Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Lopez-Rivera E, Liu YP, Verbitsky M, et al. Genetic drivers of kidney defects in the DiGeorge syndrome. N Engl J Med 2017;376:742-54. DOI: 10.1056/NEJMoa1609009

SUPPLEMENTARY APPENDIX

Genetic dissection of the renal disease of DiGeorge syndrome

Esther Lopez-Rivera[1*], Yangfan P. Liu[2*], Miguel Verbitsky[1*], Blair R. Anderson[2*], Valentina P. Capone[1*], Edgar A. Otto[3], Zhonghai Yan[1], Adele Mitrotti[1], Jeremiah Martino[1], Nicholas J. Steers[1], David A. Fasel[1], Katarina Vukojevic[4], Rong Deng[1], Silvia E. Racedo[5], Qingxue Liu[6], Max Werth[1], Rik Westland[7], Asaf Vivante[8], Gabriel S. Makar[1], Monica Bodria[1,9,10], Matthew G. Sampson[11], Christopher E. Gillies[11], Virginia Vega-Warner[11], Mariarosa Maiorana[10], Donald S. Petrey[12], Barry Honig[13], Vladimir J. Lozanovski[14], Rémi Salomon[15,16], Laurence Heidet[15], Wassila Carpentier[17], Dominique Gaillard[18], Alba Carrea[9], Loreto Gesualdo[19], Daniele Cusi[20], Claudia Izzi[21], Francesco Scolari[22], Joanna A.E. van Wijk[7], Adela Arapovic[23], Mirna Saraga-Babic[4], Marijan Saraga[4,23], Nenad Kunac[24], Ali Samii[25], Donna M. McDonald-McGinn[26], Terrence B. Crowley[26], Elaine H. Zackai[26], Dorota Drosdz[27], Monika Miklaszewska[28], Marcin Tkaczyk[29], Przemyslaw Sikora[30], Maria Szczepanska[31], Malgorzata Mizerska-Wasiak[32], Grazyna Krzemien[32], Agnieszka Szmigielska[32], Marcin Zaniew[33], John M. Darlow[34], Prem Puri[35], David Barton[36], Emilio Casolari[37], Susan L. Furth[38], Bradley A. Warady[39], Zoran Gucev[40], Hakon Hakonarson[41], Hana Flogelova[42], Velibor Tasic[40], Anna Latos Bielenska[43], Anna Materna-Kiryluk[43], Landino Allegri[10], Craig S. Wong[44], Iain A. Drummond[45], Vivette D'Agati[46], Akira Imamoto[47], Jonathan M. Barasch[1], Friedhelm Hildebrandt[8], Krzysztof Kiryluk[1], Richard P. Lifton[48], Bernice E. Morrow[5], Cecile Jeanpierre[15], Virginia E. Papaioannou[6], Gian Marco Ghiggeri[9], Ali G. Gharavi[1], Nicholas Katsanis[2$], and Simone Sanna-Cherchi[1$]

[1]Division of Nephrology, Columbia University, New York, USA.
[2]Center for Human Disease Modeling, Duke University, Durham, North Carolina, USA.
[3]University of Michigan School of Medicine, Department of Internal Medicine-Nephrology, Ann Arbor, MI, USA.
[4]Department of Anatomy, Histology, and Embryology, School of Medicine, University of Split, Split, Croatia.
[5]Department of Genetics, Albert Einstein College of Medicine, Bronx, NY, USA.
[6]Department of Genetics and Development, Columbia University Medical Center, New York, NY.
[7]Department of Pediatric Nephrology, VU University Medical Center, Amsterdam, The Netherlands.
[8]Department of Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA.
[9]Division of Nephrology, Dialysis, Transplantation, and Laboratory on Pathophysiology of Uremia, Istituto G. Gaslini, Genoa, Italy.
[10]Department of Clinical and Experimental Medicine, University of Parma, Parma, Italy.
[11]University of Michigan School of Medicine, Department of Pediatrics-Nephrology, Ann Arbor, MI, USA.
[12]Department of Systems Biology, Columbia University, New York and Howard Hughes Medical Institute, USA.
[13]Departments of Systems Biology; Biochemistry and Molecular Biophysics; Division of Nephrology in Medicine and Zuckerman Mind Brain Behavior Institute, Columbia University, New York and Howard Hughes Medical Institute, USA.
[14]Department of General and Transplant Surgery, University Hospital of Heidelberg, Germany.
[15]Department of Pediatric Nephrology, Centre de référence des Maladies Rénales Héréditaires de l'Enfant et de l'Adulte (MARHEA), Necker-Enfants Malades hospital, Paris, France.
[16]INSERM UMR1163, Laboratory of Hereditary Kidney Diseases, Necker-Enfants Malades hospital, Paris Descartes - Sorbonne Paris Cite University, Imagine Institute, Paris, France.
[17]Sorbonne Universités, UPMC Univ Paris 06, Plateforme Post-génomique de la Pitié- Salpêtrière (P3S), UMS 2 Omique, Inserm US029, Paris France.
[18]Department of Genetics, Centre Hospitalier Universitaire de Reims, UFR Medecine, Reims, France.
[19]Section of Nephrology, Department of Emergency and Organ Transplantation, University of Bari, Bari, Italy.
[20]Dept. of Medical Sciences, University of Milano, and Institute of Biomedical Technologies, Italian National Institute of Research ITB-CNR, Milan, Italy.
[21]Dipartimento Ostetrico-Ginecologico e Seconda Divisione di Nefrologia ASST Spedali Civili e Presidio di Montichiari, Brescia, Italy
[22]Cattedra di Nefrologia, Università di Brescia, Seconda Divisione di Nefrologia Azienda Ospedaliera Spedali Civili di Brescia Presidio di Montichiari, Brescia, Italy.
[23]Department of Pediatrics, University Hospital of Split, Croatia.

[24]Department of Pathology, University Hospital of Split, Croatia.
[25]Department of Neurology, University of Washington School of Medicine, and Northwest VA PADRECC (Parkinson Disease, Research, Clinical and Education Center), Seattle, WA, USA.
[26]Division of Human Genetics, Department of Pediatrics, 22q and You Center, The Children's Hospital of Philadelphia and Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA.
[27]Dialysis Unit, Jagiellonian University Medical College, Krakow, Poland.
[28]Department of Pediatric Nephrology, Jagiellonian University Medical College, Krakow, Poland
[29]Department of Pediatrics, Immunology and Nephrology, Polish Mother's Memorial Hospital Research Institute, Lodz, Poland.
[30]Department of Pediatric Nephrology Medical University of Lublin, Lublin, Poland
[31]Chair and Department of Pediatrics, School of Medicine with the Division of Dentistry in Zabrze, Medical University of Silesia in Katowice, Poland
[32]Department of Pediatrics and Nephrology, Medical University of Warsaw, Poland.
[33]Krysiewicza Children's Hospital, Poznań, Poland.
[34]Department of Clinical Genetics, Our Lady's Children's Hospital Crumlin, Dublin, 12, Ireland; National Children's Research Centre, Our Lady's Children's Hospital Crumlin, Dublin, 12, Ireland.
[35]National Children's Research Centre, Our Lady's Children's Hospital Crumlin, Dublin, 12, Ireland; National Children's Hospital Tallaght, Dublin, 24, Ireland.
[36]Department of Clinical Genetics, Our Lady's Children's Hospital Crumlin, Dublin, 12, Ireland; University College Dublin UCD School of Medicine, Our Lady's Children's Hospital Crumlin, Dublin, 12, Ireland.
[37]Pediatric Surgery Unit, University Hospital of Parma, Parma, Italy.
[38]Departments of Pediatrics and Epidemiology, Perelmen School of Medicine at the University of Pennsylvania, Division of Nephrology, Children's Hospital of Philadelphia (CHOP), Philadelphia, Pennsylvania, USA.
[39]Division of Pediatric Nephrology, Children's Mercy Hospital, Kansas City, Missouri, USA.
[40]University Children's Hospital, Medical Faculty of Skopje, Skopje, Macedonia.
[41]Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA.
[42]Faculty of Medicine, Palacky University, Olomouc 77200, Czech Republic.
[43]Department of Medical Genetics, Poznan University of Medical Sciences, and NZOZ Center for Medical Genetics GENESIS, Poznan, Poland.
[44]Division of Pediatric Nephrology, University of New Mexico Children's Hospital, Albuquerque, New Mexico.
[45]Nephrology Division, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Charlestown, MA.
[46]Department of Pathology, Columbia University, New York, NY.
[47]Ben May Department for Cancer Research, University of Chicago, Chicago, IL.
[48]Department of Genetics, Howard Hughes Medical Institute, and Yale Center for Mendelian Genomics, Yale University, New Haven, CT.

[$]To whom correspondence should be addressed
*E.L.-R., Y.P.L., M.V., B.R.A, and V.P.C. contributed equally to this manuscript

ADDRESS CORRESPONDENCE TO:
Simone Sanna-Cherchi, MD, Division of Nephrology, Columbia University, College of Physicians and Surgeons, New York, NY 10032, USA; Phone: 212-851-5554; Fax: 212-851-5461; Email: ss2517@cumc.columbia.edu.
Or
Nicholas Katsanis, PhD, Center for Disease Modeling, 466A Nanaline Duke Building, Box 3709, Duke University Medical Center, Durham, North Carolina 27710, USA. Phone: 919.613.4694; Fax: 919.684.1627; E-mail: katsanis@cellbio.duke.edu.

SUPPLEMENTARY APPENDIX: TABLE OF CONTENTS

SUPPLEMENTARY PATIENTS AND METHODS

CAKUT patients and controls. Columbia University cohort: patient recruitment was performed at different research units in Italy (Genoa, Brescia, Parma and Foggia), Macedonia (Skopje), Poland (Poznan, the Polish Registry of Congenital Malformations), the Netherlands (Amsterdam), Croatia (Split), Czech Republic, and New York (Columbia University) after informed consent. The inclusion criteria for enrollment included presence of congenital anomalies of the kidney and urinary tract documented by pre- or post-natal imaging studies such as renal agenesis, renal hypoplasia/dysplasia, duplex kidney, ureteropelvic junction (UPJ) obstruction, duplicated ureter, vesicoureteral reflux. Available family members were screened to identify familial forms of disease and all patients were investigated for extra-renal manifestations. Additional cohorts included: French Cohort (Dr. Jeanpierre, et al.), Dublin Cohort (Dr. Barton, et al.), CKiD Cohort (Dr. Wong, et al.), Boston Cohort (Dr. Hildebrandt), and Children Hospital of Philadelphia (CHOP) Cohort (Dr. Hakonarson).

CNV analysis. Genome wide genotyping for copy number variation (CNV) analysis was conducted using high-density Illumina (n=1,820) or Affymetrix (n=260) single nucleotide polymorphisms (SNP) microarrays for 2,080 (1,752 in the discovery, 328 in the replication) CAKUT patients and 22,094 population controls as previously described[1,2 3]. Briefly, Raw intensity data were processed in GenomeStudio v2011 (Illumina). PennCNV[4] was used to determine copy number variant (CNV) calls. CNVs were mapped to the human reference genome hg19 and annotated with UCSC RefGene and RefExon using the CNVision program[5]. Only CNVs with confidence scores ≥ 30 were considered in the analyses based on experimental validation from our prior study[1]. Similar to our prior studies[1 ,2 3], two CNVs were considered to be identical if they had the same copy number value and had ≥70% reciprocal overlap; otherwise they were considered to be distinct.

Exome sequencing. Exome sequencing and analysis was performed as previously described[6-8]. Briefly, for each capture experiment, 3 µg of genomic DNA was fragmented, linkers were ligated to the ends and a library was prepared. Genomic DNA was annealed to capture probes, and bound genomic DNA was eluted and subjected to sequencing. Next-Gen sequencing was then performed on an Illumina HiSeq 2500 machine. Sequence reads were converted to FASTQ format and mapped to the reference genome. Reads that aligned to the targeted exome were extracted and statistics on coverage were collected using a Perl script. Positions found to harbor heterozygous or homozygous variants that deviate from the reference sequence were identified and rare or novel SNPs were identified by comparison to the reference genome, 1000 Genomes data and dbSNP. Low-probability SNVs were identified by empiric methods that we have found significantly reduce false-positive calls: low-quality bases (quality scores <45), heterozygous calls based on low read coverage (<8X), variants that appear exclusively or with high frequency at the same read position on the same strand (implying a preponderance of non-independent reads), and low quality genotype calls using samtools (<40). Quality score >30 for SNVs, >60 for indels, and coverage >8X identified high-quality variants. Allelic frequencies were compared to dbSNP, 1000 genome, the NHLBI Exome Variant Server (http://evs.gs.washington.edu/EVS/), and the Exome Aggregation Consortium (http://exac.broadinstitute.org/).

High-throughput next generation sequencing. High-throughput next generation sequencing for the 9 genes included in the 370 Kb DiGeorge MRO was conducted using the Fluidigm microfluidic PCR capture (www.fluidigm.com) coupled to next-generation sequencing on an Illumina HiSeq 2500 (Illumina). The targeted enrichment and sequencing protocols have been previously described[9,10]. Briefly, In order to cover all 107 coding exons and intron/exon boundaries of the DGS MRO candidate genes, we designed target specific primer pairs according to the guidelines of the Access Array™ user guide (Fluidigm). The maximal amplicon size is set to 290 bp anticipating subsequent NGS paired-end reads of 2x150 bases. Universal primer sequences required for Illumina-compatible amplicon tagging and for downstream indexing were added at the 5' end to all target specific primers. We then generated 48x10-plex primer pools. 48 DNA-containing samples were combined with each of the 48 primer solutions into 2,304 separate microreaction chambers. PCR products were then harvested and transferred to a 96-well microtiter-plate. In a second PCR reaction, Illumina sequence specific adaptors and sample barcodes were attached. We indexed up to 96 total DNA samples by attaching 96 different barcodes after processing 2 different 48.48 Access Arrays. Subsequently, all 96 barcoded samples were pooled and submitted for next-generation sequencing on a single lane of an Illumina HiSeq 2500 instrument. In order to sequence the Fluidigm specific barcodes and perform a multiplexed paired-end run, we substituted the Illumina index sequencing primer with custom Fluidigm specific index primer, and extended the index read length in order to decipher the whole Fluidigm barcode.

Validation and annotation of variants. We performed Sanger sequencing to validate all variants and to test segregation in available family members. We determined allele frequencies using the Exome Aggregation Consortium (ExAC) database (http://exac.broadinstitute.org/), computed scores for potential pathogenicity using Polymorphism Phenotyping, version 2 (Polyphen-2)[11], Combined Annotation Dependent Depletion (CADD)[12], and PhastCons score by alignment with 30 vertebrate species[13].

Zebrafish lines and handling. The NaK ATPase alpha1A4:GFP transgenic[14] line was raised and maintained according to standard husbandry procedures. Embryos were kept at 28.5°C in embryo medium.

Zebrafish morpholino knockdown. Morpholinos against candidate genes were designed by and ordered from Gene Tools, LLC (Philomath, OR). The sequences for the morpholino were: aifm3_SB (splicing blocker): 5'-TGTACAGATCTTACTGTCCATTCTC-3'; crkl_TB (translation blocker): 5'-AGGAGTCGAACCGTGCAGACGACAT-3'; ret_SB: 5'-AGTATAAACCCTTACCTTCAGACAG-3'; snap29_SB: 5'-TAACCAGTGTTAAATCTGACCTGCT-3'; lztr1_SB: 5' - GTAACATATAAAACTCACCGACACC - 3'; pi4ka_SB; 5'-AACAGGTAGATGATTCATACCTCGT-3'; serpind1_SB: 5'-TCTAGAGTGTGGTGCTTTACCTGTT-3'; slc7a4_SB: 5'-CCTGTGATATTTTAAACTCACCTGA-3'. A mixture of morpholino and/or human mRNA was microinjected into zebrafish embryos at 1-4 cell stage, then the larvae were harvested 4.5 dpf for immunostaining.

Genome-editing of ret, snap29 and crkl loci using the CRISPR/Cas9 system. Guide-RNA (gRNA) was produced by synthesizing and annealing two oligonucleotides, ret gRNA F: TAGGGTGGAGTCGTCGACCGCAAA, R: AAACTTTGCGGTCGACGACTCCAC; snap29 gRNA F: TAGGGGACTCAGCGAACTCTCCTC, R: AAACGAGGAGAGTTCGCTGAGTCC; crkl gRNA F: TAGGGGAATGGGAGATGGAGTCTG, R: AAACCAGACTCCATCTCCCATTCC. The annealed oligos were then ligated to a T7cas9sgRNA2 vector by performing the ligation and digestion in a single step in a thermal cycler as previously described[15]. Prior to transcription, the gRNA vector was linearized with BamHI restriction enzyme. The gRNA was transcribed using the MEGAshortscript T7 kit (Life Technologies, AM1354) and purified using alcohol precipitation. A total of 100 pg of gRNA and 200 pg of Cas9 purified protein (PNA Bio) was co-injected into individual cells of one-cell stage embryos. For T7 endonuclease I assay, genomic DNA was prepared from 1 dpf embryos as previously described[16]. A short stretch of the genomic region flanking each gRNA target site was PCR amplified from the genomic DNA (ret, Fwd: TTCTTCAATTGTGCCATGTAGC, Rev: TTTGATTTGCACGTCCTTGTAG; snap29, Fwd: GGCTTCAGCACAGGATGTCT, Rev: GTCTCACCTCTGCGGTTTCA; crkl, Fwd: GCATGATTCCTGTTCCCTATGT, Rev: TTAGCCACTTACCTCCAGTGCT). The PCR amplicon was then denatured slowly and reannealed to facilitate heteroduplex formation. The reannealed amplicon was then digested with 5 units of T7 endonuclease I (New England Biolabs) at 37°C for 45 minutes. The samples were resolved by electrophoresis through a 3.0% agarose gel and visualized by ethidium bromide staining.

Pronephros convolution studies and measurements. All measurements of proximal tubule convolution were performed in NaK ATPase alpha1A4:GFP transgenic fish. To visualize the entire proximal pronephros, embryos were fixed and stained with an antibody against NaK ATPase (a6F, DSHB. Embryos of 4.5 dpf were fixed in Dent's fixative (80% methanol, 20% dimethylsulfoxide) overnight at 4 °C. After rehydration with decreasing series of methanol in PBS, embryos were washed with PBST and bleached for 15-20 minutes in PBST+3%H_2O_2+5%KOH. Embryos were then washed twice with IF buffer (0.1% Tween-20, 1% BSA in 1x PBST) for 10 min at room temperature. After incubation in blocking solution (10% FBS, 1% BSA in 1x PBST) for 1 hour at room temperature, embryos were incubated at 4°C overnight in primary antibody (a6F serum, 1:20 dilution). After two washes in IF buffer for 10 min, embryos were incubated with 488 Alexa Fluor goat anti-mouse IgG (1:500; A11001, Life Technologies) for 2 hours at room temperature. Image acquisition and analysis was performed using Nikon NIS-Elements software. The images were imported into NIH ImageJ software and curved line lengths were measured using polygon approximation. Measurements were normalized to control embryos and body length (relative length = a/b)**.** Significant difference from controls was determined using the Holm-Šídák multiple comparisons test following ANOVA.

Tissue studies. We used human embryonic and fetal (6^{th} and 21^{th} developmental week, respectively) and postnatal kidney (1.5 year old) to investigate expression of CRKL, SNAP29, and AIFM3. Normal human conceptuses, without signs of anomaly or maceration, were obtained after tubal pregnancy or spontaneous abortion as well as normal postnatal kidney from the Department of Pathology, University Hospital of Split,

Croatia. Embryonic and fetal tissues were treated as postmortem material with permission of the Ethics and Drug Committee of the Clinical Hospital of Split in accordance with the Helsinki Declaration[17]. Snap-frozen fresh mouse embryo tissue was embedded in OCT Compound in cryomolds for Crkl staining. Frozen blocks were cut at 25 µm thickness. Sections were mounted on Superfrost plus slides and stored at -80 ˚C. Cryosections were washed in PBS-Tween, permeabilized in PBS 0.1% Triton-X 100, and blocked in blocking buffer (PBS, 0.05% Triton X-100 and 3% BSA). Antibodies were diluted in blocking buffer. Conjugated secondary antibodies (Jackson ImmunoResearch) were diluted in blocking buffer. Images were taken on a Nikon N-STROM/TIRF/SD Green Spin Disk Confocal Miscroscope (Herbert Irving Comprehensive Cancer Center) and on an Olympus DP80 fluorescent microscope. For both human and mouse tissue staining, a rabbit polyclonal primary antibody was used for Crkl (1:100, SC-319, Santa Cruz). The following antibodies were used for human tissues studies only: SNAP-29 (1:500, OSS00034W, Thermo Fisher Scientific); AIFM3 (1:300, TA313242, Origene). As negative control, the irrelevant immunoglobulin G (MA5-16384, Thermo Fisher Scientific, Waltham, MA) was used.

The entire pronephros was visualized following embryo fixation and staining with an antibody against NaK ATPase (a6F, DSHB). To visualize deposition of crkl in whole-mount 96hpf larvae, an anti-crkl antibody (Abcam catalog # ab129318) was used at a dilution of 1:25. Embryos of 4.5 dpf were fixed in Dent's fixative (80% methanol, 20% dimethylsulfoxide) overnight at 4°C. After rehydration with a decreasing series of methanol in PBS, embryos were washed with PBST (0.5% Triton X-100) and bleached for 15-20 minutes in PBST+3%H2O2+5%KOH. Embryos were then washed 3x with PBST for 10 min at room temperature and incubated in blocking solution (10% NDS in 1x PBST) for either 2 hours at room temperature or overnight at 4°C. Larvae were then incubated overnight at 4°C with primary antibodies (a6F serum, 1:10 dilution; Crkl, 1:25 dilution). After three washes in PBST for 10 min/wash, embryos were incubated in secondary antibodies (Alexa Fluor® 350 donkey anti-mouse IgG (1:100, Life Technologies catalog # A10035); Alexa Fluor® 594 donkey anti-goat (1:250, Jackson ImmunoResearch Laboratories catalog # 705-585-003)) for 2 hours at room temperature. Samples were whole-mounted in 50% glycerol-PBS and imaged in Olympus cellSens software using an Olympus IX73 inverted microscope equipped with an Olympus DP80 digital camera. The images were then exported as *.tiff files and contrast-adjusted for maximum visibility using either Volocity (v6.3, Perkin Elmer) or Adobe Photoshop CC (Adobe Systems Inc.). Figures were constructed using Adobe Illustrator CC (Adobe Systems Inc.).

Flow cytometry and cell sorting. Cells were analyzed on an BD LSR II flow cytometer before sorting using BD Influx cell sorter, based on the same laser configuration. Briefly, cells were re-suspended in a single cell suspension in 1x PBS, 2% FBS, and incubated with DAPI (300 nM) for 10 minutes at room temperature. The single-cell suspension was washed in 5 ml of 1x PBS, 2% FBS. Before the addition of the NaK ATPase primary antibody (a6F, DSHB), cells were incubated at room temperature for 30 minutes before washing, then the secondary antibody was added and incubated for 30 minutes at room temperature (APC anti-mouse IgG2a (clone RMG2a-62)). Cells were washed and re-suspended in 500 µl of 1x PBS, 2% FBS, before being sorted into three separate populations, negative, GFP positive, and NaK ATPase positive. Cells were collected in lysis buffer containing RNAse inhibitor and immediately frozen on dry ice prior to RNA isolation using standard Trizol method.

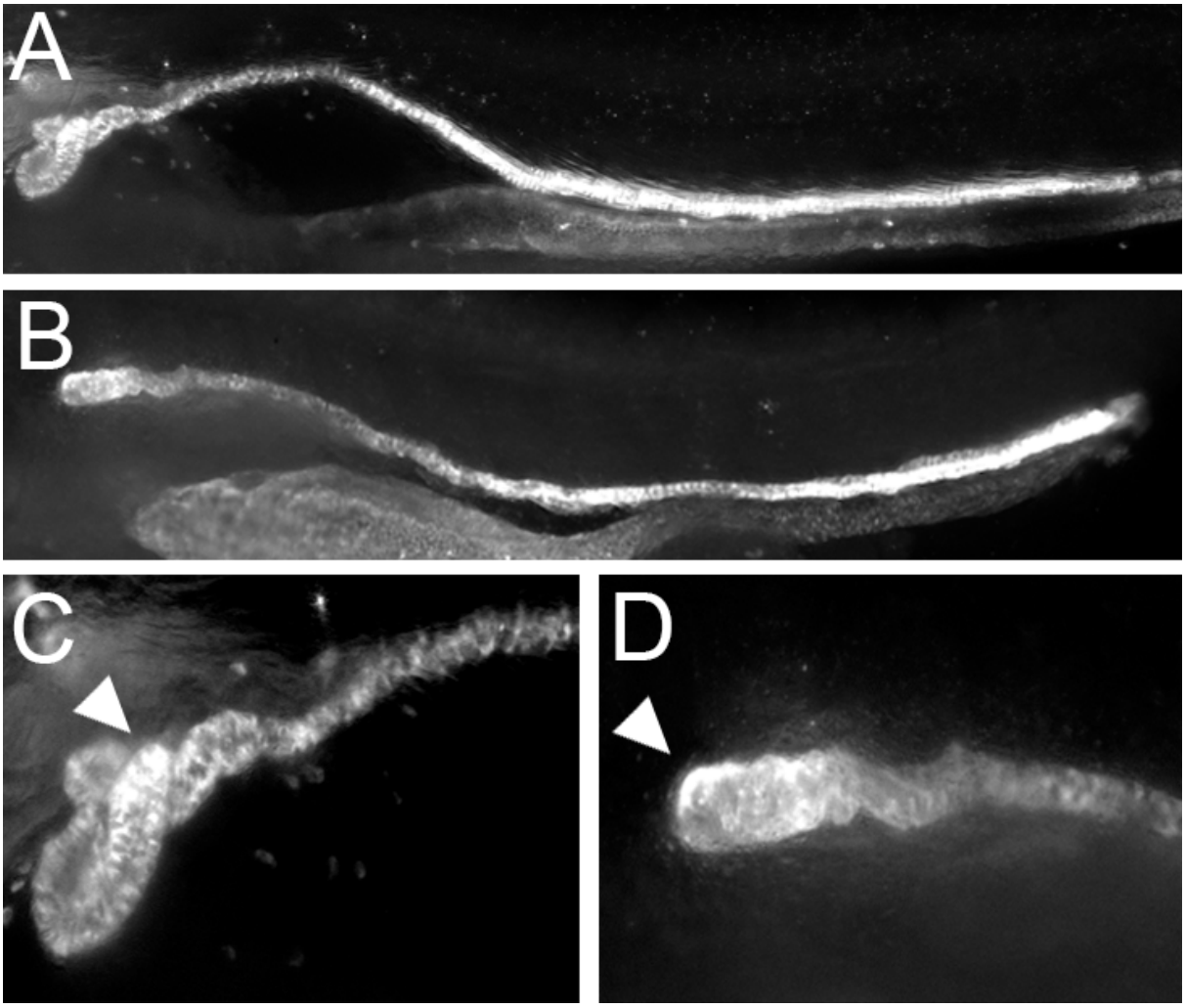
Quantitative PCR primers. *gapdh*: forward 5'- GTGGAGTCTACTGGTGTCTTC -3'; reverse 5'- GTGCAGGAGGCATTGCTTACA -3'. crkl: forward 5'- TGCACGGTTCGACTCCTC -3'; reverse 5'- CGCTGCCAGTGAAGTCATAA -3'.

In situ hybridization probes. The zebrafish in situ hybridization for pax2a was performed using the same probe and protocol as described by Drummond et al[18]. Probe sequence for crkl: 5'-
GGACAGAGACACGGGATGTTTTTGGTGCGAGATTCGTCCACTTGTCCTGGTGATTATGTACTTTCAGTGTC
CGAAAACTCCAAAGTTTCGCACTATATCATCAACTCTTTGCCAAGCAAGAGATTCAAGATAGGCGATCAAG
AGTTTGATAATCTACCCGGCCTTTTGGAGTTCTATAAAATACATTATTTGGATACGACCACCCTCATAGAAC
CAGCACCAAGGTACCCCAGTACTGCTCTGCCCAGTGGTCCTATTCAGCCATCTGGAGGACTGGGTGATG
AGAACCAGGAGTATGTGCGGACTCTTTATGACTTCACTGGCAGCGATGCTGAAGACCTTCCTTTCAAGAA
GGGTGAAATCCTAATAATTATGGACAAGCCTGAGGAGCAGTGGTGGAGTGCCAAGAACAAAGAAGGCCG

AACAGGCATGATTCCTGTTCCCTATGTAGAAAAGCTTGTGAGATCTTCGCCTCATCCTGGTCAGTCCATCC
ATGGTTCACGGAATTCCAATAGCTATGGCATCCCTGAGCCATCACACGCCTATGCCCAGCCTCAGACTC-3’

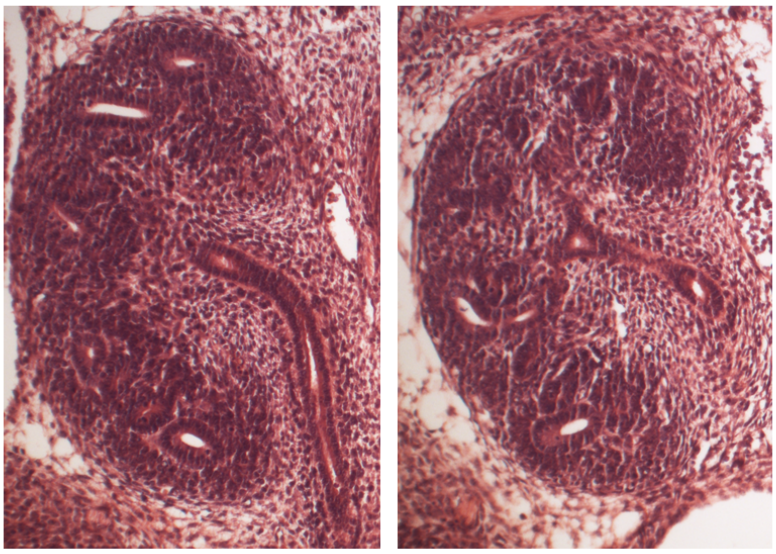
Crkl mouse models. To generate a *Crkl* Cre-dependent conditional mutant strain, we used three ES clones targeted at the *Crkl* locus obtained from EUCOMM/Helmholtz Zentrum München (Design ID 44027; E11, F09, and G10). The JM8.N4 ES cells (in a C57Bl/6N inbred background) were used for targeting. Chimeras were generated using a standard blastocyst injection followed by embryo transfer to pseudopregnant females. Germ line chimeras were crossed with C57BL/6J females to obtain F1 heterozygous animals. After one additional generation of backcross with C57BL/6J mice, heterozygous mice were then crossed with a C57BL/6J congenic line of the FLP strain B6;129S4-Gt(ROSA)26Sortm2Dym/J (FLPeR mice) in order to convert the original targeted allele to a Cre-dependent conditional mutation without lacZ reporter or neo cassettes. Pups that were double heterozygous for FLPeR and the targeted mutation were then backcrossed with wild type C57BL/6J mice to segregate out the FLPeR. The *Crkl* allele in the resultant Cre-dependent conditional mice has one FRT and one loxP sites approximately 500 bases upstream of Exon 2 and one loxP site approximately 200 bases downstream of the exon. The *Crkl* conditional mutant line was maintained as homozygotes by brother-sister intercrosses. Homozygous *Crkl* conditional mutant mice ($Crkl^{f/f}$) do not have an overt phenotype and reproduce normally. Mice harboring the floxed *Crkl* allele (Crklf/+; Crklf/f) were bred with transgenic mice harboring Cre recombinase under the control of three different promoters: Six2-Cre, HoxB7-Cre, and E2a-Cre to drive Cre-madiated deletion of *Crkl* exon 2 in capped mesenchyme, ureteric bud, and in all lineages (global knockout), respectively. Basic morphologic characterization of kidney and urinary tract pathology was conducted on paraffin-embedded tissues, stained with hematoxylin and eosin, in embryos at embryonic development day E13.5, E14.5, E15.5, and E16.5. All animal work for generation of the conditional mutant line was performed in strict concordance with an animal protocol approved by the IACUC of the University of Chicago and Columbia University.

Supplementary Figure S1. Close-up of proximal tubule convolution in NaK ATPase:GTP stained embryos.



(A-B) The proximal tubule of 4.5 dpf zebrafish larvae visualized by Na^+/K^+ ATPase staining in (A) uninjected control embryos and (B) morpholino (MO) knockdown of *ret* (8.0 ng/nl). (C-D) Increased magnification of the most proximal portion of the pronephros adjacent to the glomerulus (arrowheads) highlight the distinct convolution defects seen in morphant embryos.

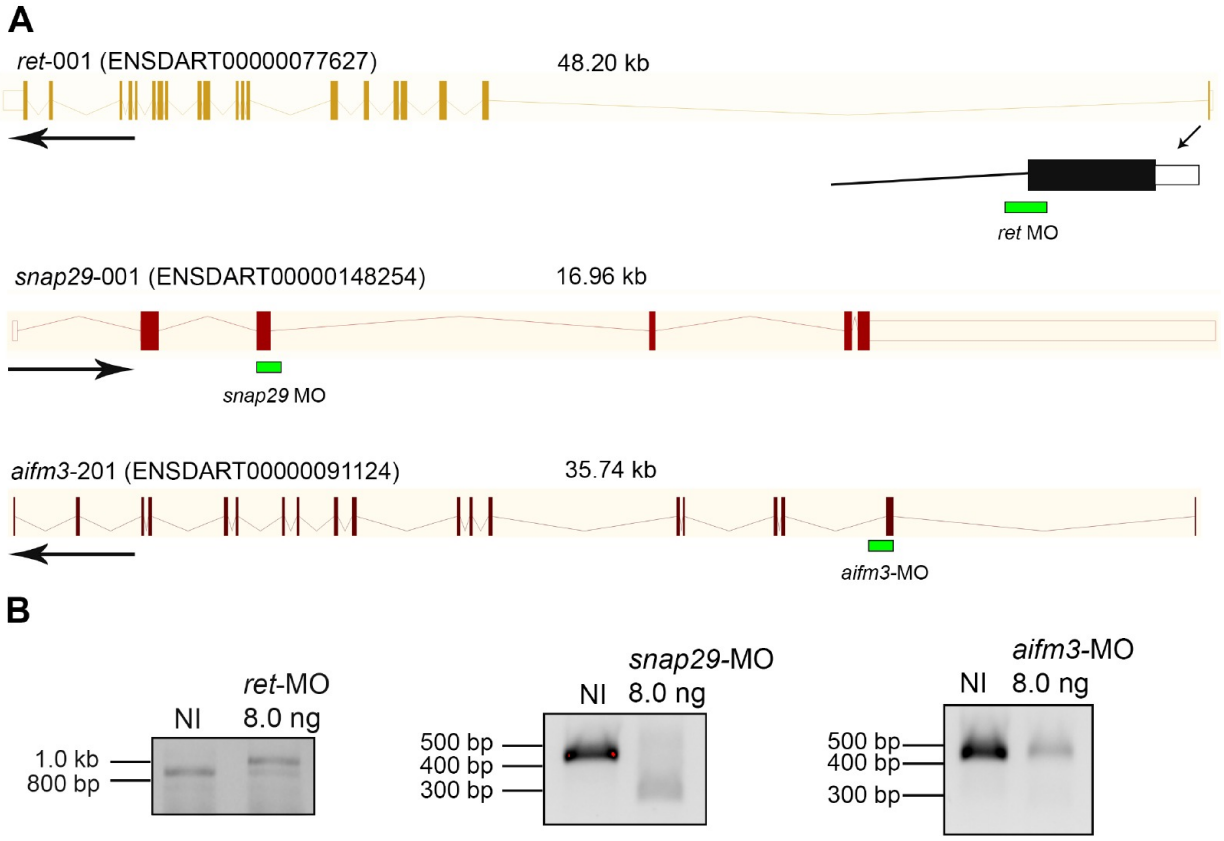
Supplementary Figure S2. Normal renal development in *Tbx1* null mice.



Tbx1 +/- Tbx1 -/-

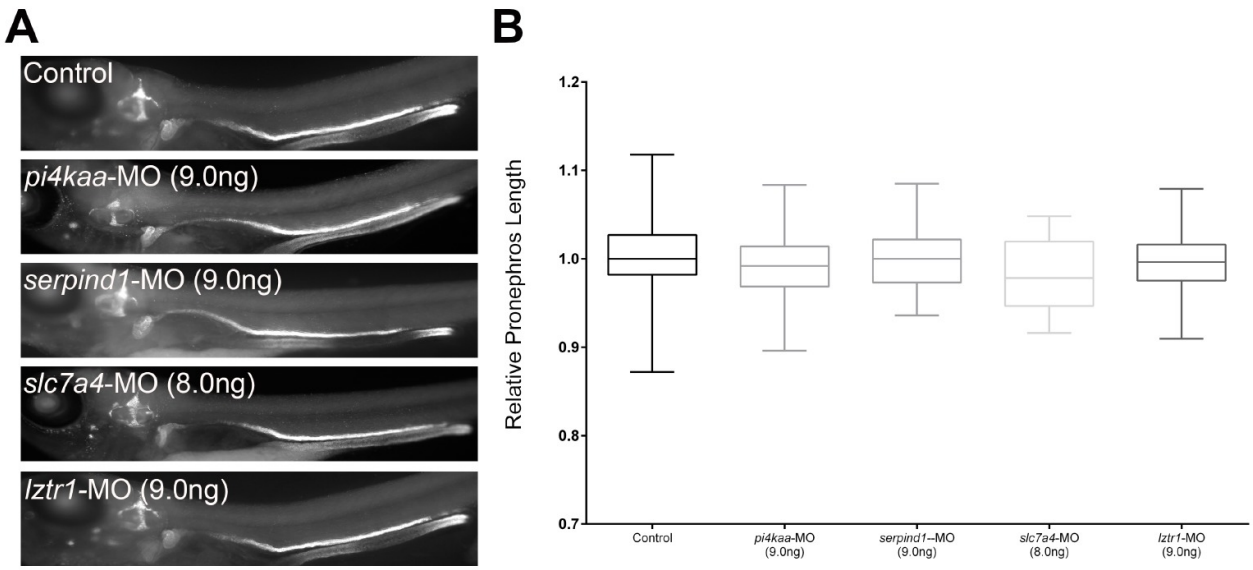
Mice heterozygous for *Tbx1* null mutation are viable, fertile and do not show any anatomical abnormalities, while homozygous mice recapitulate most of the phenotypes of DiGeorge syndrome[19]. In this figure is shown a representative image of normal kidney development in both heterozygous and homozygous Tbx1 mutant mouse embryos at E12.5.

Supplementary Figure S3. Efficiency of ret, snap29, and aifm3 morpholinos.



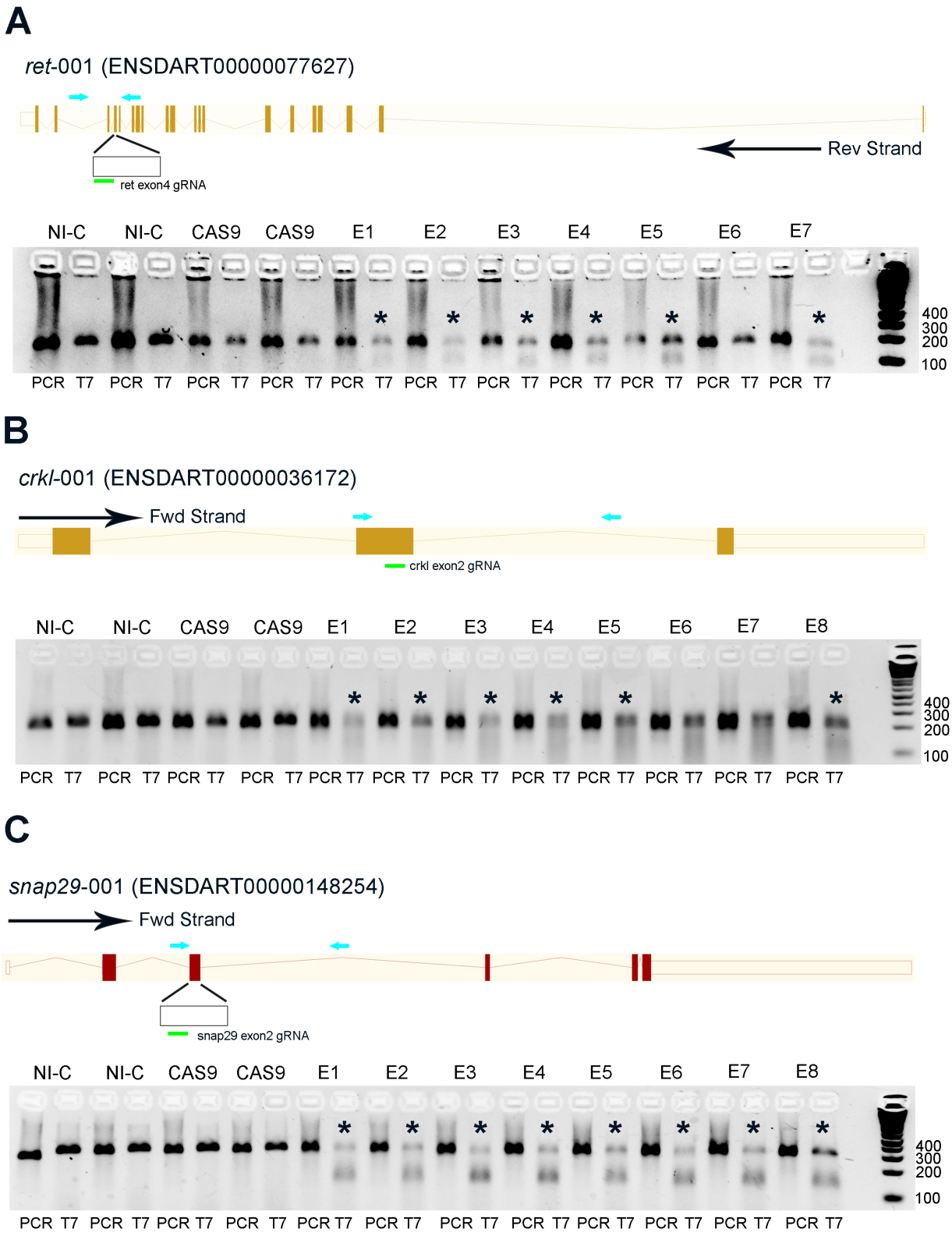
(A) Schematic of *ret*, *snap29* and *aifm3* zebrafish transcripts and the location of each respective splice-blocking morpholino (MO) target (green bar). (B) cDNA PCR amplification of the region surrounding the exon-intron junction of each MO target yield aberrant splicing in MO-injected embryos compared to non-injected controls (NI). Amplification of a 791 base pair (bp) *ret* fragment spanning exons 1-4 reveals partial retention of intron 2 in *ret*-MO injected embryos. *snap29* morphants display a near complete loss of wild-type transcript (92%) with the resulting amplification (508 bp) favoring a deletion of exon 3 (308 bp). *aifm3*-MO injected embryos reduces normal splicing of wild-type message by 80% and favors aberrant splicing resulting in a transcript where exon 2 (214 bp) is deleted.

Supplementary Figure S4. Functional modeling of *pi4ka, serpind1, slc7a4* and *lztr1* does not display convolution defects.



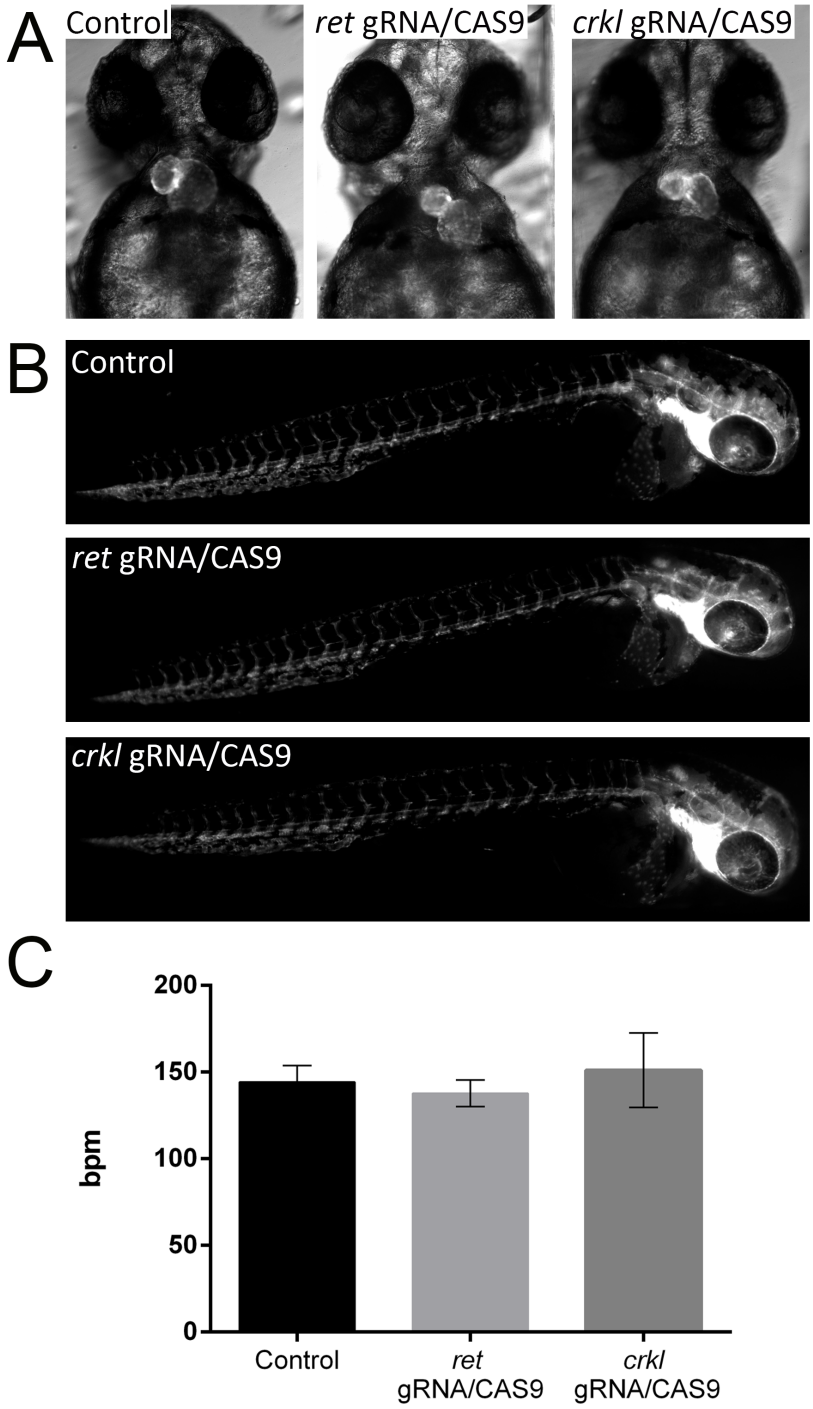
(A) The proximal tubule of 4.5 dpf zebrafish larvae is visualized by Na^+/K^+ ATPase staining. Morpholino (MO) knockdown of *pi4kaa* (9.0 ng/nl), *serpind1* (9.0 ng/nl), *slc7a4* (8.0 ng/nl) and *lztr1* (9.0 ng/nl) do not display convolution defects different from non-injected control (Control) embryos. (B) Relative pronephros length was defined as the ratio of the length of the pronephros divided by the length of the body axis in individual larvae. Quantitative assessment among multiple batches of MO-injected embryos against *pi4kaa, serpind1, slc7a4* and *lztr1* do not display a significant difference in relative pronephros length from control embryos. Control, sham-injected control (n=177); *pi4kaa*-MO (n=27); *serpind1*-MO (n=34); *slc7a4*-MO (n=26); *lztr1*-MO (n=27).

Supplementary Figure S5. Establishment of CRISPR/Cas9 zebrafish models.



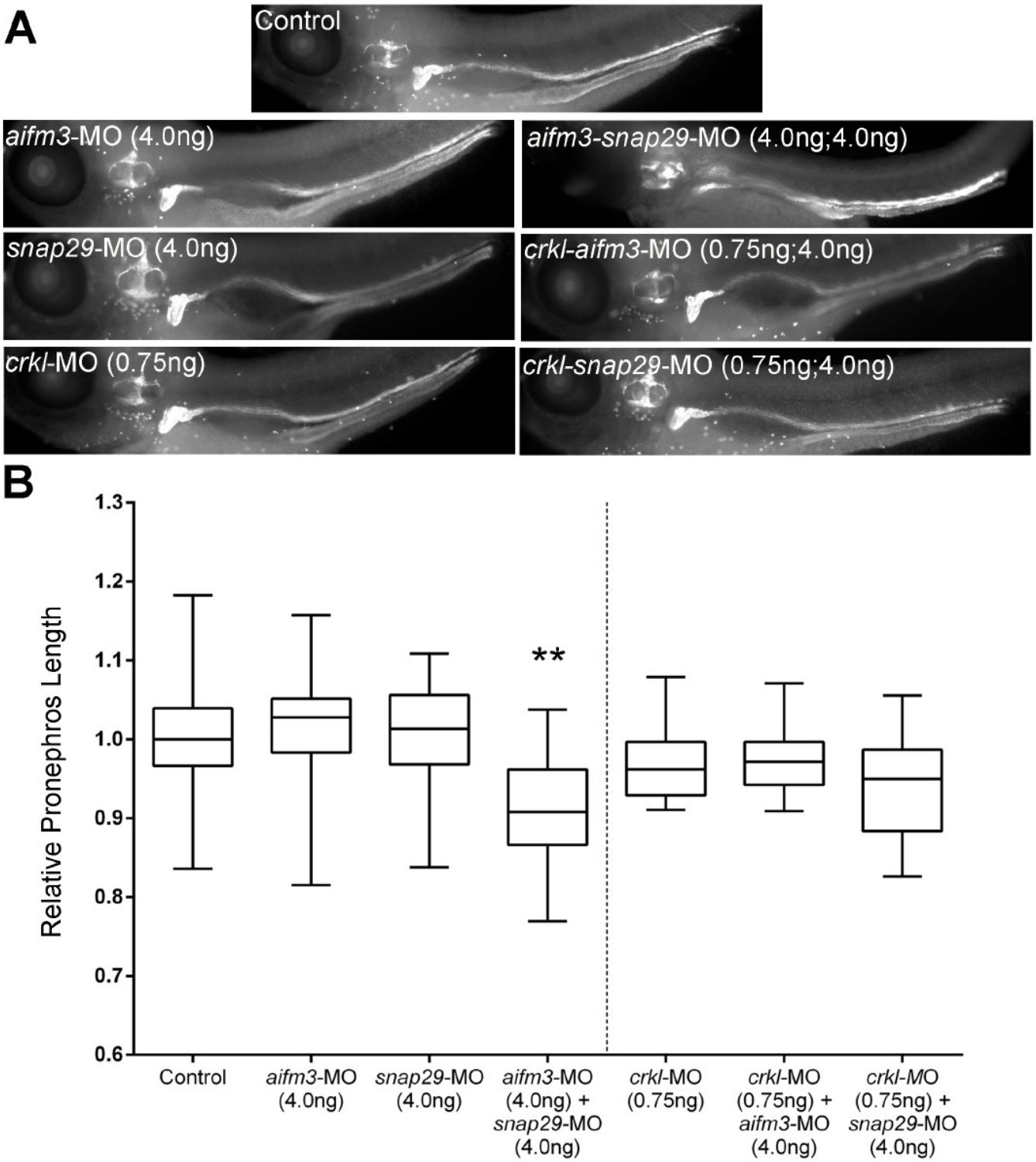
Schematics of zebrafish (A) *ret*, (B) *crkl*, and (C) *snap29* loci and location of the guide RNA (gRNA) target used for CRISPR experiments; the primers used to PCR-amplify the target region are shown (arrowheads). At 1 dpf, a representative sampling of 7-8 founders (E1-E8), non-injected controls (NI-C), or Cas9-only injected embryos (Cas9) were selected and subjected to T7 endonuclease 1 (T7E1) assay. The appearance of multiple T7E1 fragments indicated positive gRNA targeting of each respective locus (*asterisks). No T7E1 fragments were detected in either non-injected or Cas9 injected embryos.

Supplementary Figure S6. Assessment of heart and vasculature defects in *ret* and *crkl* CRISPR/CAS9 models.



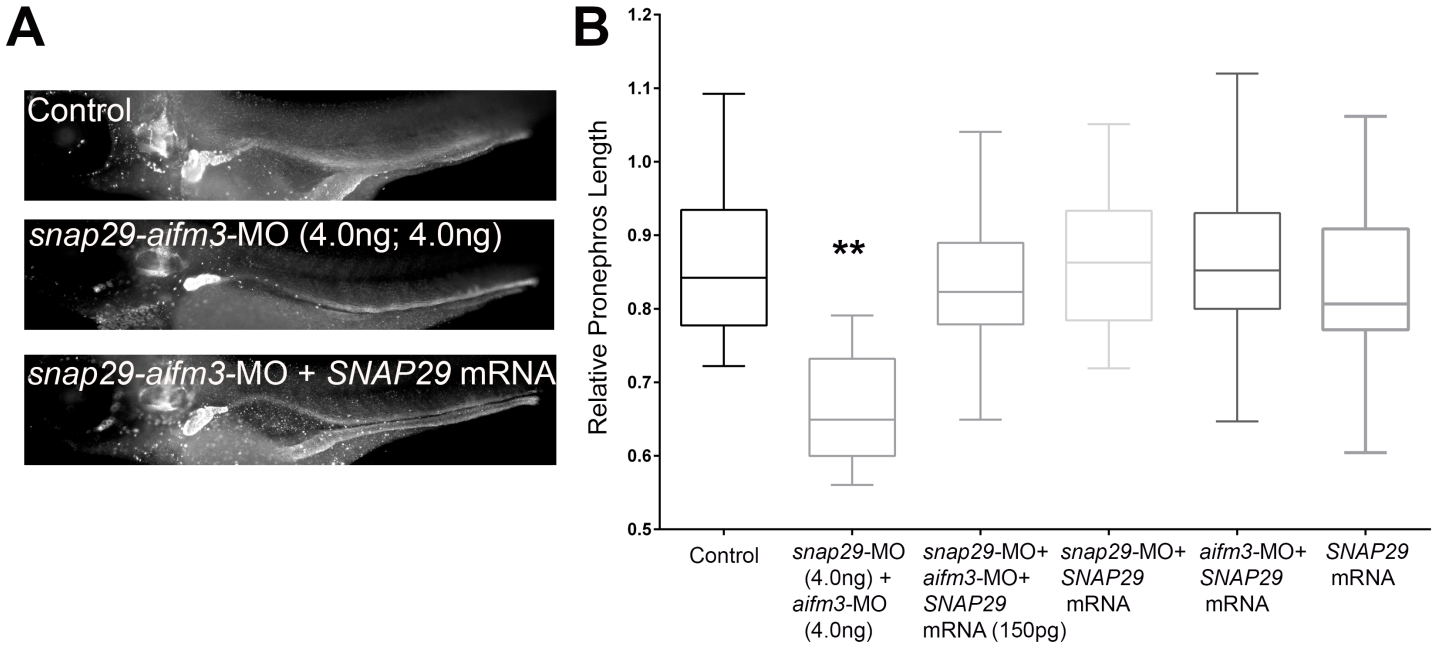
(A) 48 hpf embryos were assessed for heart-looping defects using a *cmlc2*-GFP transgenic line, which allows for specific visualization of myocardial cells around the heart chambers and atrioventricular canal.[20] Embryos were injected with either sham (Control), *ret* gRNA/CAS9 or *crkl* gRNA/CAS9. No significant defects were observed in either *ret* or *crkl* CRISPR/CAS9 F0 models. (B) Fli:eGFP transgenic embryos[21] were evaluated at 48 hpf for gross vasculature defects in sham-injected Control, *ret* gRNA/CAS9 or *crkl* gRNA/CAS9 F0 models. (C) Heart beat rate was measured at 48 hpf in sham-injected Control, *ret* gRNA/CAS9 or *crkl* gRNA/CAS9 injected embryos. No significant difference between conditions was observed. N=22-26, repeated two times.

Supplementary Figure S7. *aifm3* and *snap29* interact in a genetic context.



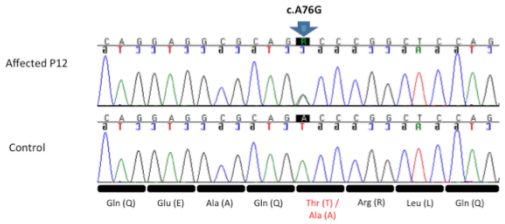
(A) Morpholino knockdown at sub-effective doses of *aifm3*-MO (4.0 ng/nl), *snap29*-MO (4.0 ng/nl) and *crkl* (0.75 ng/nl) display proper convolution of the anterior pronephros in 4.5 dpf embryos. Morpholinos were co-injected at sub-effective doses to elucidate convolution defects only apparent under co-suppression. *crkl* co-suppressed with either *aifm3* or *snap29* did not produce a noticeable convolution defect compared to sham-injected controls or single MOs. In contrast, *aifm3* co-suppressed with *snap29*, induced a significant defect in proximal tubule convolution. (B) Co-injection of *aifm3*-MO and *snap29*-MO produce a significant convolution defect when compared to either morphant alone. Control, sham-injected control (n=93); *aifm3*-MO (n=89); *snap29*-MO (n=90); *aifm3*-MO+*snap29*-MO (n=59); *crkl*-MO (n=41); crkl-MO+*aifm3*-MO (n=45); *crkl*-MO+*snap29*-MO (n=52). **p<0.01.

Supplementary Figure S8. *In vivo* interaction of *aifm3* and *snap29* is specific.

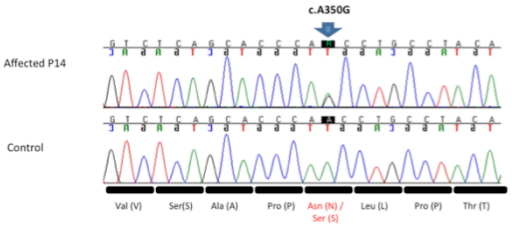


(A) Co-suppression of *aifm3* (4.0 ng/nl) and *snap29* (4.0 ng/nl) in 4.5 dpf zebrafish larvae by morpholino (MO) injection at sub-effective doses induces a significant defect in proximal tubule convolution as visualized by $Na^{+/}K^{+}$ ATPase staining. To test that this interaction is specific, and not due to toxicity caused by the presence of multiple MOs, *SNAP29* human mRNA (150 pg/nl) was injected in the presence of *aifm3-snap29* co-suppression. (B) Complementation of *aifm3-snap29* co-suppression by *SNAP29* human mRNA significantly rescues convolution defects when quantified relative to embryo body length. Control, sham-injected control (n=25), *snap29*-MO + *aifm3*-MO (n=24), *snap29-aifm3*-MO + *SNAP29* mRNA (n=36), *snap29*-MO + *SNAP29* mRNA (n=23), *aifm3*-MO + *SNAP29* mRNA (n=27), *SNAP29* mRNA (n=21). Repeated two times. **p<0.01.

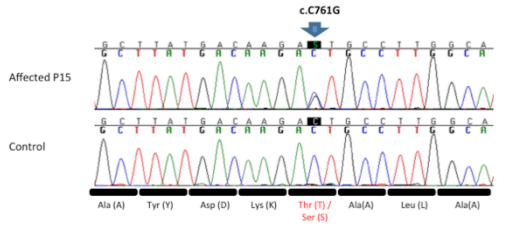
Supplementary Figure S9. Protein conservation for the four rare CRKL missense variants.



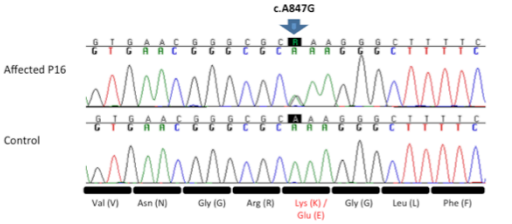
Species	Protein sequence (p.T26A)
H.sapiens	QEAQ T RLQG
P.troglodyte	QEAQ T RLQG
M.mulatta	QEAQ T RLQG
N.leucogenys	QEAQ T RLQG
E.caballus	QEAQ T RLQG
S.scrofa	QEAQ T RLQG
C.familiaris	QEAQ T RLQG
M.musculus	QEAQ T RLQG
R.norvegicus	QEAQ T RLQG
G.gallus	AEAQ T RLQG
X.laevis	QEAL S RLQG
D.rerio	QEAQ N RLQG



Species	Protein sequence (p.N117S)
H.sapiens	VSAP N LPTA
P.troglodyte	VSAP N LPTA
M.mulatta	VSAP N LPTA
N.leucogenys	VSAP N LPTA
E.caballus	VSAP S LPTA
S.scrofa	VSAP N LPTA
C.familiaris	VSAP N LPTA
M.musculus	VSAP N LPTA
R.norvegicus	VSAP N LSTA
G.gallus	GSAP A MSAA
X.laevis	GTAP S VPVT
D.rerio	PSGP I QPSG



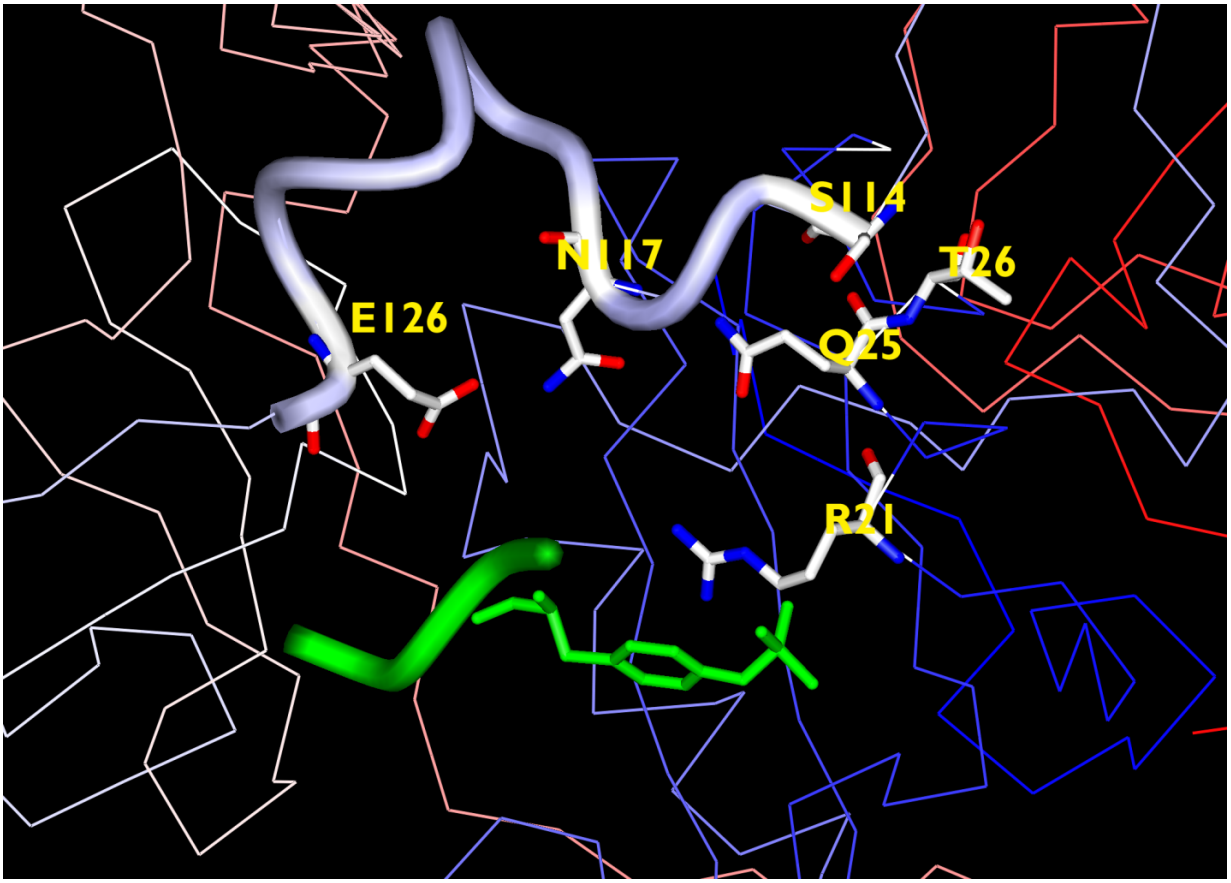
Species	Protein sequence (p.T254S)
H.sapiens	AYDK T ALAL
P.troglodyte	AYDK T ALAL
M.mulatta	AYDK T ALAL
N.leucogenys	AYDK T ALAL
E.caballus	AYDK T ALAL
S.scrofa	AYDK T ALAL
C.familiaris	AYDK T ALAL
M.musculus	AYDK T ALAL
R.norvegicus	AYDK T ALAL
G.gallus	AYDK T ALAL
X.laevis	AYDK T ALAL
D.rerio	AYDK T ALAL



Species	Protein sequence (p.K283E)
H.sapiens	VNGR K GLFP
P.troglodyte	VNGR K GLFP
M.mulatta	VNGR K GLFP
N.leucogenys	VNGR K GLFP
E.caballus	VNGR K GLFP
S.scrofa	VNGR K GLFP
C.familiaris	VNGR K GLFP
M.musculus	VNGR K GLFP
R.norvegicus	VNGR K GLFP
G.gallus	VNGR K GLFP
X.laevis	VNGR K GLFP
D.rerio	VNNR K GLFP

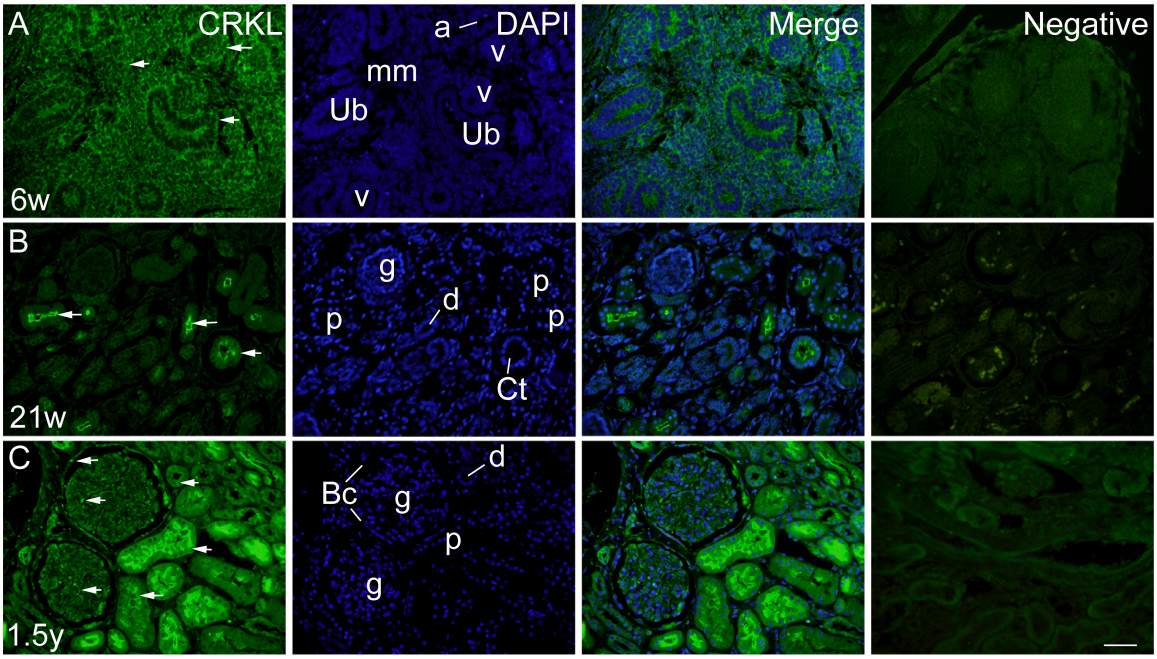
Chromatograms of the four novel missense variants identified in RHD patients, compared to unaffected controls. Protein alignment across twelve vertebrate orthologs is reported.

Supplementary Figure S10. Protein modeling for CRKL missense variants.



N-terminal SH2 and central SH3 domains. Both T26 and N117 are making tight hydrogen bonds. The sidechain oxygen of T26 has a H-bond with S114 backbone nitrogen (distance 3Å). N117 makes two H-bonds. Its sidechain oxygen has a H-bond with the sidechain oxygen of Q25 (distance 3.4Å) and its sidechain nitrogen has a H-bond with the sidechain oxygen of E126 (distance 2.7Å). The N117 H-bonds are buried. The H-bonds seem be “anchoring” the loop from residues 114 to 126 to the N-terminal SH2 domain. These residues are predicted to be critical in determining the relative orientation of the N-terminal SH2 domain and the central SH3 domain. Therefore, loss of the H-bonds on mutation would be expected to be highly disruptive. The loop is also right near the phosphopeptide binding site (R21 is the arginine in the SH2 binding pocket and phosphotyrosine from a cocrystal of another SH2 domain with its partner is shown in green). In summary, by disrupting the conformation of the 114-126 loop, these mutations are predicted to have a considerable effect on both affinity and specificity. The T254 and K283 mutations, although could probably be structurally accommodated, they are near the poly-proline binding site of the SH3 domain.

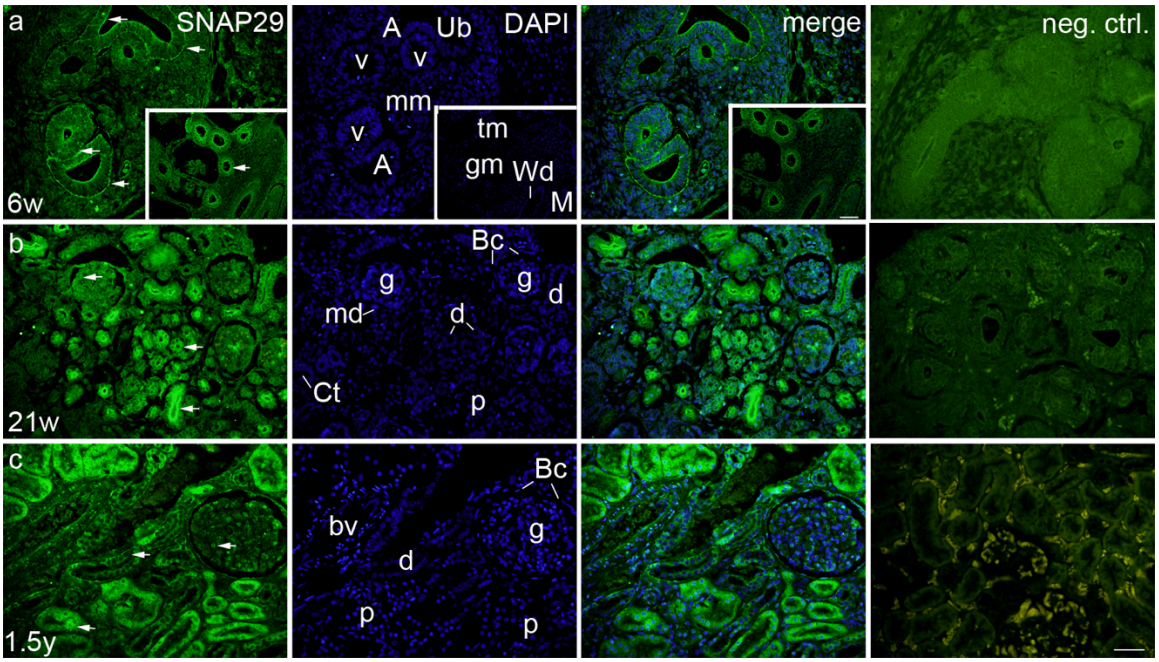
Supplementary Figure S11. CRKL immunostaining in human developing and pediatric kidney.



Localization of Crkl in the developing human, urinary tract. Transversal sections through the human metanephros in the 6th (A) and 21st (B) developmental week, and postnatal 1.5 years old human kidney (C). Positive cells – arrows; Ub – ureteric bud; A – ampulla; c – metanephric cup; Ct – collecting tubule; v – renal vesicle; mm – metanephric mesenchyme; g – glomerulus; p – proximal tubule; d – distal tubule; Bc – Bowman's capsule; neg. ctrl. – negative isotype control. Immunofluorescence staining to CRKL (green), and DAPI (blue), scale bar a-c = 25μm. (D)

Supplementary Figure S12. SNAP29 immunostaining in human developing and pediatric kidney.

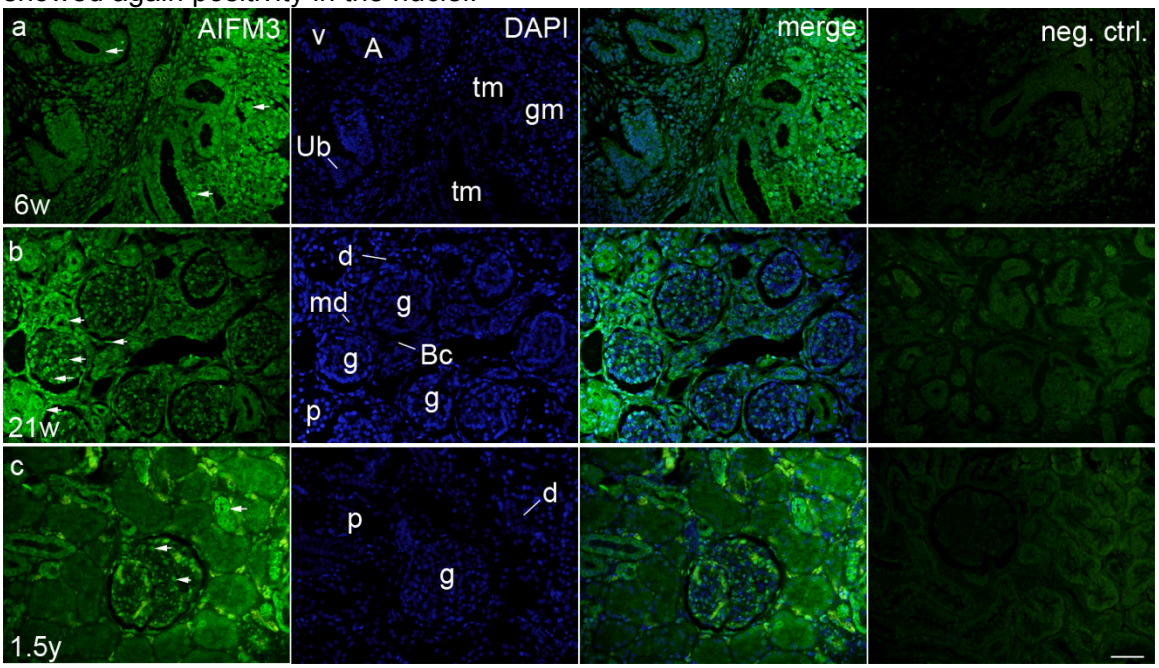
SNAP29 appeared to be both membrane-bound at the basal and apical side of developing nephrons, but also dispersed in the cytosol of other structures derived from metanephric mesenchyme at the 6th week of development. In the 21st week of development and in the 1.5 years old pediatric kidney, proximal and distal tubules and glomeruli showed diffuse SNAP29 positivity in the cytoplasm, while collecting tubules were negative. The proximal tubules were also highly positive for SNAP29.



Transversal sections through the human metanephros and mesonephros (inset) in the 6th (a), metanephros in the 21st (b), developmental week and postnatal 1.5 years old human kidney (c): positive cells – arrows, Ub – ureteric bud, A – ampulla, Ct – collecting tubule, v – renal vesicle, mm – metanephric mesenchyme, g – glomerulus, p – proximal tubule, d – distal tubule, Bc – Bowman's capsule, tm- tubules of mesonephros, gm – glomerulus of mesonephros, bv – blood vessel, md – macula densa, M – Mullerian duct, Wd – Wolffian duct, neg. ctrl. – negative isotype control. Immunofluorescence staining to SNAP29 (green), and DAPI (blue), scale bar a-c = 25μm.

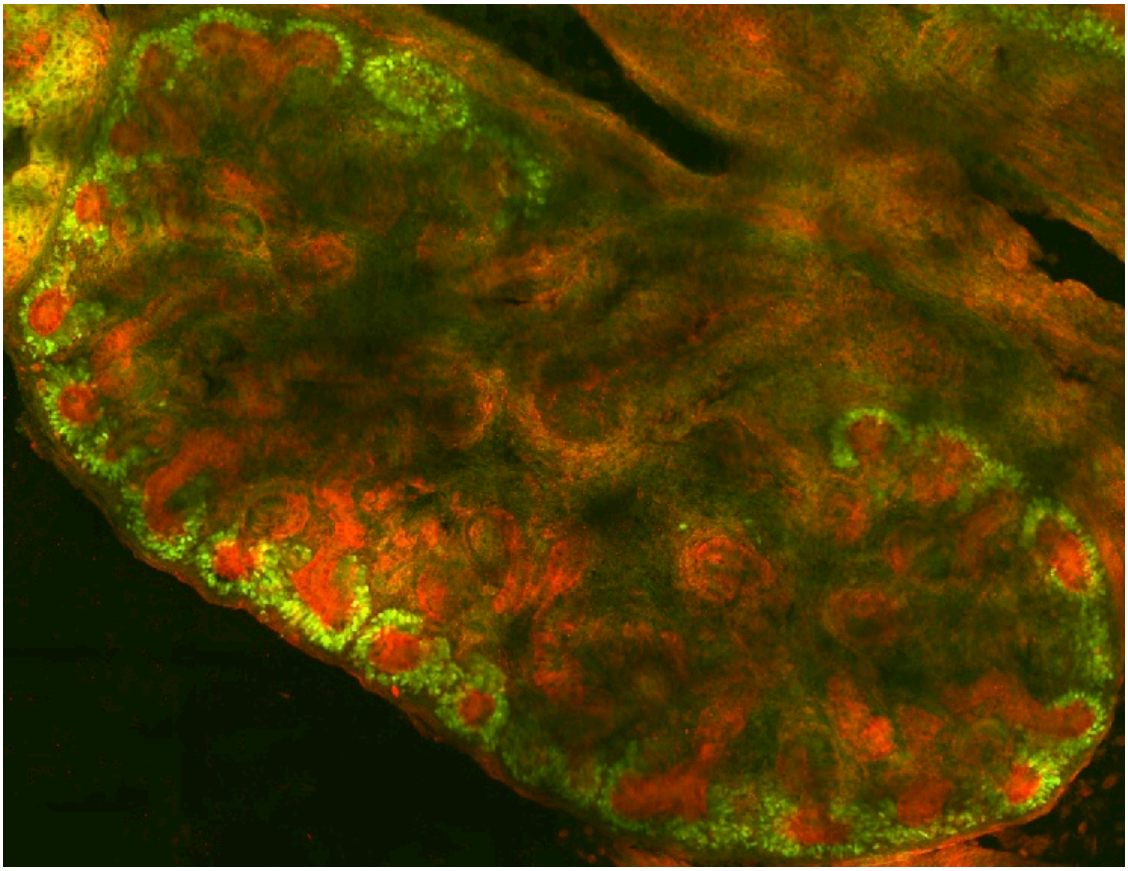
Supplementary Figure S13. AIFM3 immunostaining in human developing and pediatric kidney.

During the 6th week of development AIFM3 appeared to be mildly positive, only at the apical side in developing nephrons, and positive in the nuclei of forming glomeruli and tubules. In the 21st week of development, proximal and distal tubules showed diffuse positivity in the cytoplasm, while in the glomeruli AIFM3 showed nuclear localization. Postnatal 1.5 years old kidney displayed strong AIFM3 positivity in the cytoplasm of distal tubules while proximal tubules were only occasionally AIFM3 positive in the cytoplasm. In the glomeruli, AIFM3 showed again positivity in the nuclei.



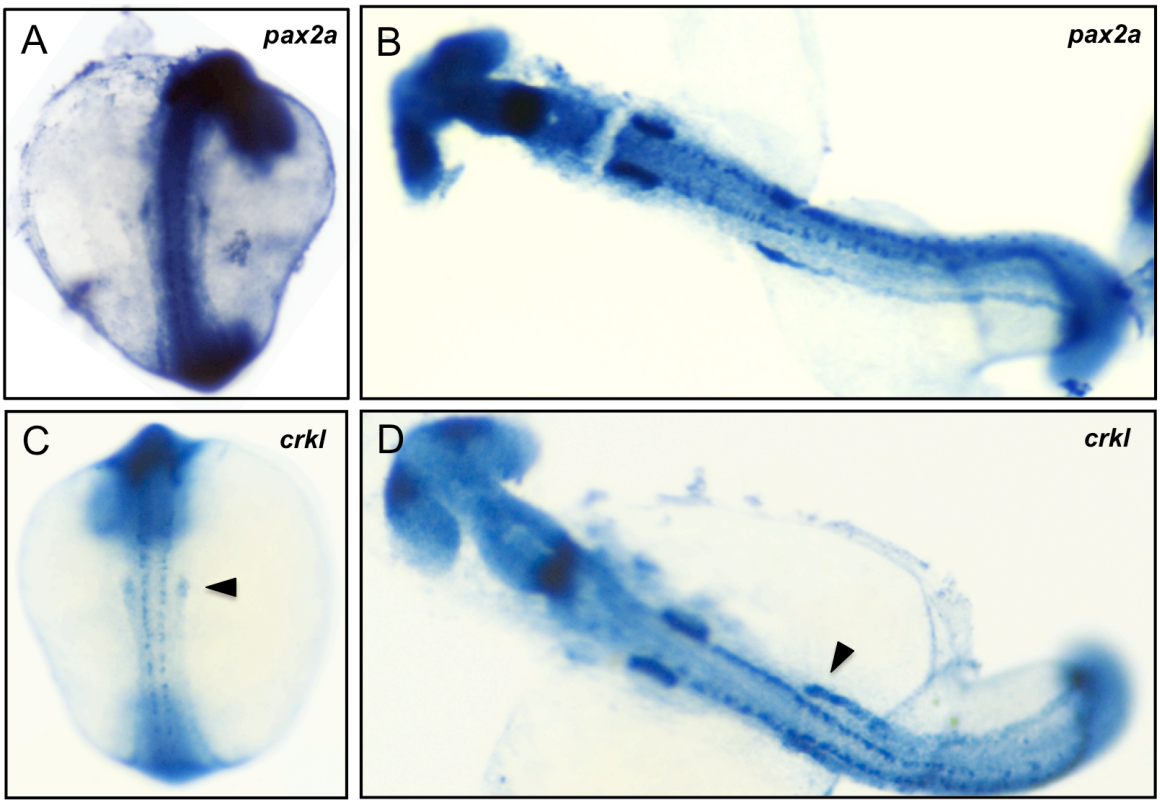
Transversal sections through the human metanephros and mesonephros. a) 6th developmental week; b) 21th developmental week; and c) postnatal 1.5 years old human kidney. Positive cells – arrows, Ub – ureteric bud, A – ampulla, v – renal vesicle, g – glomerulus, p – proximal tubule, d – distal tubule, Bc – Bowman's capsule, tm- tubules of mesonephros, gm – glomerulus of mesonephros, md – macula densa, neg. ctrl. – negative isotype control. Immunofluorescence staining to AIFM3 (green), and DAPI (blue), scale bar a-c = 25µm.

Supplementary Figure S14. Crkl immunostaining in E15.5 Six2-GFP mouse kidney.



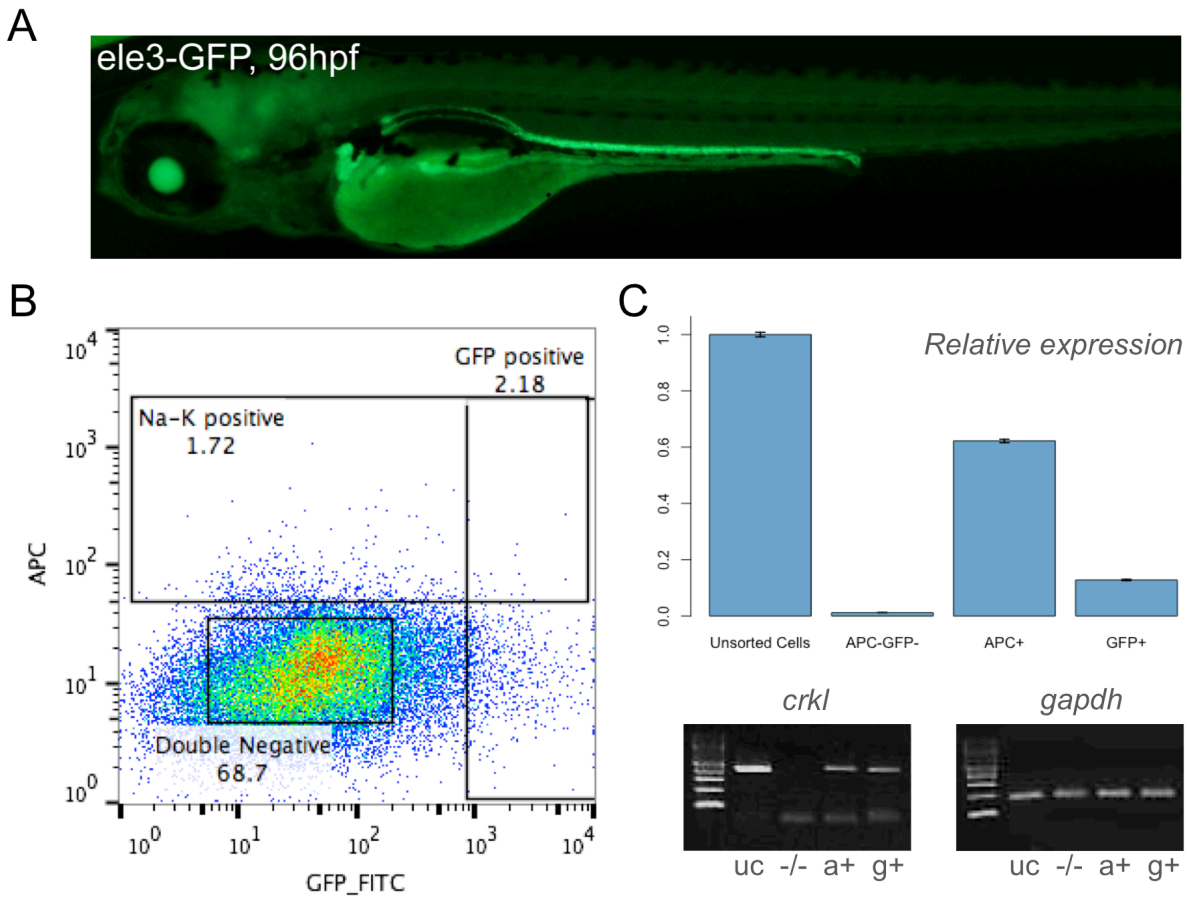
Snap-frozen E15.5 embryonic kidney from Six2-GFP mouse after immunostaining with anti-Crkl antibodies (red) on thick (~30 mm) cryostatic section. High and specific Crkl signal in ureteric bud derived structures, surrounded by Six2-positive cap mesenchyme cells; occasional positivity in S-shape bodies and mesenchyma-derived structures.

Supplementary Figure S15. Whole-mount *in situ* hybridization shows specific crkl mRNA expression in 20hpf and 24hpf zebrafish pronephros.



In situ hybridization in 20hpf (A,C) and 24hpf (B,D) zebrafish embryos for *pax2a* (A,B) and *crkl* (C,D). Arrowheads indicate strong and specific *crkl* expression in the proximal pronephric tubule, comparable to the expression and localization of *pax2a*.

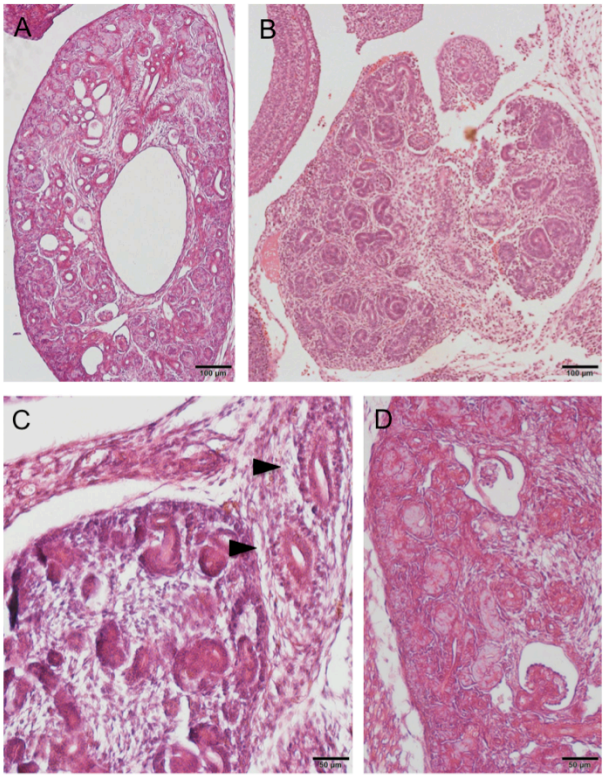
Supplementary Figure S16. *crkl* mRNA expression from pronephric tubule cells after flow cytometry cell sorting.



To assess specificity for *crkl* mRNA expression in pronephros we conducted flow cytometry cell sorting in 96hpf zebrafish using ele3-GFP transgenic fish (labeling the entire pronephros) (A), or via selection with NaK ATPase antibody (a6F, DSHB). (B) Flow cytometry plot showing the three populations sorted for RNA studies: NaK ATPase positive cells (APC+ or A+); GFP positive cells (GFP+ or G+); double negative cells (APC-GFP- or -/-). (C) Quantitative PCR and relative electrophoresis gels show *crkl* expression in both NaK ATPase and GFP positive cells, indicating specific *crkl* expression in the zebrafish pronephros.

As per the other two candidates, snap29 and aifm3, we have found them to be expressed at extremely low levels during zebrafish development and lie below the limit of detection given the small number of cells that we can obtain during cell sorting.

Supplementary Figure S17. Developmental abnormalities observed in *Crkl* mutant mice.



(A) E15.5 Crkl$^{f/f}$-Six2-Cre: The kidney displays dilatation of the pelvis, multifocal tubular cysts and failure of medullary development. (B) E15.5 Crkl$^{f/+}$-Six2-Cre: The kidney appears septated with lobulated external renal contours and disorganized nephrogenic zone. (C) E14.5 Crkl $^{f/+}$-E2a-Cre: The nephrogenic zone is thinned and double ureters are identifiable in the perirenal region (arrowheads). (D) E14.5 Crkl$^{f/f}$-Hoxb7-Cre: glomerular cysts (arrowhead), severe dysplasia.

Supplementary Table S1. CAKUT cohorts.

Cohort	Number of Patients	Phenotype (N)	Platform
Discovery CAKUT CNV cohort	1,752	RHD (765) OU (397) VUR (473) DCS (63) LUTM (44) Others (10)	SNP microarrays (Illumina 550, 610Quad, 660, OmniExpress, Omni1, Omni2.5; Affymetrix 6.0)
Replication RHD CNV cohort	328	RHD (328)	SNP microarrays (Illumina MEGA Power Chip V1.0)
RHD exome sequencing	60	RHD (60)	Exome sequencing (Roche V2 Capture Kit; 60X average depth)
RHD targeted resequencing	526	RHD (526)	Fluidigm microfluidic PCR capture

CAKUT= congenital anomalies of the kidney and urinary tract; CNV= copy number variation; RHD= renal hypodysplasia; OU= obstructive uropathy; VUR= vesicoureteral reflux; DCS= duplicated collecting system; LUTM= lower urinary tract malformations; SNP= single nucleotide polymorphism; PCR= polymerase chain reaction.

Supplementary Table S2. Prevalence of the chromosomes 22q11.2 and 17q12 deletions in patients with CAKUT as compared to population controls.

We confirmed that the 17q12 RCAD deletion is the most common microdeletion syndrome in human CAKUT and we identified deletions at the 22q11.2 DGS locus as the second most common structural variant predisposing to kidney and urinary tract malformations; the frequencies were estimated based on the cohort of 2,080 CAKUT cases, including 1,093 patients with RHD, and 22,094 population controls.

Genomic Disorder	Controls (%) N=22,094	Discovery Cohort (%) N=1,752	P-value	OR (CI)	RHD Discovery (%) N=765	P-value	OR(CI)	RHD Combined (%) N=1,093	P-value	OR(CI)
Chr. 22q11.2 (DGS)	3 (0.014)	11 (0.627)	9.7×10^{-11}	45.4 (12-253)	9 (1.176)	1.2×10^{-11}	85.3 (21-491)	12 (1.098)	4.5×10^{-14}	81.5 (23-451)
Chr. 17q12 (RCAD)	0 (0)	19 (1.084)	$<2.2 \times 10^{-16}$	Inf (58-inf)	18 (2.353)	$<2.2 \times 10^{-16}$	Inf (127-inf)	24 (2.196)	$<2.2 \times 10^{-16}$	Inf (124-inf)

CAKUT=congenital anomalies of the kidney and urinary tract; RHD=renal hypodysplasia; DGS=DiGeorge syndrome; RCAD=renal cysts and diabetes syndrome; OR=odds ratio; CI=confidence interval.

Supplementary Table S3. Genes included in the DGS C-D smallest region of overlap.

Gene	Position (hg19)	Number of exons	HI probability	HI percentile	Number of individuals with LOF in Exac (%)	Human Disease (mode of inheritance)	Mouse Model
PI4KA	21061979-21213100	55	0.194	44.2%	79 (0.1301)	Polymicrogyria, perisylvian, with cerebellar hypoplasia and arthrogryposis (autosomal recessive)[22]	NA
SERPIND1	21128383-21142008	4	0.632	10.4%	16 (0.0263)	Thrombophilia due to heparin cofactor II deficiency, 612356 (autosomal dominant)[23]	Viable and fertile; No kidney phenotype reported[24]
SNAP29	21213292-21245501	5	0.032	97.9%	9 (0.0148)	Cerebral dysgenesis, neuropathy, ichthyosis, and palmoplantar keratoderma syndrome, 609528 (autosomal recessive)[25]	NA
CRKL	21271714-21308037	3	0.942	2.4%	1 (0.0016)	NA	Embryonic lethality at E15; cardiac and craniofacial malformations[26]
AIFM3	21319418-21335649	20	0.118	65.4%	82 (0.1351)	NA	NA
LZTR1	21336302-21353327	21	0.164	50.4%	213 (0.3509)	Noonan syndrome 10 (autosomal dominant)[27]; Susceptibility to Schwannomatosis (autosomal dominant)[28]	NA
THAP7	21354061-21356404	4	0.282	31.3%	6 (0.0099)	NA	NA
P2RX6	21369442-21382302	12	0.051	93.5%	674 (1.1104)	NA	NA
SLC7A4	21383007-21386847	4	0.403	21.0%	191 (0.3245)	NA	NA

HI=haploinsufficiency: scores and percentiles calculated according to Huang et al[29].

Supplementary Table S4. Patients with C-D deletion and kidney and urinary tract malformations from the 22Q and You database.

Patient	Deletion type	Kidney Phenotype	Side (R,L,B)	Urinary Tract Phenotype	Extra-renal Phenotype	Sex	Ethnicity	Age at Diagnosis	Outcome
12164-A	C-D	Renal agenesis	R	Pelvic Kidney (L)	-	M	Caucasian	13 months	Speech delay
12283-A	C-D	Renal hyperechogenicity	B	Hydronephrosis (L)	TOF, pulmonary artery stenosis	M	Caucasian	1 year	Dysphagia

The 22Q and You Center database at the Children Hospital of Philadelphia currently comprises 1,305 patients with deletions at the chromosome 22q11.2 locus. For 694 individuals abdominal imaging studies were available. Of those, ten patients had the C-D deletion and two (20%) were found to have kidney and urinary tract defects. R=Right; L=Left; B=Bilateral; TOF=tetralogy of Fallot.

Supplementary Table S5. Population controls with chromosome 22q11.2 deletions.

Control	Start (Mb)	End (Mb)	Size (Mb)	Deletion type	Kidney Phenotype	Extra-renal Phenotype	Sex	Race	Age	Outcome
C1	18.88	21.47	2.59	A-D	CKD stage IV	Early-onset Parkinson's disease; congenital hypoparathyroidism	M	Caucasian	41	Deceased
C2	20.72	21.48	0.76	B-D	NA	NA	F	African American	NA	NA
C3	21.06	21.47	0.40	C-D	NA	NA	M	Caucasian	NA	NA

Clinical data were available for individual C1. Chart review showed that the patient was affected by congenital hypoparathyroidism, early-onset Parkinson's disease, and advanced CKD (serum creatinine 2.7mg/dL, eGFR by CKD-EPI estimation = 28 mL/min, corresponding to a CKD stage IV). The individual did not suffer from other predisposing factors to CKD such as diabetes mellitus, uncontrolled hypertension, or primary glomerulonephritis. This individual, therefore, represented a patient with undiagnosed DGS, strongly implicating variants at 22q11.2 as deterministic for kidney disease. After removal of this patient from the controls dataset, the strength of the statistical association between 22q11.2 deletions and renal malformations further improved (12/1,093 patients vs 2/22,093 controls, $p = 8.5 \times 10^{-15}$; OR = 123.7).
Location is reported in Mb (megabases) coordinates from the Human Genome 19 release. CKD= chronic kidney disease; M= male, F= female; NA= not available.

Supplementary Table S6. C-D MRO genes variants from exome sequencing in 60 RHD patients.

Gene	Position (hg19)	Variant	Consequence	PolyPhen-2 Prediction (Score)	CADD Score	rsID	ESP MAF % (Eur)	Exac MAF % (Tot)	ID	Sex	Urinary Tract Phenotype	Extrarenal Phenotype
PI4KA	21067061	c.1852C>T	p.D1866N	0.009	14.96	rs368442314	0.0132	0.00489	P17*	M	RHD, VUR	NA
									P18	F	MCDK, bladder diverticuli	Arrhythmia, presacral fistula
PI4KA	21119138	c.2675T>C	p.Y892C	1	26.6	NOVEL	NA	NA	P19	M	RHD	Left ventricular hypertrophy, clubfeet
SERPIND1	21138311	c.941G>A	p.R314Q	0.988	31	rs142500721	0.0116	0.00163	P20	M	RHD, VUR	NA
SERPIND1	21141256	c.1402C>T	p.R468C	0.991	25.5	NOVEL	NA	0.00163	P21	M	RHD, pelvis dilation	NA
									P22	F	RHD, DCS	NA
SNAP29	21235350	c.448A>G	p.I150V	0.131	4.176	NOVEL	NA	0.0033	P23	F	RA	NA
AIFM3	21330753	c.956C>T	p.T319M	1	23.9	rs139810844	0.0116	0.00165	P20	M	RHD, VUR	NA
AIFM3	21332226	c.1409C>G	p.T476S	0.043	6.79	NOVEL	NA	NA	P24	M	RHD	NA
LZTR1	21342407	c.509G>A	p.R170Q,Splice	0.977	24.2	NOVEL	NA	0.00327	P17*	M	RHD, VUR	NA
LZTR1	21350033	c.19432A>G	Splice	NA	19.34	NOVEL	NA	0.00165	P25	M	RHD	Atrial septal defect
THAP7	21354705	c.394C>G	p.G132R	0.118	13.76	NOVEL	NA	0.00569	P26	M	RA	NA
P2RX6	21369548	c.85A>T	p.K29*	NA	25.9	rs148541070	0.6047	0.47335	P27	M	RHD	NA
									P24	M	RHD	NA
P2RX6	21380135	c.854G>A	p.R285H	1	32	NOVEL	NA	0.0077	P28	M	RA	NA
SLC7A4	21385954	c.148T>C	p.M50V	0.868	15.15	rs143734260	0.3721	0.22333	P29	M	RHD, VUR	NA

MRO= minimal region of overlap; RHD= renal hypodysplasia; RA= renal agenesis; VUR= vesicoureteral reflux; DCS= duplicated collecting system. *Non segregating variants.

Supplementary Table S7. LOF variants identified via targeted resequencing of 526 RHD patients.

Gene	Position (hg19)	Variant	Consequence	CADD Score	rsID	ESP MAF %	Exac MAF %	ID	Sex	Urinary Tract Phenotype	Extrarenal Phenotype
SERPIND1	21134015	c.415C>T	R139*	37	NOVEL	NA	NA	P30	M	RHD	NA
SERPIND1	21138259	c.890-1G>A	Splice	21.4	NOVEL	NA	0.0032	P31	F	RHD	NA
CRKL	21272313	c.91C>T	Gln31*	44	NOVEL	NA	NA	P13	F	RA	NA
AIFM3	21327721	c.157G>T	Glu53*	38	NOVEL	NA	0.0009	P32	F	RHD	Chest deformity, hemangioma, growth retardation
P2RX6	21369548	c.85A>T	K29*	25.9	rs148541070	0.4152	0.472	P33	F	RHD, OU	NA
								P34	M	RA, VUR	NA
								P35	F	RHD	NA
								P36	M	RA	NA
								P37	F	RA	NA
P2RX6	21377325	c.557+1G>A	Splice	16.94	rs149671550	0.0384	0.0374	P38	F	RHD	NA
								P39	F	RHD	NA

LOF= loss-of-function; RHD= renal hypodysplasia; RA= renal agenesis; VUR= vesicoureteral reflux; OU= obstructive uropathy

Supplementary Table S8. Rare *CRKL* coding variants identified in patients with renal agenesis or hypodysplasia (RHD).

hg19 position	Variant	Consequence	CADD	PhastCons	Exac MAF (%)	CUMC Controls MAF (%)	Patient	Sex	Kidney Phenotype	Side (R, L, B)	Urinary Tract Phenotype	Extra-renal Phenotype
21272298	c.A76G	p.T26A	11.41	1	Absent	Absent	P12	M	Renal hypodysplasia	R	Vesicoureteral reflux	Preauricular appendix
21272313	c.C91T	p.Q31*	44	1	Absent	Absent	P13	F	Renal agenesis	R	-	-
21288105	c.A350G	p.N117S	6.299	0.996	Absent	Absent	P14	F	Renal agenesis	L	-	Polycystic ovary syndrome
21288516	c.C761G	p.T254S	24	0.995	Absent	Absent	P15	M	Multicystic dysplastic kidney	L	-	-
21304068	c.A847G	p.K283E	11.28	1	Absent	Absent	P16	F	Renal hypodysplasia	R	-	-

Variant position is reported in bp (basepairs) coordinates from the Human Genome release 19. CADD= Combined Annotation Dependent Depletion; MAF=minor allele frequency estimated from over 60,500 individuals from the Exome Aggregation Consortium (http://exac.broadinstitute.org); R=Right, L=Left, B= Bilateral; M=Male, F=Female. CUMC=Columbia University Medical Center controls (N=1,728).

Supplementary Table S9. LOF variants identified in patient P13 by exome sequencing.

Gene	Chr	Position (hg19)	Variant	Consequence	CADD Score	rsID	Exac MAF % (Tot)	Number of individuals with LOF in Exac (%)
MAMDC4	9	139751468	c.G1947A	p.W649*	20.2	NOVEL	0.002545	547 (0.9011)
NRAP	10	115393885	c.C1513T	p.Q505*	39	NOVEL	NA	290 (0.4777)
DPH1	17	1943798	c.922-1G>T	splice	15.79	NOVEL	0.0008293	164 (0.2702)
ZNF227	19	44732650	c.C112T	p.R38*	15.88	NOVEL	NA	145 (0.2389)
CEP250	20	34054899	c.599+2T>C	splice	20.5	NOVEL	NA	120 (0.1977)
TMPRSS15	21	19715974	c.1278-1G>A	splice	18.57	NOVEL	NA	245 (0.4036)
CRKL	22	21272313	c.C91T	p.Q31*	44	NOVEL	NA	1 (0.0016)

Supplementary Table S10. *CRKL* rare variants burden tests.

	RHD (%) N=586	CUMC Controls (%) N=1,728	P-value	OR (CI)	Eur. ExAC Controls (%) N=33,552	P-value	OR(CI)
CRKL	5 (0.85)	1 (0.058)	4.9×10^{-3}	14.8 (1.6-700.7)	55 (0.164)	3.7×10^{-3}	5.2 (1.6-12.9)

Tests for burden of rare (MAF<0.001) functional variants (LOF + missense) by Fisher's exact test in 586 RHD cases compared to 1,728 ethnically and geographically matched controls, and to 33,552 controls individuals of European origin from the ExAC database. Since the association for the chromosome 22q11.2 deletion exceeds genome-wide significance, the Bonferroni-corrected threshold for the nine genes in the 22q11.2 C-D locus is 5.5×10^{-3} (0.05/9). Our test statistics exceeds Bonferroni-corrected significance threshold in both cases. RHD= renal hypodysplasia; CUMC= Columbia University Medical Center; OR=Odds ratio; CI=Confidence interval; ExAC=Exome Aggregation Consortium.

REFERENCES

1. Sanna-Cherchi S, Kiryluk K, Burgess KE, et al. Copy-number disorders are a common cause of congenital kidney malformations. Am J Hum Genet 2012;91:987-97.
2. Verbitsky M, Sanna-Cherchi S, Fasel DA, et al. Genomic imbalances in pediatric patients with chronic kidney disease. J Clin Invest 2015;125:2171-8.
3. Westland R, Verbitsky M, Vukojevic K, et al. Copy number variation analysis identifies novel CAKUT candidate genes in children with a solitary functioning kidney. Kidney Int 2015;88:1402-10.
4. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 2007;17:1665-74.
5. Sanders SJ, Ercan-Sencicek AG, Hus V, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 2011;70:863-85.
6. Sanna-Cherchi S, Sampogna RV, Papeta N, et al. Mutations in DSTYK and dominant urinary tract malformations. N Engl J Med 2013;369:621-9.
7. Westland R, Bodria M, Carrea A, et al. Phenotypic expansion of DGKE-associated diseases. J Am Soc Nephrol 2014;25:1408-14.
8. Boyden LM, Choi M, Choate KA, et al. Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities. Nature 2012;482:98-102.
9. Halbritter J, Diaz K, Chaki M, et al. High-throughput mutation analysis in patients with a nephronophthisis-associated ciliopathy applying multiplexed barcoded array-based PCR amplification and next-generation sequencing. J Med Genet 2012;49:756-67.
10. Halbritter J, Porath JD, Diaz KA, et al. Identification of 99 novel mutations in a worldwide cohort of 1,056 patients with a nephronophthisis-related ciliopathy. Hum Genet 2013;132:865-84.
11. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-9.
12. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310-5.
13. Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 2005;15:1034-50.
14. Liu Y, Pathak N, Kramer-Zucker A, Drummond IA. Notch signaling controls the differentiation of transporting epithelia and multiciliated cells in the zebrafish pronephros. Development 2007;134:1111-22.
15. Jao LE, Wente SR, Chen W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proc Natl Acad Sci U S A 2013;110:13904-9.
16. Wang D, Jao LE, Zheng N, et al. Efficient genome-wide mutagenesis of zebrafish genes by retroviral insertions. Proc Natl Acad Sci U S A 2007;104:12428-33.
17. Williams JR. The Declaration of Helsinki and public health. Bull World Health Organ 2008;86:650-2.
18. Drummond IA, Majumdar A, Hentschel H, et al. Early development of the zebrafish pronephros and analysis of mutations affecting pronephric function. Development 1998;125:4655-67.
19. Jerome LA, Papaioannou VE. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. Nat Genet 2001;27:286-91.
20. Huang CJ, Tu CT, Hsiao CD, Hsieh FJ, Tsai HJ. Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. Developmental dynamics : an official publication of the American Association of Anatomists 2003;228:30-40.
21. Lawson ND, Weinstein BM. In vivo imaging of embryonic vascular development using transgenic zebrafish. Developmental biology 2002;248:307-18.
22. Pagnamenta AT, Howard MF, Wisniewski E, et al. Germline recessive mutations in PI4KA are associated with perisylvian polymicrogyria, cerebellar hypoplasia and arthrogryposis. Human molecular genetics 2015;24:3732-41.
23. Kondo S, Tokunaga F, Kario K, Matsuo T, Koide T. Molecular and cellular basis for type I heparin cofactor II deficiency (heparin cofactor II Awaji). Blood 1996;87:1006-12.
24. Vicente CP, He L, Tollefsen DM. Accelerated atherogenesis and neointima formation in heparin cofactor II deficient mice. Blood 2007;110:4261-7.

25. Sprecher E, Ishida-Yamamoto A, Mizrahi-Koren M, et al. A mutation in SNAP29, coding for a SNARE protein involved in intracellular trafficking, causes a novel neurocutaneous syndrome characterized by cerebral dysgenesis, neuropathy, ichthyosis, and palmoplantar keratoderma. Am J Hum Genet 2005;77:242-51.
26. Guris DL, Fantes J, Tara D, Druker BJ, Imamoto A. Mice lacking the homologue of the human 22q11.2 gene CRKL phenocopy neurocristopathies of DiGeorge syndrome. Nat Genet 2001;27:293-8.
27. Yamamoto GL, Aguena M, Gos M, et al. Rare variants in SOS2 and LZTR1 are associated with Noonan syndrome. J Med Genet 2015;52:413-21.
28. Piotrowski A, Xie J, Liu YF, et al. Germline loss-of-function mutations in LZTR1 predispose to an inherited disorder of multiple schwannomas. Nat Genet 2014;46:182-7.
29. Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and predicting haploinsufficiency in the human genome. PLoS Genet 2010;6:e1001154.