.   Rankin, C.A., et al., *Matrix metalloproteinases and TIMPS in cultured C57BL/6J-cpk kidney tubules.* Kidney Int, 1996. **50**(3): p. 835-44.

31.   Harding, M.A., et al., *The SGP-2 gene is developmentally regulated in the mouse kidney and abnormally expressed in collecting duct cysts in polycystic kidney disease.* Dev Biol, 1991. **146**(2): p. 483-90.

32.   Rocco, M.V., et al., *Attenuated expression of epithelial cell adhesion molecules in murine polycystic kidney disease.* Am J Physiol, 1992. **262**(4 Pt 2): p. F679-86.

33.   Shao, X., S. Somlo, and P. Igarashi, *Epithelial-specific Cre/lox recombination in the developing kidney and genitourinary tract.* J Am Soc Nephrol, 2002. **13**(7): p. 1837-46.

34. Shao, X., et al., *A minimal Ksp-cadherin promoter linked to a green fluorescent protein reporter gene exhibits tissue-specific expression in the developing kidney and genitourinary tract.* J Am Soc Nephrol, 2002. **13**(7): p. 1824-36.

35. Mandell, J., et al., *Congenital polycystic kidney disease. Genetically transmitted infantile polycystic kidney disease in C57BL/6J mice.* Am J Pathol, 1983. **113**(1): p. 112-4.

36. Chou, C.L., et al., *Regulation of aquaporin-2 trafficking by vasopressin in the renal collecting duct. Roles of ryanodine-sensitive Ca2+ stores and calmodulin.* J Biol Chem, 2000. **275**(47): p. 36839-46.

37. Elliot, S., et al., *Urinary excretion of aquaporin-2 in humans: a potential marker of collecting duct responsiveness to vasopressin.* J Am Soc Nephrol, 1996. **7**(3): p. 403-9.

38. Holthofer, H., *Lectin binding sites in kidney. A comparative study of 14 animal species.* J Histochem Cytochem, 1983. **31**(4): p. 531-7.

39. Nakanishi, K., et al., *Proximal tubular cysts in fetal human autosomal recessive polycystic kidney disease.* Journal of the American Society of Nephrology, 2000. **11**(4): p. 760-3.

40. Lovric, S., et al., *Genetic testing in steroid-resistant nephrotic syndrome: when and how?* Nephrol Dial Transplant, 2016. **31**(11): p. 1802-1813.

41. MacArthur, D.G., et al., *Guidelines for investigating causality of sequence variants in human disease.* Nature, 2014. **508**(7497): p. 469-76.

42. van der Ven, A.T., et al., *Whole-Exome Sequencing Identifies Causative Mutations in Families with Congenital Anomalies of the Kidney and Urinary Tract.* J Am Soc Nephrol, 2018. **29**(9): p. 2348-2361.

43. Vivante, A. and F. Hildebrandt, *Exploring the genetic basis of early-onset chronic kidney disease.* Nat Rev Nephrol, 2016. **12**(3): p. 133-46.

44. Gattone, V.H., 2nd, K.A. MacNaughton, and A.L. Kraybill, *Murine autosomal recessive polycystic kidney disease with multiorgan involvement induced by the cpk gene.* Anat Rec, 1996. **245**(3): p. 488-99.

45.   Ricker, J.L., et al., *Development of autosomal recessive polycystic kidney disease in BALB/c-cpk/cpk mice.* J Am Soc Nephrol, 2000. **11**(10): p. 1837-47.

46.   Fry, J.L., Jr., et al., *A genetically determined murine model of infantile polycystic kidney disease.* J Urol, 1985. **134**(4): p. 828-33.

47.   De Conti, L., et al., *Complexities of 5'splice site definition: implications in clinical analyses.* RNA Biol, 2012. **9**(6): p. 911-23.

48.   Hori, T., et al., *Molecular basis of two-exon skipping (exons 12 and 13) by c.1248+5g>a in OXCT1 gene: study on intermediates of OXCT1 transcripts in fibroblasts.* Hum Mutat, 2013. **34**(3): p. 473-80.

49.   Poudrier, J., et al., *Frequency of the IVS12 + 5G-->A splice mutation of the fumarylacetoacetate hydrolase gene in carriers of hereditary tyrosinaemia in the French Canadian population of Saguenay-Lac-St-Jean.* Prenat Diagn, 1996. **16**(1): p. 59-64.

50.   BinEssa, H.A., et al., *Functional analysis of 22 splice-site mutations in the PHEX, the causative gene in X-linked dominant hypophosphatemic rickets.* Bone, 2019. **125**: p. 186-193.

51.   Benita, Y., et al., *Regionalized GC content of template DNA as a predictor of PCR success.* Nucleic Acids Res, 2003. **31**(16): p. e99.

52.   Meienberg, J., et al., *New insights into the performance of human whole-exome capture platforms.* Nucleic Acids Res, 2015. **43**(11): p. e76.

53.   Avner, E.D., et al., *Congenital murine polycystic kidney disease. I. The ontogeny of tubular cyst formation.* Pediatr Nephrol, 1987. **1**(4): p. 587-96.

54.   Nauta, J., et al., *Renal and biliary abnormalities in a new murine model of autosomal recessive polycystic kidney disease.* Pediatr Nephrol, 1993. **7**(2): p. 163-72.

55.   Yoder, B.K., X. Hou, and L.M. Guay-Woodford, *The polycystic kidney disease proteins, polycystin-1, polycystin-2, polaris, and cystin, are co-localized in renal cilia.* J Am Soc Nephrol, 2002. **13**(10): p. 2508-16.

56.   Woo, D.D., A.P. Tabancay, Jr., and C.J. Wang, *Microtubule active taxanes inhibit polycystic kidney disease progression in cpk mice.* Kidney Int, 1997. **51**(5): p. 1613-8.

57.   Kathem, S.H., A.M. Mohieldin, and S.M. Nauli, *The Roles of Primary cilia in Polycystic Kidney Disease.* AIMS Mol Sci, 2014. **1**(1): p. 27-46.

58.   Trudel, M., *c-Myc Signalling in the Genetic Mechanism of Polycystic Kidney Disease*, in *Polycystic Kidney Disease*, X. Li, Editor. 2015: Brisbane (AU).

59.   Trudel, M., et al., *Polycystic kidney disease in SBM transgenic mice: role of c-myc in disease induction and progression.* Am J Pathol, 1998. **152**(1): p. 219-29.

60.   Trudel, M., N. Chretien, and V. D'Agati, *Disappearance of polycystic kidney disease in revertant c-myc transgenic mice.* Mamm Genome, 1994. **5**(3): p. 149-52.

61.   Burtey, S., et al., *Overexpression of PKD2 in the mouse is associated with renal tubulopathy.* Nephrol Dial Transplant, 2008. **23**(4): p. 1157-65.

62.   Kurbegovic, A. and M. Trudel, *Progressive development of polycystic kidney disease in the mouse model expressing Pkd1 extracellular domain.* Hum Mol Genet, 2013. **22**(12): p. 2361-75.

63.   Parrot, C., et al., *c-Myc is a regulator of the PKD1 gene and PC1-induced pathogenesis.* Hum Mol Genet, 2019. **28**(5): p. 751-763.

64.   Tao, S., et al., *Glycogen synthase kinase-3beta promotes cyst expansion in polycystic kidney disease.* Kidney Int, 2015. **87**(6): p. 1164-75.

65.   Hohenstein, P., et al., *High-efficiency Rosa26 knock-in vector construction for Cre-regulated overexpression and RNAi.* Pathogenetics, 2008. **1**(1): p. 3.

66.   O'Connor, A.K., et al., *An inducible CiliaGFP mouse model for in vivo visualization and analysis of cilia in live tissue.* Cilia, 2013. **2**(1): p. 8.

67.   Patel, V., et al., *Acute kidney injury and aberrant planar cell polarity induce cyst formation in mice lacking renal cilia.* Hum Mol Genet, 2008. **17**(11): p. 1578-90.

68. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.

69. Reese, M.G., et al., *Improved splice site detection in Genie.* J Comput Biol, 1997. **4**(3): p. 311-23.

70. Yeo, G. and C.B. Burge, *Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals.* J Comput Biol, 2004. **11**(2-3): p. 377-94.

71. Desmet, F.O., et al., *Human Splicing Finder: an online bioinformatics tool to predict splicing signals.* Nucleic Acids Res, 2009. **37**(9): p. e67.

72. Rentzsch, P., et al., *CADD: predicting the deleteriousness of variants throughout the human genome.* Nucleic Acids Res, 2019. **47**(D1): p. D886-D894.

73. Sim, N.L., et al., *SIFT web server: predicting effects of amino acid substitutions on proteins.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W452-7.

74. Schwarz, J.M., et al., *MutationTaster2: mutation prediction for the deep-sequencing age.* Nat Methods, 2014. **11**(4): p. 361-2.

75. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.

76. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

77. Van der Auwera, G.A., et al., *From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.* Curr Protoc Bioinformatics, 2013. **43**: p. 11 10 1-11 10 33.

78. Seelow, D., et al., *HomozygosityMapper--an interactive approach to homozygosity mapping.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W593-9.

Figure legends

Figure 1. (**A**) Schematic diagram showing the *Cys1-GFP* transgene knock-in at the *Rosa26* locus (*Rosa26-Cys1-GFP* allele) before (T^{OFF}) and after (T^{ON}) deletion of a PGK Neo cassette (yellow rectangle) flanked by LoxP sites (gray triangles). In the T^{OFF} configuration, expression of *Cys1-GFP* is prevented by the PGK Neo cassette. In cells expressing a *Ksp-Cre* transgene, Cre-mediated recombination deletes PGK Neo and *Cys1-GFP* is expressed (T^{ON}). SA: splice acceptor, PKG-Neo: Phosphoglycerate kinase promoter driving a neomycin resistance gene followed by 3 polyA signals (3XpA, red rectangle). The purple boxes flanking *Cys1-GFP* are attB sites. (**B)** PCR-based genotyping of *Rosa26-Cys1-GFP*, *Ksp-Cre* and *Cys1* alleles: lane 1 – *Cys1$^{cpk/+}$;Rosa26-Cys1-GFP/+* mice; lane 2 and 4 – *Cys1$^{cpk/cpk}$;Rosa26-Cys1-GFP; Ksp-Cre/+* mice; lanes 3 and 5 – *Cys$^{+/+}$;Rosa26-Cys1-GFP;Ksp-Cre/+* mice. (**C**) *Cys1-GFP* rescue of gross phenotypes in *cpk* mice. Six week-old wild-type (WT) and *Cys1$^{cpk/cpk}$;Rosa26-Cys1-GFP; Ksp-Cre/+* rescued (R) mice are of equivalent size. Examination of kidneys from WT, R and *cpk* mice at 14 and 21 days of age show equivalent sizes of R and WT kidneys, with both markedly smaller than *cpk* kidneys. (**D**) Western blot analysis of total kidney protein from 6 week-old mice: lane 1 – C57Bl/6 wild-type mice; lane 2 – *Cys$^{cpk/+}$;Rosa26-Cys1-GFP* mice; lane 3 – R mice; lane 4 – C mice. Mouse cystin is 145 amino acids long but migrates aberrantly at ~25 kDa on SDS-PAGE. Cystin-GFP (arrow, ~50 kDa) and endogenous cystin (arrowhead, ~25 kDa) were detected using polyclonal rabbit anti-cystin antibody. GAPDH served as an internal protein loading and transfer control. The asterisk indicates non-specific bands.

Figure 2. Immunohistochemical detection of cystin-GFP fusion protein (green, **A-C**), AQP2 (red, **D-F**) and merged (**G-I**) in kidney tissues from 6 week-old mice of the indicated genotypes. The *Rosa26-Cys1-GFP/+* mice (**A,D,G**) do not express cystin-GFP fusion protein due to the absence of the *Ksp-Cre* transgene. The *Rosa26-Cys1-GFP/+; Ksp-Cre/+* mice (**B,E,H**) have a wild-type Cys1 gene and express cystin-GFP fusion protein in Cre-positive cells. The *cpk/cpk; Rosa26-*

Cys1-GFP/+;Ksp-Cre/+ rescue mice (**C,F,I**) express cystin-GFP fusion protein in Cre-positive cells. Cell nuclei are stained with DAPI (blue). Scale bars equal 10 µm. Images are representative of tissue sections from 3 animals.

Figure 3. Kidney histology and differential lectin staining. Formalin fixed, paraffin embedded 5 µm sections of kidney tissues from 6 week-old mice of the indicated genotypes were stained with H&E and examined by light microscopy (**A-F**) or stained with lectins LTA (green, proximal tubules) and DBA (red, distal tubules) and examined by immunofluorescence microscopy (**G,H,I**). Panels A-C are 5X magnification with scale bars equal to 100 µm. Boxed areas are shown at 40X magnification in corresponding panels D-F with scale bars equal to 20 µm. Lectin staining was performed on serial sections corresponding to H&E stained samples. Cell nuclei were stained with DAPI. Scale bars in panels G-I are equal to 10 µm. The asterisk (*) identifies a large cyst that was not stained by either LTA or DBA. Images are representative of tissue sections from 3 animals.

Figure 4. Immunoblot analysis of C-MYC protein expression in kidneys from wild-type (WT), *cpk/cpk*;*Rosa26-Cys1-GFP/+* (*cpk*) and *cpk/cpk*;*Rosa26-Cys1-GFP/+;Ksp-Cre/+* (Rescue) mice. (**A**) Total kidney lysates were immunoblotted using anti-C-MYC and anti-GAPDH antibodies. (**B**) C-MYC band intensity was normalized with GAPDH. Data represents mean ± S.E; Y-axis values indicate C-MYC/GAPDH band intensity ratio; **$P < 0.01$, vs. others, ANOVA, n = 3 each group.

Figure 5. Whole genome sequencing identified a recessive mutation in the gene *CYS1* in one family with polycystic kidneys and liver fibrosis. (**A**) Genome-wide homozygosity mapping in individual B783 identifies homozygous peak regions. The gene locus for *CYS1* is located within a peak region of chromosome 2 (arrowhead). (**B**) Exon structure of *CYS1* cDNA. Positions of start codon , stop codons, and of the affected splice site (boundary exon 1-2) are indicated. The region within exon 1 deleted in the *cpk* mouse is also indicated. (**C**) Domain structure of the encoded

cystin-1 protein, which contains a myristoylation site (M), polybasic region (P) and AxEGG motif required for targeting cystin-1 to the cilium (C). The arrow indicates the boundary between protein regions encoded by exons 1 and 2. Abbreviations: AA, amino acid; BP, base pairs; C, cilium trafficking domain; M, myristoylation site; P, polybasic region.

Supplementary Figure S1: Confirmation of *CYS1* mutation in family B783 with childhood onset polycystic kidney disease and liver fibrosis. (**A**) Map showing primer location in context of *CYS1* variant for allele-specific primer PCR. (**B**) Table shows shared forward primer and allele specific reverse primers for the wildtype (WT) and mutated nucleotide (MUT). (**C**) Gel electrophoresis of *CYS1* allele-specific PCR for WT and MUT alleles from DNA of the proband (21), father (11), and mother (12) in family B783 is shown. PCR products were noted at the expected size (128 bp) for both parents for the WT PCR, while the WT PCR in the affected child yielded only non-specific bands. The MUT PCR yielded a single product of expected size in all three family members. This suggested the parents were heterozygous for the MUT allele, while the affected proband is homozygous. (**D**) The PCR products were excised at 128 bp for each of the 6 reactions in (C) and were sequenced using the Sanger approach. The sequences derived from the chromatograms were aligned to the *CYS1* locus as shown. The WT and MUT products from parental DNA (11 and 12) and the MUT product for the affected proband (21) confirmed as amplification products of this locus. The product from WT PCR of the affected proband (21) DNA did not align. (**E**) The chromatogram sequence from the WT PCR for the proband was aligned to the human genome using UCSC Browser BLAT. This yielded only one search result, indicating 98.3% similarity with an intron in chromosome 12. This supports that the WT PCR from the DNA of the affected proband yielded non-specific products, while the others yielded only the *CYS1* locus. Abbreviations: 11, father; 12, mother, 21, proband; WT, wildtype; MUT, mutant.

25.

Table 1. *CYS1* mutation in one family with an ARPKD phenotype.

Family	Change	Affect	Exon (Zyg, Seg)	*In silico* Severity Scores	gnomAD (H/h/T)	Sex	Ethnic origin	PC	Clinical Phenotype
B783	c.318+5 G>A	Splice	1 (Hom, B)	CADD 17.06[a] HSF -13.17[bc] MaxEnt -62.8%[c] NNS -71.3%[c]	0/0/ ~251,000	M	Egypt	Y	**Initial Onset: 5 years, polyuria, polydipsia Serum Studies: Cr 0.6 mg/dL RUS: bilateral cortical and medullary cysts Extra-renal: Mild congenital hepatic fibrosis**

Abbreviations: B, both parents appropriately heterozygous; CADD, Combined Annotation-Dependent Depletion prediction score; Cr, serum creatinine; gnomAD, Genome Aggregation database; H, homozygotes in gnomAD; h, heterozygous alleles in gnomAD; Hom, homozygous zygosity; HSF, HSF splice prediction score; M, male; MaxEnt, MaxEnt splice prediction score; NNS, NNSPLICE splice-site mutation prediction score; PC, parental consanguinity; RUS, renal ultrasound; Seg, segregation; T, total alleles in gnomAD; Y, yes; Zyg, zygosity.

[a]CADD scores between 10-20 are in the 1-10% most deleterious substitutions possible in human genome in terms of predicted effect on the gene product.
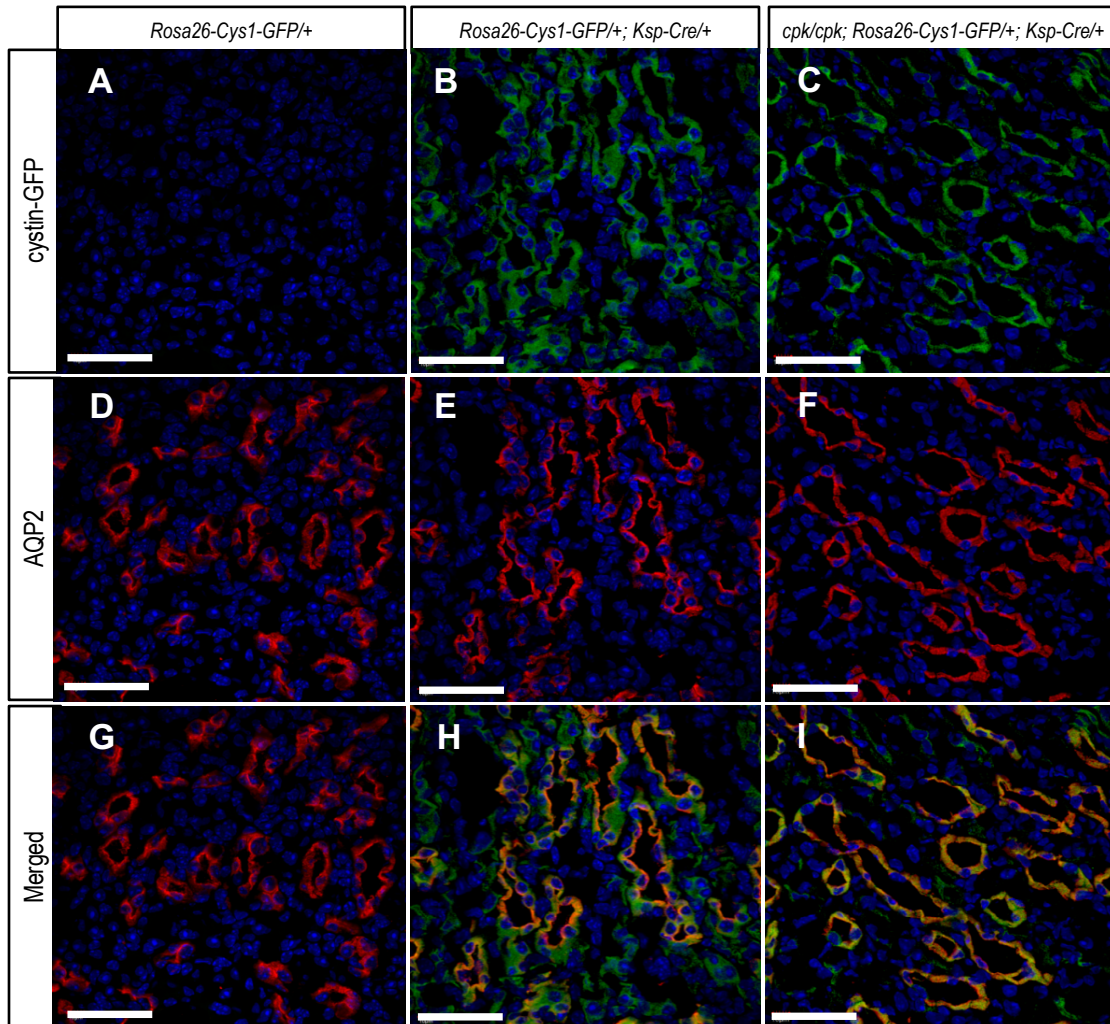
[b]Reports wildtype site broken by this change.

[c]Percentage reflects expected reduction in splicing at this donor site as consequence of change.
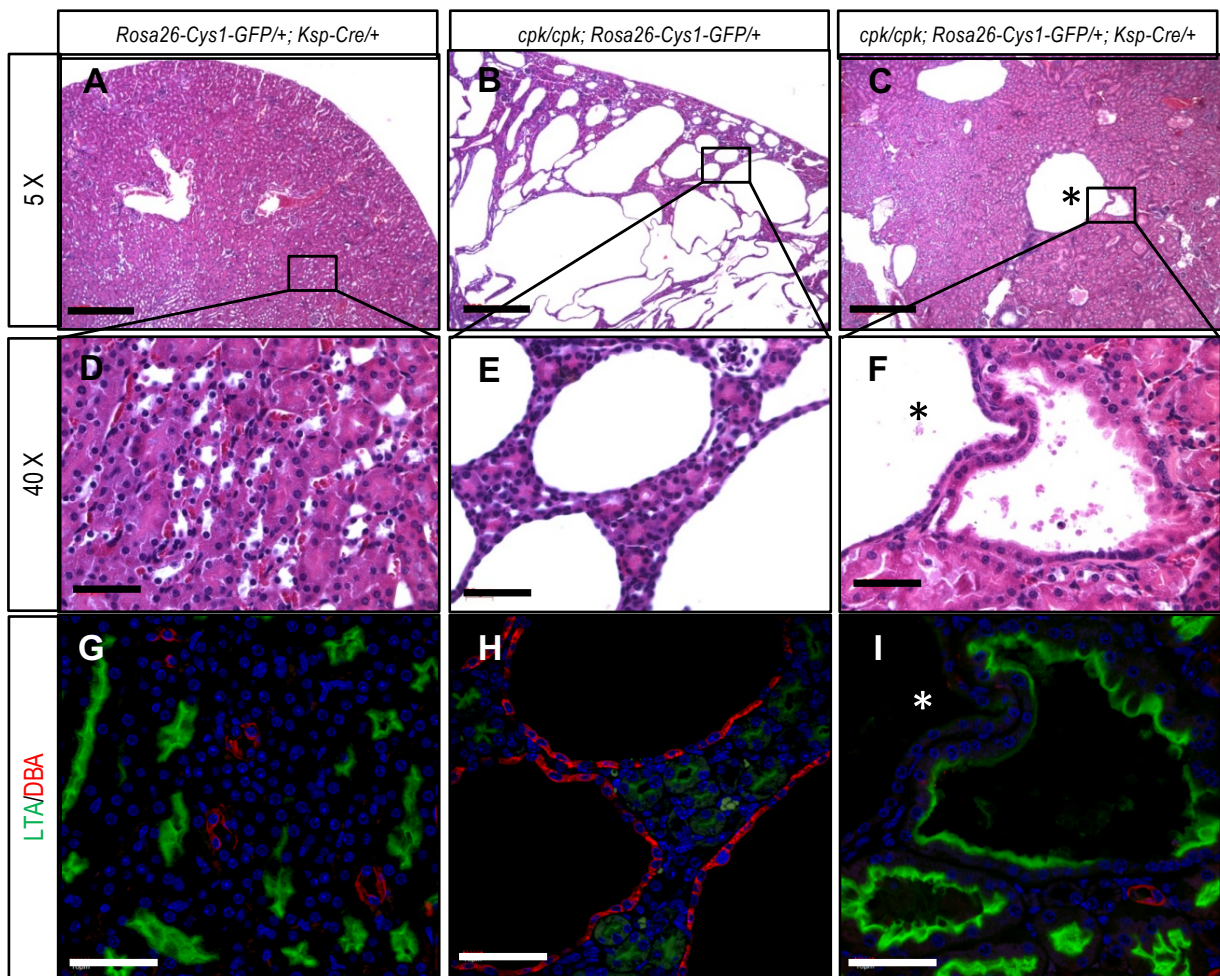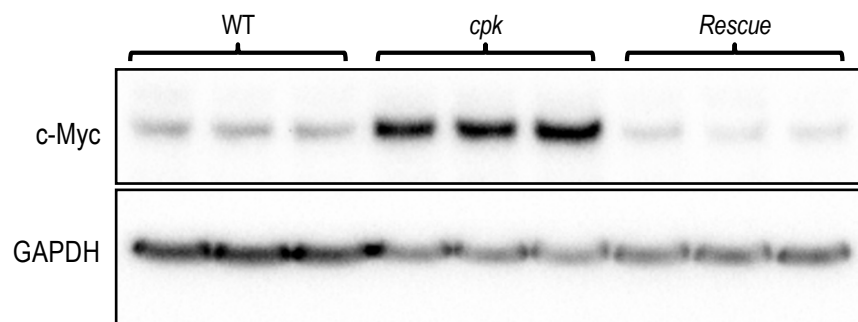
Figure 1
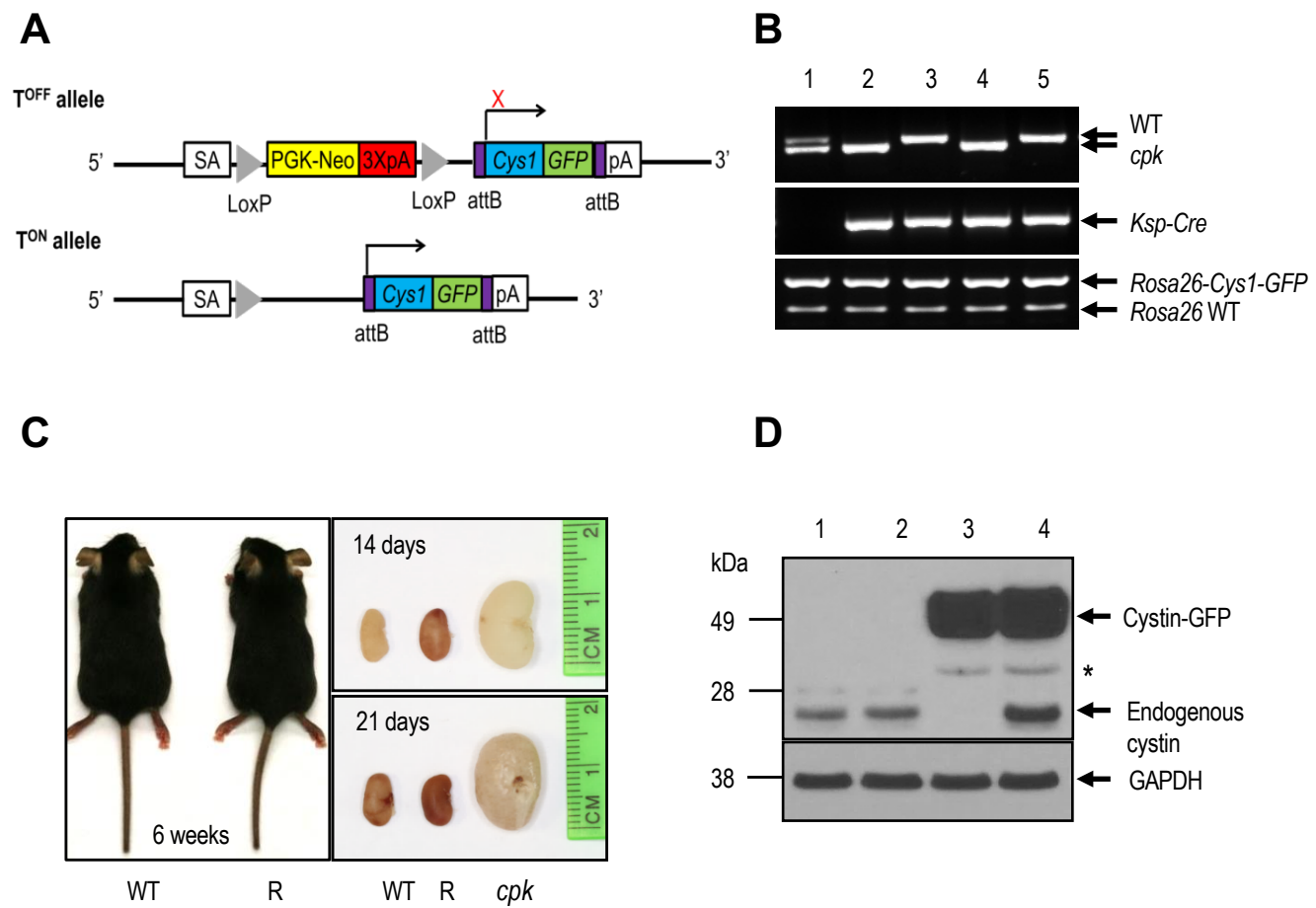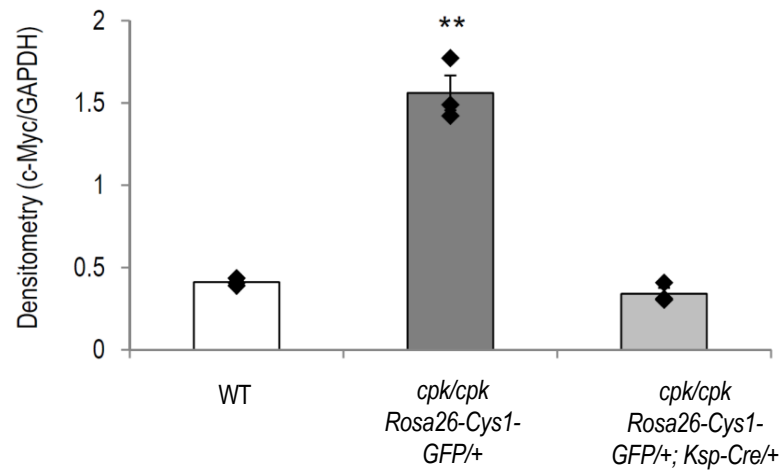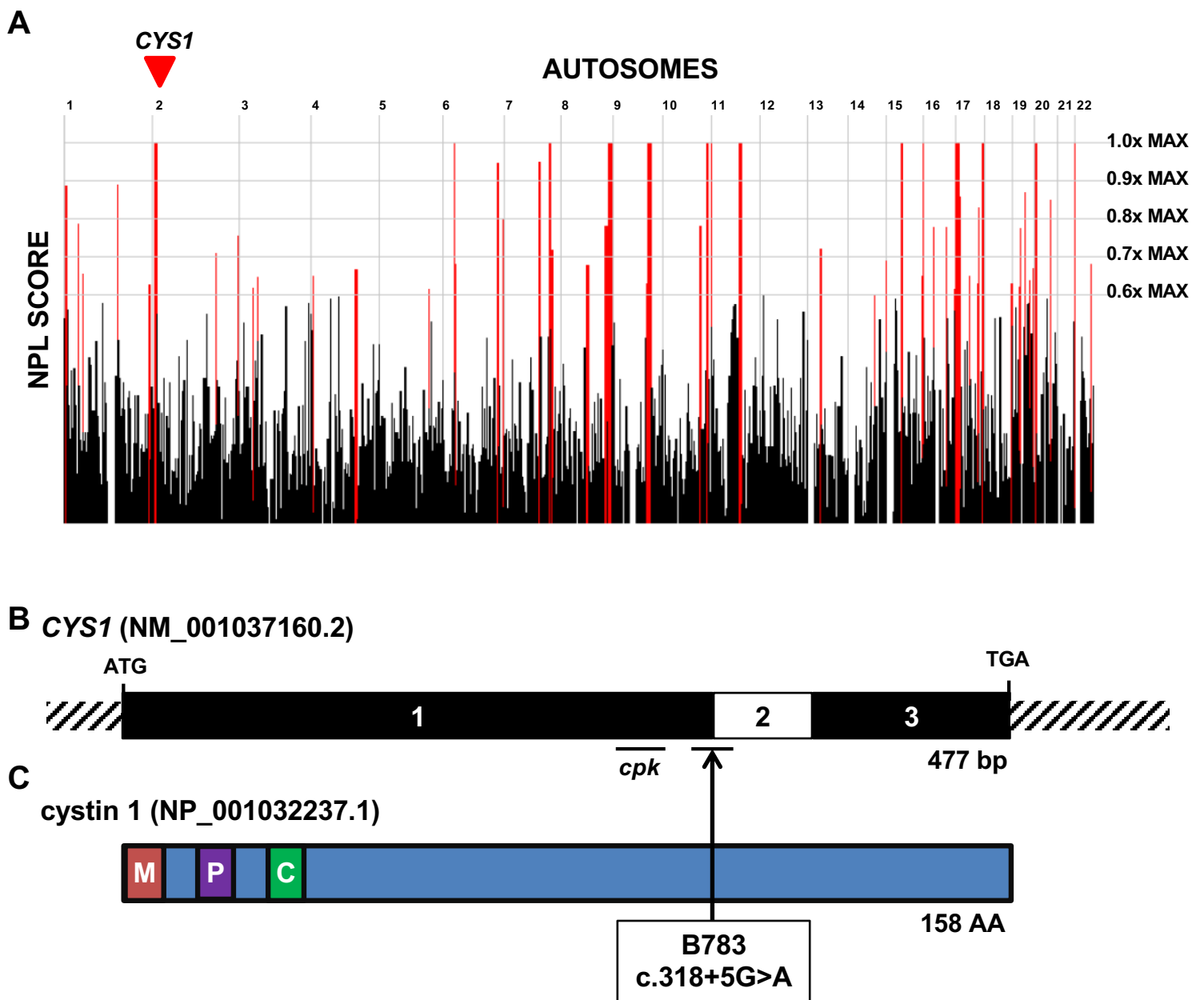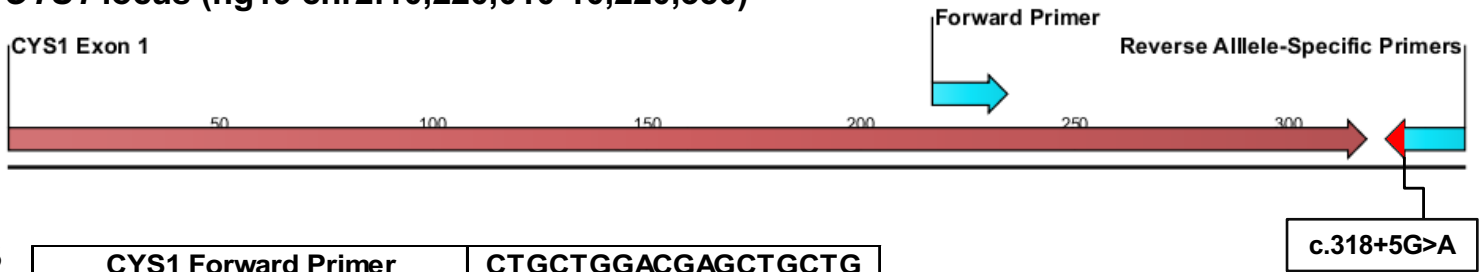
Figure 2

Figure 3

Figure 4

A

B

Figure 5

A

B *CYS1* (NM_001037160.2)

C cystin 1 (NP_001032237.1)

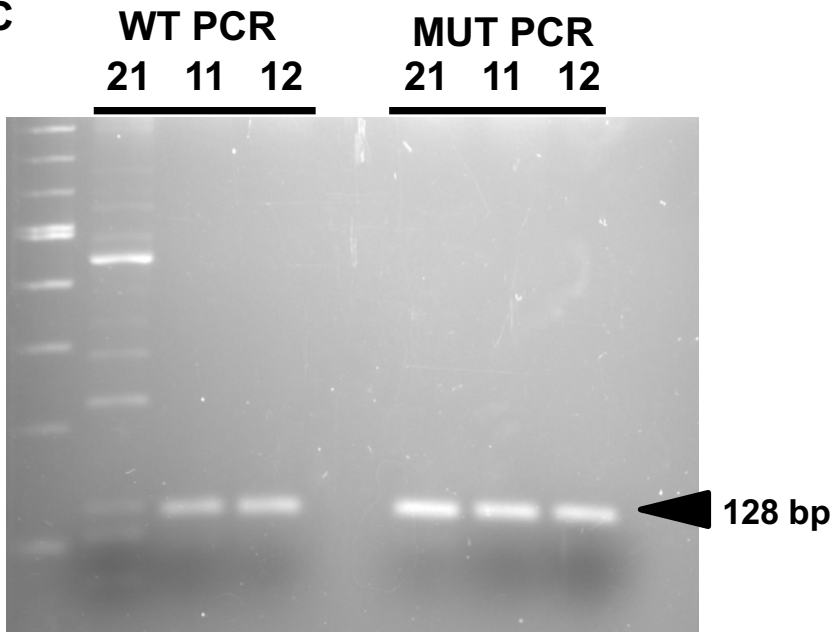B783
c.318+5G>A

Supplemental Figure

A

CYS1 locus (hg19 chr2:10,220,010-10,220,350)

B

CYS1 Forward Primer	CTGCTGGACGAGCTGCTG
CYS1 Reverse WT Primer	GCCCTGCGGCTCCCACAC
CYS1 Reverse MUT Primer	GCCCTGCGGCTCCCACAT

C

D

E

21 WT PCR PRODUCT