Identification of pathological transcription in autosomal dominant polycystic kidney disease epithelia

Sebastian Friedrich1,4,5, Hannah Müller1,5, Caroline Riesterer3,5, Hannah Schüller1, Katja Friedrich1, Carlotta Leonie Wörner1, Tilman Busch1, Amandine Viau2, E. Wolfgang Kuehn1,3, Michael Köttgen1,3* & Alexis Hofherr1,3

Autosomal dominant polycystic kidney disease (ADPKD) affects more than 12 million people worldwide. Mutations in PKD1 and PKD2 cause cyst formation through unknown mechanisms. To unravel the pathogenic mechanisms in ADPKD, multiple studies have investigated transcriptional mis-regulation in cystic kidneys from patients and mouse models, and numerous dysregulated genes and pathways have been described. Yet, the concordance between studies has been rather limited. Furthermore, the cellular and genetic diversity in cystic kidneys has hampered the identification of mis-expressed genes in kidney epithelial cells with homozygous PKD mutations, which are critical to identify polycystin-dependent pathways. Here we performed transcriptomic analyses of Pkd1- and Pkd2-deficient mIMCD3 kidney epithelial cells followed by a meta-analysis to integrate all published ADPKD transcriptomic data sets. Based on the hypothesis that Pkd1 and Pkd2 operate in a common pathway, we first determined transcripts that are differentially regulated by both genes. RNA sequencing of genome-edited ADPKD kidney epithelial cells identified 178 genes that are concordantly regulated by Pkd1 and Pkd2. Subsequent integration of existing transcriptomic studies confirmed 31 previously described genes and identified 61 novel genes regulated by Pkd1 and Pkd2. Cluster analyses then linked Pkd1 and Pkd2 to mRNA splicing, specific factors of epithelial mesenchymal transition, post-translational protein modification and epithelial cell differentiation, including CD34, CDH2, CSF2RA, DLX5, HOXC9, PIK3R1, PLCB1 and TLR6. Taken together, this model-based integrative analysis of transcriptomic alterations in ADPKD annotated a conserved core transcriptomic profile and identified novel candidate genes for further experimental studies.

Autosomal dominant polycystic kidney disease (ADPKD) is the most common hereditary nephropathy in humans. Most patients with ADPKD are born healthy, but progressive cystic transformation of both kidneys induces a continuous decline in renal function leading to kidney failure. Mutations in two genes, PKD1 and PKD2, cause ADPKD. The protein products of these genes, polycystin-1 and polycystin-2, form a receptor-ion channel complex that is essential to establish renal tubular morphology and prevent cystogenesis.

Many cellular signalling pathways have been implicated in the pathogenesis of ADPKD, including Wnt, mTOR, JAK/STAT and Hippo. Yet, the precise molecular function of PC-1/TRPP2 and the signaling pathways critical for disease initiation have remained unclear. Unbiased, discovery-based approaches offer unique opportunities to address these key questions of ADPKD pathology. RNA sequencing technology has made large scale analysis of gene expression a powerful approach to systematically investigate disease mechanisms. Consequently, multiple studies have explored transcription in cystic kidneys from Pkd-deficient mice and...
ADPKD patients. These studies have reported a large number of differentially expressed genes in grossly cystic samples. However, the concordance between studies for specific gene sets has been low. This complex variability is likely to be explained by differences in experimental design, tissue sampling and model system. Hence, transcriptomic profiling of ADPKD has confirmed broad cellular alterations in cystic tissue, but has provided limited novel insights into PKD-dependent tubule-cell-specific factors causing cystic kidney disease.

A key challenge of specific profiling of PKD1- and PKD2-dependent transcription has been the genetic mosaicism of organisms with ADPKD. In ADPKD, cysts arise from individual renal epithelial cells upon somatic loss of heterozygosity for PKD1 or PKD2. This focal process has been modeled in mice by homozygous Pkd inactivation in kidney tubular epithelial cells. However, in both cases only a small fraction of renal cells harbor the homozygous PKD mutations required to trigger cystogenesis. The majority of cells in polycystic kidneys are still heterozygous for PKD gene mutations—or wildtype in the case of tubule-specific knockout mice. Furthermore, due to secondary processes, such as proliferation of cyst epithelia, inflammation and fibrosis, cystic kidneys show considerable secondary changes in cellular composition. These processes contribute to disease progression, but further confound the transcriptomic profiling of homozygous mutant tubular cells at cyst initiation, a prerequisite to gain insights into PKD gene function.

Here, we investigate cell-autonomous ADPKD-dependent transcriptional changes based on two ground truths: (1) mutations in PKD1 and PKD2 cause ADPKD, and (2) renal cystogenesis is initiated in PKD-deficient tubular epithelial cells. First, we characterized transcriptomic profiles of genetically engineered, clonal ADPKD cells in a well-controlled, reductionist approach. RNA-seq of Pkd1- and Pkd2-deficient mouse inner medullary collecting duct 3 (mIMCD3) epithelial cells identified a core ADPKD data set of 178 reproducible and concordantly regulated genes. Using this core data set, we next performed a cross-species meta-analysis to integrate the wealth of transcriptomic data from Pkd-deficient animals and human ADPKD kidneys, which confirmed 31 highly consistent, previously described, transcriptionally regulated genes. Notably, we also identified 61 novel genes that are differentially expressed in Pkd1 and Pkd2 deficiency (Fig. 1). Gene ontology and functional pathway enrichment analyses linked specific factors of epithelial mesenchymal transition (EMT), post-translational protein modification and epithelial cell differentiation to ADPKD pathogenesis. This combination of cellular PKD models with transcriptomic data from cystic kidneys showed the power of integrative transcriptomics to validate candidate targets for in-depth studies of pathogenic mechanisms in ADPKD, including CD34, CDH2, CSF2RA, DLX5, HOXC9, PIK3R1, PLCB1 and TLR6.

Results
Transcriptional characterization of ADPKD by deep RNA sequencing of Pkd1- and Pkd2-deficient renal tubular cells.

Mouse inner medullary collecting duct 3 (mIMCD3) cells represent a well-established, differentiated in vitro model of the distal nephron. mIMCD3 cells are highly differentiated, phenotypically stable and show low heterogeneity, with preserved apico-basolateral polarity, formation of primary cilia, collecting duct-like transport activity and tubule formation in 3D culture. This enables stringent analyses of gene-dependent transcription in a controlled environment, free of systemic variations (Fig. S1a).

Analysis of active gene regulatory networks as major molecular determinants of cell-type identity further confirmed the kidney transcriptional signature of mIMCD3 cells (Figure S1b). For the targeted evaluation of ADPKD-dependent cellular pathophysiology, we here compared wild-type mIMCD3 cells with constitutive Pkd1- and Pkd2-deficient cell lines.

Repeated deep RNA-seq of wild-type, Pkd1−/− and Pkd2−/− cells yielded rich and highly consistent transcriptomic profiles (Fig. 2a). Sequencing depth, read distribution between samples, sample-to-sample-distances, dispersion of variance and histogram of p-values verified excellent quality of data (Figure S2a–e).

Collected RNA-seq data confirmed the genomic deletion of Pkd1 and Pkd2, and showed significant mutual interdependence of Pkd1 and Pkd2 mRNA expression, indicating a transcriptional Pkd1/2 feedback loop (Figs. 2b,c, S1c,d). Full gene deletion removed all Pkd1 transcripts in mutant cells and was associated with a decrease in Pkd2 expression (−31.7%) (Figs. 2b,c, S1c). Deletion of Pkd2, on the other hand, was associated with a significant transcriptional up-regulation of Pkd1 mRNA (+54.6%) (Fig. 2b,c). RNA-seq of Pkd2−/− cells also confirmed that genomic deletion of chromosome 5, bases 104,478,281–104,490,878, removed Pkd2 exons 4–9 (of a total 15) and, hence, TRPP2 channel transmembrane domains 2–6 (of a total 6), which generated biochemically TRPP2-deficient cells (Figs. 2c, S1c).

Unbiased statistical integration of Pkd1- and Pkd2-dependent transcription identified a core ADPKD data set of 178 consistently regulated genes (CD178).

Mutations in Pkd1 and Pkd2 cause ADPKD. However, to date, no comprehensive analysis of PKD1- and PKD2-dependent transcription in ADPKD has been reported; several studies have focused on Pkd1 and limited data is available on Pkd2-
Healthy       ADPKD

Wt          $Pkd1^{-/-}$  $Pkd2^{-/-}$

Cell Culture Model

RNA-seq

Data Analysis

76            178         323

Functional Enrichment

$Pkd1^{-/+}$  $Pkd2^{-/-}$

Transcriptomics

31 Confirmed and 61 New ADPKD Related Genes
responsive transcription\textsuperscript{18,19}. We hypothesized that the integrative analysis of PKD1- and PKD2-deficiency may substantially improve the high-confidence identification of consistent transcriptional changes in ADPKD pathogenesis.

Technically, the low biological variability of wild-type, Pkd1\textsuperscript{−/−} and Pkd2\textsuperscript{−/−} mIMCD3 cells provided a sound basis to delineate ADPKD-specific renal tubular transcription (Figs. 2a, S2a–e). This was important, as gene expression is an inherently stochastic process and is known to vary even between cells of the same population\textsuperscript{29}. In general, observed variations in gene expression can be decomposed as:

\[ \text{Var(Expr)} = \text{Across group variability} + \text{Measurement error} + \text{Biological variability}. \]

With group variability being here the variation in gene expression due to the three genotypes under consideration. Measurement error being rather low in next-generation RNA-seq\textsuperscript{32}. And biological variability being in our experiments restricted to stochastic cell-to-cell variation; and, therefore, minimal in comparison to tissue or cell type variability in whole organ analyses\textsuperscript{31}. Thus, our approach of repeated sampling, high data quality and model simplicity should minimize the confounding effects of measurement error and biological variability, focusing generated data on the primary outcome of interest: the estimated Pkd-dependent difference in gene expression between genotypes.

Figure 2. Transcriptomic profiling of genetically engineered ADPKD cells showed significant interdependence of Pkd1 and Pkd2 mRNA expression. (a) Unsupervised hierarchical clustering showed low variability and genotype-dependent grouping of mIMCD3 Wt, Pkd1\textsuperscript{−/−} and Pkd2\textsuperscript{−/−} cells. (b) Pkd1 knockout was generated by a full genomic deletion. Loss of Pkd2 was associated with a significant transcriptional up-regulation of Pkd1 mRNA (782.25 [SEM = 97] mean normalized counts in wildtype vs. 1,209 [SEM = 175] in Pkd2\textsuperscript{−/−} cells). (c) Pkd2 knockout was based on genomic deletion of Pkd2 exons 4–9 and no counts were detected for these exons. + : adjusted p-values for single exons 4–9 were < 1.75 × 10\textsuperscript{−5} in Pkd2\textsuperscript{−/−} versus Wt and Pkd2\textsuperscript{−/−} versus Pkd1\textsuperscript{−/−} cells.
Two complementary mathematical methodologies were used to query the raw RNA-seq data for ADPKD-specific gene expression: (1) statistical analysis of differential gene expression (DGE); and (2) unbiased data simplification by principal component analysis (PCA)33–35.

For DGE, statistical comparison of wild-type, Pkd1−/− and Pkd2−/− cells identified significant differences in gene expression (Fig. 3a,b). Given that the likelihood of statistically significant results to reflect a true effect is dependent on the respective effect size, we defined DGE positive results by false discovery rate (FDR) < 0.05, and log2 fold change ≥ 135,36. By this definition, 781 genes in Pkd1−/− and 1150 in Pkd2−/− were identified as differentially expressed vs. wildtype (Fig. 3a,b, Tables S1, S2). From these, the DGE-based Pkd-specific renal tubular gene expression profile was compiled by filtering for concordant changes. Hence, DGE identified 254 (38 up- and 216 down-regulated; DGE254) Pkd-specific genes (Fig. 3c and Table S3).

For PCA, raw RNA-seq data was input to unsupervised eigenvector-based multivariate analyses. This dimensionality reduction confirmed the close correlation of individual genotypes (Fig. 3d). Two principal components (PC) explained most sample variance, PC1 with 69% and PC2 with 29%. The biological meaning of this data structure may be inferred by the two key factors differentiating the evaluated mIMCD3 cell lines: (1) random clonal variation; and (2) Pkd-dependent gene expression. Because P2C correlated with Pkd1- and Pkd2-deficiency, genes were ranked by PC2 weighting and the top 20% (n = 501; PCA501) were selected for comparison with the DGE254 data set (Fig. 3e and Table S4).

Comparison of the independent DGE254 and PCA501 data sets identified 178 distinct genes specifically and concordantly responding to Pkd1- and Pkd2-deficiency in mIMCD3 cells (core data set; CD178) (Figs. 3f, S3 and Table S5). Crucially, the integration of Pkd1 and Pkd2 signatures should strongly focus CD178 on transcripts relevant for ADPKD pathogenesis.

**Function-based clustering of the CD178 core ADPKD data set isolated key pathways mediating cyst pathophysiology.** Conceptually, CD178 provided an unbiased look at the transcriptional activity of Pkd-deficient kidney cells. To uncover the cellular signaling pathways represented by these core ADPKD genes, we next linked prior biological knowledge to the newly generated CD178 expression data. Analyses were performed on protein and transcript levels.

Protein–protein associations within CD178 were queried using the STRING platform37. 176 CD178 genes matched with proteins in the STRING database (Fig. 4a). Notably, STRING network analysis separated 3 highly connected k-Means clusters within CD178 (A) with 10 proteins (A10); (B) with 8 proteins (B8); and (C) with 7 proteins (C7; total = ABC25) (Fig. 4a). To interpret the identified protein interactions we performed unguided as well as cluster-guided gene set enrichment analysis (GSEA) using a FDR < 0.0538.

Unguided GSEA correlated previously defined hallmark gene sets to CD178; with Epithelial Mesenchymal Transition (EMT) being the most significantly correlated hallmark gene set (FDR = 1.79 × 10−11), supporting the previously described role of EMT in ADPKD cystogenesis (Fig. 4b)16,39. Given the smaller size of ABC25 and, hence, smaller possibility space, cluster-guided GSEA was complemented with gene ontology (GO) analysis, transcription factor target analysis and evaluation of the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Fig. 4c)39–42. Interestingly, cluster A10 correlated mainly with EMT; and also with factors related to the Endoplasmic Reticulum Lumen as well as Post Translational Protein Modification. Cluster B8, instead comprised genes related to Epithelial Cell Differentiation, Animal Organ Morphogenesis and Sequence Specific DNA Binding, suggesting a role in tissue formation. And cluster C7 seemed exclusively related to phosphorylation-dependent cellular signaling (Fig. 4c). For full GSEA results, see Table S6.

The transcriptional regulation of CD178 was probed by transcription factor enrichment analysis using the ShinyGO and GeneSetDB meta-databases41,42. Seven transcription factors were identified to be significantly enriched in the regulation of CD178 (FDR < 0.05; Ar, Gbp6, Esr1, Ets2, Glis1, Tp53 and Rbpj; TF-), controlling at least 17 of the 178 genes in CD178 (Fig. 4d,e). Targeted expression analysis of TF- demonstrated for Esr1 a significant down-regulation in Pkd1−/− and Pkd2−/− cells (log2 fold change = −1.04, FDR = 0.003 and log2 fold change = −1.54, FDR = 1.78 × 10−6, respectively); none of the other 6 TF- factors showed consistent, Pkd1-and Pkd2-dependent differences in gene expression (Table S7). Because the PC-1/TRPP2 receptor-ion channel structure makes a direct control of cellular transcription unlikely, it is intriguing to speculate that non-transcriptional polycystin signaling regulates TF- activity, which in turn may control PKD-dependent gene expression in kidney tubular epithelial cells43. Ca2+-regulation and phosphorylation-dependence, for example, have been described for the zinc-finger protein GLI1 and the cellular tumor antigen p5344,45. TF- transcription factors may, hence, provide a mechanistic link from PC-1/TRPP2 channel activity to downstream regulation of cellular gene expression.

Another important contributor to both protein–protein associations and control of gene expression levels is alternative splicing of pre-mRNAs46. To date, alternative splicing has not been reported in ADPKD. Analysis of differential exon usage in our wild-type, Pkd1−/− and Pkd2−/− RNA-seq data identified 176 Pkd-responsive exons in 107 distinct genes (Fig. 5a,b). Within CD178, three genes (Pik3r1, Bex4 and Slc16a3) showed significant differences in exon usage (Fig. 5c). Similarly, TF- Rbpj exon usage seemed to be Pkd-dependent (Fig. 5d). Notably, Pkd1 exons 11, 15, 34 and 46 were differentially spliced in Pkd2-deficient cells (Fig. 5e,f). This indicated that functional TRPP2 is required for both, the amount and the splice variant of PKD1 gene expression (Figures S1c, 5e,f). Although direct derivation of biologically meaningful insights from these data is challenging, it reinforced the intimate mutual regulation of PKD genes and PC-1/TRPP2 proteins.

Furthermore, integration of protein association- and transcription factor-derived networks delineated putative Pkd-dependent regulatory cascades in mIMCD3 cells: 5 ABC25 genes (Cd34, Sp1, Cdi2, C3 and Mmp2) were linked to 6 of the TF- transcription factors, thus, further connecting phenotypic ADPKD pathogenesis to the unknown cellular functions of PKD1 and PKD2 (Fig. 4e).
Figure 3. Complementary mathematical analysis of Pkd1- and Pkd2-responsive transcripts identified a core ADPKD data set of 178 consistently regulated genes (CD178). To identify genes involved in ADPKD pathogenesis, two independent mathematical approaches were used: (1) statistical analysis of differential gene expression (DGE); and (2) unbiased data simplification by principal component analysis (PCA). (a) and (b) Applying RNA-seq on Pkd1- and Pkd2-deficient mIMCD3 cells revealed a total of 791 differentially expressed genes in Pkd1−/− and 1250 in Pkd2−/− cells (large, dark-yellow dots: FDR < 0.05, log2 fold change ≥|1|). In Pkd1−/− cells, > 54% (429 genes) and in Pkd2−/− cells, > 63% (790 genes) were downregulated. (c) Filtering DGE of Pkd1−/− and Pkd2−/− for concordant changes provided 254 genes (DGE_{254}), of which 38 were up- and 216 downregulated genes (orange dots). (d) 10% most variant genes were input for PCA (n = 2,506, with reads obtained for 25,062 genes). Principal component 1 (PC1) mainly accounted for non-Pkd-related differences between wildtype, Pkd1 and Pkd2 cells. PC2 discriminated Pkd1−/− and Pkd2−/− from wildtype, accounting for 29% of total variance. (e) Since PC2 correlated with Pkd1- and Pkd2-deficiency, genes were ranked according to their PC2 contribution, and the top 20% (n = 501, marked in blue) were selected for further data analysis (PCA_{501}). (f) Comparison of DGE_{254} and PCA_{501} identified 178 candidate genes specifically responding to Pkd1- and Pkd2-deficiency in mIMCD3 cells (core data set; CD178).
Figure 4. Function-based mapping of CD178 identified ADPKD-related protein networks. (a) CD178 was used as input for STRING network analysis. No annotations were available for 2610203C20Rik and 6330403L08Rik. Application of the k-means algorithm to CD178 revealed 3 distinct clusters: A with 10 proteins (A10); B with 8 proteins (B8); and C with 7 proteins (C7; total = ABC25). (b) and (c) to facilitate interpretation of the observed clusters, gene set enrichment analysis (GSEA) was applied to both CD178 and ABC25 (FDR < 0.05). ABC25 genes were subjected to an extended analysis, comprising hallmark gene sets, gene ontology (GO) analysis, transcription factor target analysis and evaluation of the Kyoto Encyclopedia of Genes and Genomes (KEGG). See Table S7 for full results. (d) To elucidate CD178 transcriptional regulation, ShinyGO and GeneSetDB meta-databases were queried (FDR < 0.05). This revealed a total of 17 CD178 genes regulated by 7 transcriptional factors (TFs). (e) 6 of 7 TFs were shown to form a k-means based cluster. Also, ABC25 identified several regulatory cascades: 5 ABC25 genes (Cd34, Spp1, Cdh2, C3 and Mmp2) were linked to 6 of the TFs (Ar, Ets, Gli1, Rbpj, Tp53). Spp1 and Cdh2 were regulated by more than one TF, whereas Ets2, Rbpj and Tp53 regulated more than one ABC25 gene.
Figure 5. Loss of Pkd1 and Pkd2 caused significant alteration in exon usage, including differential splicing of Pkd1. (a) Differential exon usage in Pkd1−/− and Pkd2−/− cells was assessed using DEXSeq. Surprisingly, the number of transcripts with differential exon usage compared to wildtype was >70-fold higher in Pkd2-deficient cells than in Pkd1−/− cells (n = 16,480 vs. n = 403). (b) 176 differentially used exons overlapped between both genotypes. Identified exons mapped to 107 annotated genes and 39 genomic loci without annotation. 107 characterized genes are thus differentially spliced in Pkd1−/− and Pkd2−/− cells. (c) Within the CD178 core set of genes with Pkd-dependent mRNA expression, Pik3r1, Bex4 and Slc16a3 showed additional Pkd-dependent alterations in exon usage. Interestingly, Pik3r1 was associated with protein phosphorylation in functional cluster C7 (Fig. 4c). (d) TF7 transcription factor Rbpj showed differential splicing in Pkd1−/− and Pkd2−/− cells. Rbpj activity had been linked to the expression of 4 CD178 genes: Cdh2, Ghul, Mt2 and Spp1 (Fig. 4e). (e) and (f) further corroborated the functional interdependence of PKD1 and PKD2, Pkd1 exons 11, 15, 34 and 46 showed reduced usage in Pkd2−/− cells. 4 domains of Polycystin-1 (PC-1) were affected: exons 11 and 15 project to the long extracellular N-terminus of PC-1, with exon 15 being part of the receptor of egg jelly module; exon 34 contributes to formation of the transmembrane domains; and exon 46 forms part of the intracellular C-terminus.
Combining published and newly generated data on PKD1 and PKD2 deficiency generated a cross-species transcriptional profile of ADPKD and identified 9 novel targets for future development. The genetic contrast of wild-type and specifically engineered mutant renal tubular epithelial cells provided the opportunity to assess the intricacies of Pkd dependent transcriptional regulation in simplified cellular isolation. To weigh the relative importance of newly identified CD178 pathways in the complex ADPKD in vivo pathophysiology, we next expanded our investigation to the in vitro in vivo correlation of CD178-derived pathways.

Rental expression of CD178 genes was analyzed by cell type. As expected by the mIMCD3 data source, all CD178 genes were consistently expressed along the nephron without particular cortical or juxtamedullary enrichment (Figure S4a–c). Within the nephron, parietal cells (> 60%), collecting duct (> 40%) and medullary pelvis epithelium (> 50%) show the highest expression of CD178 genes (Figure S4a–c).

To evaluate the relevance of the newly identified CD178 genes for ADPKD pathophysiology we performed a comprehensive cross-species meta-analysis of PKD-dependent RNA expression. At the time of analysis, data were available from six studies evaluating PKD1-dependent transcription and one evaluating loss of Pkd2: one in renal tubular cells; five in mutant mouse renal tissue; and one in cystic human kidneys (Table S8). To non-orthologous models of ADPKD were not included in the analysis. For internal validation, we further performed RNA-seq of whole kidneys from littermate control and adult-onset conditional Pkd1/2−/− mice (n = 7; Pkd1tm2Ggg/Pax8−/−; induction from post-natal day 28 to 42; RNA isolation in week 12) (Figure S5a–c). Pooling analysis of aggregated data showed that expression of 177 CD178 genes had been described previously as PKD-dependent (Figure S6a). Of those, 31 replicated in ≥ 3 independent transcriptional profiles (Rep31), with six genes (ALDH2, APNL, GALST1, Hs6ST2, PANK1, and PCSK9) showing concordant regulation across studies, substantiating their pathophysiological relevance in ADPKD (Figs. 6a,b, S6b,c). 61 Pkd1/2−/− dependent genes in CD178 on the other hand, had not been associated with ADPKD before (New61) (Fig. 6a,c).

To extend our mechanistic understanding of New61 in PC-1/TRPP2 signaling and identify novel targets for development in ADPKD, we probed the pathophysiological relevance of respective genes by applied bayesian inference. New61 genes were selected to show significant concordant changes (log2 fold change ≥ |1|) in Pkd1- and Pkd2-deficient mIMCD3 cells (Fig. 6c). (2) Analysis of renal cell type-specific mRNA profiles and comparison to Rep31 expression validated disease relevant tissue expression for 49 New61 genes (Figure S4a–c). Within the nephron, parietal cells (> 60%), collecting duct (> 40%) and medullary pelvis epithelium (> 50%) show the highest expression of CD178 genes (Figure S4a–c).

The potential biological relevance of CD178 in vivo was subsequently validated by meta-analysis of independent ADPKD transcriptional data, including new data from a Pkd1−/− mouse model (Fig. 6a–d). Putative confounding factors of this approach remain our incomplete understanding of the relevant ADPKD phenotype in vitro as well as missed or biased information in reference biobanks. Still, available data was sufficient to identify the Pkd1/2−/− TF7 > ABC25 hierarchical regulatory network, which may comprise core pathways amplifying...
Figure 6. Cross-species meta-analysis of independent ADPKD transcriptomic profiles. (a) 7 independent Pkd1−/− transcriptomic studies were chosen for a meta-analysis approach, comprising cell culture, mouse tissue as well as human kidney samples12–15. 6 studies had been previously published, 3 in another meta-analysis16. 1 mouse model (Pkd1−/− RNA-seq) was established for independent confirmation (Figure S5). 31 genes were identified in the core data set and at least 3 other studies. 61 genes are reported in the core data set only. (b) Of 31 confirmed genes Rep31, 2 showed concordant upregulation, 4 downregulation. Interestingly, 10 genes are concordantly regulated in cell culture models with discordant regulation in animal/human models. (c) Complementing published data sets, CD178 combined regulation by Pkd1−/− and Pkd2−/−. A total of 61 genes appeared in CD178 only (New61). (d) Combining the cross-species meta-analysis with function-based mapping and gene enrichment analysis of ABC25 highlighted 8 yet unreported and 6 confirmed genes. Both, confirmed and new genes were represented in clusters A10 and B8. Interestingly, none of the Rep31 but 3 of New61 genes were represented in functional cluster C7. Thus, combining Pkd1−/− and Pkd2−/− identified new targets for in-depth analysis of ADPKD pathogenesis.
transient polycystin-dependent ion flux into stable cellular differentiation and renal tubular structure (Figs. 4e, 6d). Derived from the largest volume of ADPKD mRNA expression data available, the PKD1/2 > TF; > ABCG2 signaling cascade—and in particular the 6 replicated and 8 newly identified genes—generated novel, testable hypotheses that may broaden our understanding of PKD-dependent tissue homeostasis in health and disease.

For example, sex has been described as significant determinant of ADPKD disease severity and disease.

Men with ADPKD often have worse renal cystic disease but less severe liver cystic disease. Supported by the male origin of mLmCD3 cells, PKD1/2 > AR > CDH2 signaling in cluster A9 may contribute to this phenotype (Figs. 4e, 6d and Table S9). The AR gene in TF-encodes the steroid-hormone activated androgen receptor, which regulates eukaryotic gene expression and affects cellular proliferation and differentiation in many target tissues. Consistent with a putative input from PC-1/TRPP2, AR transcription factor activity is modulated by bound coactivator and corepressor proteins as well as post-translational modifications. AR-regulated CDH2 encodes cadherin-2, a Cα-dependent cell adhesion protein that is known to be important for the establishment and maintenance of the multicellular structure of organs. Notably, renal proximal tubular cells are the only epithelial cells in the adult organism expressing cadherin-2 as major cell–cell adhesion protein. Up-regulation of CDH2 is a hallmark of EMT, which has been described extensively for cyst-lining cells (Figure S6d). Further investigation of PKD1/2 > AR > CDH2 may hence link several important disease-modulating factors in men with ADPKD.

Within the Epithelial Cell Differentiation cluster B6, PKD1/2 > ETS2 > CD34 may provide another interesting model to test (Figs. 4e, 6d). ETS2 encodes the epithelium-specific, winged helix-turn-helix E26 transformation-specific proto-oncogene 2 transcription factor (ETS2) that regulates diverse cellular activities, including apoptosis, cell growth, adhesion, the extracellular matrix and other transcription factors. ETS2 is a labile protein with a short half-life. ETS2 transcription factor activity is regulated by E3 ligase-dependent ubiquitination and subsequent proteasomal degradation. Interestingly, PC-1/TRPP2 has already been shown to regulate the activity of two E3 ligases SIAH-1 and c-CBL. Also, ETS2 is a downstream effector of the RAS/RAF/ERK pathway that promotes disease pathogenesis in ADPKD. CD34 is regulated by ETS2 and is a common marker for diverse progenitor cells with progressive downregulation on more mature cells. Functionally, CD34 is a sialomucin-type single-transmembrane glycoprophosphoprotein closely related to podocalyxin that seems to alter cellular adhesive properties. CD34 is furthermore intracellularly coupled to CRKL, which has been associated with congenital kidney anomalies in DiGeorge Syndrome, including microcystic tubules and glomeruli. The observed PKD-dependent over-expression of CD34 may therefore contribute to cystogenesis by causing cellular de-differentiation that interferes with renal tubular stability (Figure S6d).

In cluster C, we identified CSF2RA, PIK3R1 and PLCB1 as novel genes responsive to PKD1 and PKD2, but not enough biobank information was available to automatically develop transcription-based regulatory signaling models (Figs. 4e, 6d). However, functional links to ADPKD have been described previously for PIK3R1 and PLCB1. PIK3R1 encodes phosphoinositide-3-kinase regulatory subunit 1 (PIK3R1), the major suppressive regulatory subunit of phosphoinositide 3-kinase (PI3K). PI3K participates in a broad range of regulatory processes, including cell growth, proliferation, metabolism and secretion. In cancer, loss of PIK3R1 has been shown to promote EMT and cellular proliferation by over-activation of downstream AKT signaling. In ADPKD, increased PI3K/AKT signaling has been described as disease modifying, but both, loss and gain of PIK3R1 expression may reconcile these divergent findings, but additional experimental evidence is needed to clarify the molecular function of PIK3R1 in ADPKD. PLCB1 encodes phosphoinositide C beta 1 (PLCB1), which catalyzes the formation of inositol 1,4,5-trisphosphate (InsP3) and diacylglycerol (DAG) from phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P2). PLCB1 uses Ca2+ as a cofactor and plays an important role in the intracellular transduction of many extracellular signals. Goq-mediated activation of PLC has been described downstream of PC-1.

Our transcriptomic analyses yielded novel, testable hypotheses and highlighted our still rather limited understanding of cellular physiology. Publicly available biobank data allowed for the identification and cross-correlation of ABCG2, but were insufficient to develop unbiased hypotheses for most variations in PKD-dependent gene expression. To date, combination of sophisticated model systems and high-resolution methods yield volumes of relevant transcriptomic data. Yet, interpretation has proven challenging when it comes to deriving biological meaning, as most types of systemic analysis, such as over-representation analysis, GSEA and signaling pathway impact analysis, rely to some degree on a priori knowledge of the pathways, the biological role, or the molecular function of genes. As such, annotations have been instrumental for understanding molecular signatures of many diseases. But annotation-based approaches have significant drawbacks that limit the insights that can be gleaned from respective analyses. For instance, there is a strong underlying assumption that orthologous genes share similar biological functions. This may overemphasize highly conserved cellular processes and potentially overlook important species-specific and/or tissue-specific functions. Similar network effects cause well-studied genes to continuously accrue data, while the non-randomness of missing annotations is more likely to disadvantage prime specific genes or new targets without a wealth of literature. With more data becoming publicly available, future in silico re-analyses of CD25 have therefore the intriguing potential to further clarify core functional connections currently still hidden in the ADPKD transcriptomic profile.

Taken together, our data indicate the robust power of integrative transcriptomics to select targets for subsequent functional studies in ADPKD, including the 6 replicated (C3, FSTL1, PCOLCE, PCSK9, SPP1 and ZFP37) and 8 newly identified (CD34, CDH2, CSF2RA, DLX5, HOXC9, PIK3R1, PLCB1 and TLR6) genes (Fig. 6d).
Methods

Contact for reagent and resource sharing. Further information and requests for resources should be directed to the Lead Contacts, Michael Köttgen (michael.koettgen@uniklinik-freiburg.de) and Alexis Hofherr (alexis.hofherr@uniklinik-freiburg.de).

Key resources. Details on key resources are listed in the supplementary information Key Resources Table.

Experimental model details. Cell culture. Murine inner-medullary collecting duct cells (mIMCD3) were isolated from healthy mice expressing the early region of simian virus SV40 (Tg(SV40E)Bri/7) to sustain continuous proliferation in vitro.22–25. Wild-type mIMCD3 cells are diploid, consistently differentiate and replicate collecting duct-like transport activity22–25. Pkd1−/− and Pkd2−/− mIMCD3 cells have been described previously28. Independent deletion of Pkd1 (CRISPR) and Pkd2 (TALEN) was confirmed by DNA sequencing and validated by RNA-Seq: mIMCD3 Pkd1−/− (Δ-chr. 17: 24,550,055–24,594,963/Δ-chr. 17: 24,550,106–24,594,903); mIMCD3 Pkd2Δc1/2 (Δ-chr. 5: 104,478,281–104,490,878)29. RNA-seq read mapping and de novo transcriptome assembly did not indicate significant genome-editing induced off-target effects in Pkd1/2 cells. Flow cytometry-based DNA content measurement showed normal ploidy in all cell lines. Loss of respective PC-1 and TRPP2 expression was demonstrated by Western blot. Overall viability, proliferation and differentiation, including ciliation, basolateral formation and transepithelial electrical resistance, were similar in wild-type and mutant cells29.

Cells were cultured in a mix of Dulbecco’s modified eagle’s medium (DMEM-F12; Lonza), which was supplemented with 10% fetal bovine serum and 1% Penicillin–Streptomycin (10,000 U Penicillin and 10 mg/ml Streptomycin; both Sigma Aldrich). Cells were grown at 37 °C and 10% CO2. PCR-based testing for mycoplasma contamination was performed regularly. New ampules were started after ~25 passages. Cells were harvested using 0.25% Trypsin–EDTA (Gibco). Cell counts were performed by Countess II cell counter (Thermo Fisher).

Method details. Protein isolation, SDS-PAGE, Western blot and ECL detection. Cells were grown to epithelial confluency (5 days). Cells were lysed, proteins were isolated, processed in SDS-PAGE and Western blot and detected by electrochemiluminescence as described previously101. For TRP2-immunoprecipitation, mouse anti-TRPP2 antibody was used (Santa Cruz Biotechnologie Inc., Dallas, USA). Beta-actin was targeted using anti-b-actin antibody (Sigma Aldrich, St. Louis, USA). Luminescence was detected by a 16-bit ChemoCam system (Intas).

Immunofluorescence. Indirect immunofluorescence staining of cells has been described previously102. Briefly, after reaching full confluency cells were permeabilized using Triton X-100 (Carl Roth GmbH, Karlsruhe, Germany) and fixed by parafomaldehyde (Electron Microscopy Sciences, Hatfield, PA, USA). Cells were stained by primary (1:50 anti-ZO1 rat, Santa Cruz Biotechnology, Dallas, USA) and secondary antibodies (1:1000 Anti-rat IgG, Cy3, Jackson Lab. Inc., Waltham, USA; anti-rabbit IgG, Cy3, Jackson Lab. Inc., Waltham, USA in PBS. DAPI (Sigma-Aldrich) was added in a dilution of 1:1000 to visualize DNA. Bright-field images were recorded using an Axio Vert Observer microscope Z1 (Zeiss).

RNA isolation. For wildtype/Pkd comparison, six 10 cm cell culture dishes were grown per genotype to full confluency and ciliation, starting with 5 × 10⁶ cells per dish. RNA Isolation was performed 4 days after full confluency according to manufacturer’s protocol using the RNeasy Plus Mini Kit (Qiagen), harvesting 2 × 10⁶ cells per dish. Purification of RNA samples was performed by solvent extraction with 100% ethanol followed by precipitation with ammonium acetate. 10 µl glycogen was added to 100 µl RNA solution. RNA was solved in 10 µl water.

To minimize false discovery, mIMCD3-based transcriptomic profiling of ADPKD was hence restricted to continuous proliferation in vitro.22. Wild-type mIMCD3 cells are diploid, consistently differentiate and replicate collecting duct-like transport activity.22–25. Pkd1−/− and Pkd2−/− mIMCD3 cells have been described previously.28. Independent deletion of Pkd1 (CRISPR) and Pkd2 (TALEN) was confirmed by DNA sequencing and validated by RNA-Seq: mIMCD3 Pkd1−/− (Δ-chr. 17: 24,550,055–24,594,963/Δ-chr. 17: 24,550,106–24,594,903); mIMCD3 Pkd2Δc1/2 (Δ-chr. 5: 104,478,281–104,490,878)29. RNA-seq read mapping and de novo transcriptome assembly did not indicate significant genome-editing induced off-target effects in Pkd1/2 cells. Flow cytometry-based DNA content measurement showed normal ploidy in all cell lines. Loss of respective PC-1 and TRPP2 expression was demonstrated by Western blot. Overall viability, proliferation and differentiation, including ciliation, basolateral formation and transepithelial electrical resistance, were similar in wild-type and mutant cells.

Cells were cultured in a mix of Dulbecco's modified eagle's medium (DMEM-F12; Lonza), which was supplemented with 10% fetal bovine serum and 1% Penicillin–Streptomycin (10,000 U Penicillin and 10 mg/ml Streptomycin; both Sigma Aldrich). Cells were grown at 37 °C and 10% CO2. PCR-based testing for mycoplasma contamination was performed regularly. New ampules were started after ~25 passages. Cells were harvested using 0.25% Trypsin–EDTA (Gibco). Cell counts were performed by Countess II cell counter (Thermo Fisher).

Method details. Protein isolation, SDS-PAGE, Western blot and ECL detection. Cells were grown to epithelial confluency (5 days). Cells were lysed, proteins were isolated, processed in SDS-PAGE and Western blot and detected by electrochemiluminescence as described previously.101. For TRP2-immunoprecipitation, mouse anti-TRPP2 antibody was used (Santa Cruz Biotechnologie Inc., Dallas, USA). Beta-actin was targeted using anti-beta-actin antibody (Sigma Aldrich, St. Louis, USA). Luminescence was detected by a 16-bit ChemoCam system (Intas).

Immunofluorescence. Indirect immunofluorescence staining of cells has been described previously.102. Briefly, after reaching full confluency cells were permeabilized using Triton X-100 (Carl Roth GmbH, Karlsruhe, Germany) and fixed by paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA, USA). Cells were stained by primary (1:50 anti-ZO1 rat, Santa Cruz Biotechnology, Dallas, USA) and secondary antibodies (1:1000 Anti-rat IgG, Cy3, Jackson Lab. Inc., Waltham, USA; anti-rabbit IgG, Cy3, Jackson Lab. Inc., Waltham, USA) in PBS. DAPI (Sigma-Aldrich) was added in a dilution of 1:1000 to visualize DNA. Bright-field images were recorded using an Axio Vert Observer microscope Z1 (Zeiss).

RNA isolation. For wildtype/Pkd comparison, six 10 cm cell culture dishes were grown per genotype to full confluency and ciliation, starting with 5 × 10⁶ cells per dish. RNA Isolation was performed 4 days after full confluency according to manufacturer's protocol using the RNeasy Plus Mini Kit (Qiagen), harvesting 2 × 10⁶ cells per dish. Purification of RNA samples was performed by solvent extraction with 100% ethanol followed by precipitation with ammonium acetate. 10 µl glycogen was added to 100 µl RNA solution. RNA was solved in 10 µl water.

DNA sequencing (RNA-seq) analysis. RNA-seq was performed in samples from wild-type as well as from one Pkd1−/− and one Pkd2−/− mIMCD3 clone. Despite multiple sampling, untargeted analysis of such clonal cell lines may be confounded by undetected, random genome editing-induced off-target effects or clonal artifacts. To minimize false discovery, mIMCD3-based transcriptomic profiling of ADPKD was hence restricted to consistent changes across the two independent Pkd genotypes. Data was analyzed with Galaxy Europe (https://usegalaxy.eu/). Reads were first trimmed and quality checked with Trim Galore (Galaxy Tool Version 0.4.3.1) and FastQC (Galaxy Tool Version 0.72), respectively. Reads were aligned to the mouse genome version mm10 (GRCm38) applying HISAT2 (Galaxy Tool Version 2.1.0 + galaxy3)104. Mapped reads were counted with htseq-count (Galaxy Tool Version 0.9.1)104. Differential gene expression (DGE) was tested using DESeq2 (Galaxy Tool Version 2.11.40.4) using a cutoff of log2 fold change of ≥ 1 for upregulated and ≤ −1 for downregulated genes.35,106 False discovery rate of DGE was controlled with a p value < 0.05 adjusted for multiple testing (Benjamini–Hochberg procedure). Principal Component Analysis (PCA) was performed using DESeq2 in R.106 Of 25'061 genes with reads obtained in the RNA-seq analysis, 10% most variant genes were selected as input for PCA (n = 2506). Data was visualized using the R packages: pcaExplorer and pheatmap.107
Alternative splicing analysis. The reads processed in the RNA-seq analysis were also aligned to the reference genome and tailored for StringTie tool with HSAT2 (Galaxy Tool Version 2.1.0 + galaxy3)39,41. De novo transcript reconstruction was conducted with StringTie (Galaxy Tool Version 1.3.6)38,42. The transcriptome was assembled using StringTie-merge (Galaxy Tool Version 1.3.6). Assembled transcripts were compared to the reference annotation (Mus musculus, mm9) with GFFCompare (Galaxy Tool Version 0.11.2)109. Exon abundancies were then estimated using DEXSeq-Count (Galaxy Tool Version 1.28.1)99,100. Differential exon usage (DEU) was tested using standard hypergeometric distribution with correction for multiple hypotheses testing using FDR < 0.05 

Kidney cell explorer analysis. Kidney Cell Explorer Analysis was performed using47 https://cello.shinyapps.io/kidneyCellExplorer/. "Adult mouse kidney" was selected as reference data set. CD178 were used as input by pasting the respective gene IDs in the search field. Expression was displayed in "Average expression (rescaled)" mode and "RdYIBu" Palette. Genes were counted as "expressed" if rescaled average expression was > 0, e.g. if color was lighter than dark blue (HEX code #4575b4ff).

Nephro cell analysis. Nephro cell analysis was performed using http://nephrocell.mikmtmc.org/. "Adult normal kidney" was selected as reference data set. Rep31 and New61 were used as input; one gene at a time. Violin plots displayed were analyzed and divided into 5 categories: 0 = not measured in the nephrocell data set; 1 = no expression; 2 = low expression, no violin plots displayed; 3 = high expression, violin plots displayed; 4 = very high expression, striking violin plots displayed. A heatmap was generated to visualize these values.

Transcription factor enrichment analysis. Testing for transcription factors regulating the core data set genes was performed with ShinyGO (Tool Version 0.51) using the function TF.Target.TFacts41,42. False discovery rate was set to < 0.05.

Pathway analysis. Functional protein network and transcriptional factor network were determined using the STRING (Search tool for the retrieval of interacting genes/proteins) database version 11.037 (https://version-11.0.string-db.org/).

Functional enrichment analysis was performed against the Molecular Signature Database (MSigDB) v7.1 using standard hypergeometric distribution with correction for multiple hypotheses testing using FDR < 0.05 according to Benjamini and Hochberg38,42. Pathway databases used: Hallmark Gene Sets, Positional Gene Sets, GO Gene Sets, KEGG, All Transcription Factor Targets.

Cross-species meta-analysis. Available PKD1-responsive transcriptional data was retrieved from a literature research covering publications between 2008 and 201812–17. Transgenic models were not considered for comparison. Studies were included if they provided lists of differentially expressed genes and raw data had been deposited for public access. No PKD2 data matching above criteria were found. Ref18 was included for PKD2-responsive expression. Lists of differentially expressed genes were obtained from the primary or supplementary data of the respective publications. Analysis of13,16 was based on the previously described PKD1 Signature by15. Criteria for differential gene expression were adopted from the respective studies.

RNA-seq and profiling of Pkd1tm2Ggg/Pax8rtTA mice. Pkd1tm2Ggg/Pax8rtTA (C57BL/6 genetic background) mice have been described previously108. All animal experiments were conducted according to the guidelines of the National Institutes of Health Guide for the Care and Use of Laboratory Animals, as well as the German law for the welfare of animals, and were approved by regional authorities (Regierungspräsidium Freiburg G-13/18, G-15/58, and G-16/28). Mice were housed in a specific pathogen-free facility, fed ad libitum, and housed at constant ambient temperature in a 12-hour day/night cycle. Breeding and genotyping were done according to standard procedures. Pkd1 deletion was induced from post-natal day 28–42. Mice received doxycycline hydrochloride (Fagron) via the drinking water (2 mg/mL with 5% sucrose, protected from light) to induce Pkd1 deletion. Littersmates (lacking either TetO-Cre or Pax8rtTA) were used as controls. 2 male control animals and 5 male Pkd1−/− animals were sacrificed at a median age of 11.8 weeks. RNA was extracted from kidneys with RNeasy Plus Mini Kit (Qiagen) according to the manufacturer’s protocol as a one-step elution with 50 μL RNAse free water per sample. Quantity and quality of RNA was checked by NanoDrop One spectrophotometer (ThermoFisher). RNA-seq was performed by GATC Biotech AG (Konstanz, D) on Illumina HiSeq Genome Sequencer. Data analysis was performed using the same work-flow as for CD178 (see "RNA Sequencing analysis"). Differentially expressed genes (FDR < 0.05) were overlapped to top 20% genes contributing to principal component 2 in the principal component analysis. Overlapping genes were summarized in the Pkd1−/− mouse data set.

Quantification and statistical analysis. Statistical analyses of RNA-seq data were carried out in Galaxy and R using DESeq239,106. False discovery rate of DGE was set to < 0.05 (Benjamini–Hochberg procedure). The number of biological replicates (n), statistical test performed and resulting significance for each experiment is
noted in the figure legend or the corresponding results section of this document. Where applicable, original data for plots is available in supplementary tables. Additional data is available upon request.

Data availability
The raw and processed data from our analyses of RNA-seq will be available in the gene expression omnibus database under the accession number: GSE179947.

Received: 17 March 2021; Accepted: 8 July 2021
Published online: 23 July 2021

References
1. Cornec-Le Gall, E., Alam, A. & Perrone, R. D. Autosomal dominant polycystic kidney disease. Lancet 393(10174), 919–935 (2019).
2. Mochizuki, T. et al. PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. Science 272(5266), 1339–1342 (1996).
3. Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. The International Polycystic Kidney Disease Consortium. Cell 82(1), 289–298 (1995).
4. Shillingford, J. M. et al. The mTOR pathway is regulated by polycystin-1, and it's inhibition reverses renal cystogenesis in polycystic kidney disease. Proc. Natl. Acad. Sci. USA 103(14), 5466–5471 (2006).
5. Kim, S. et al. The polycystin complex mediates Wnt/Ca(2+) signalling. Nat. Cell Biol. 18(7), 752–764 (2016).
6. Happe, H. et al. Altered Hippo signalling in polycystic kidney disease. J. Pathol. 224(1), 133–142 (2011).
7. Bhunia, A. K. et al. PKD1 induces p21(waf1) and regulation of the cell cycle via direct activation of the JAK-STAT signaling pathway in a process requiring PKD2. Cell 109(2), 157–168 (2002).
8. Hofherr, A. et al. The mitochondrial transporter SLC25A25 links ciliary TRPP2 signaling and cellular metabolism. PLoS Biol. 16(8), e2005651 (2018).
9. Antignac, C. et al. The future of polycystic kidney disease research–as seen by the 12 Kaplan Awardees. J. Am. Soc. Nephrol. 26(9), 2081–2095 (2015).
10. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10(1), 57–63 (2009).
11. Kim, H. & Kim, J. S. A guide to genome engineering with programmable nucleases. Nat. Rev. Genet. 15(5), 321–334 (2014).
12. Kunnen, S. J. et al. Comparative transcriptomics of shear stress treated Pkd1(−/−) cells and pre-cystic kidneys reveals pathways involved in early polycystic kidney disease. Biomed. Pharmacother. 108, 1123–1134 (2018).
13. Menezes, L. F. et al. Network analysis of a Pkd1-mouse model of autosomal dominant polycystic kidney disease identifies HNF4alpha as a disease modifier. PLoS Genet. 8(11), e1003053 (2012).
14. Chen, W. C., Tseng, Y. S. & Li, H. Gene expression in early and progression phases of autosomal dominant polycystic kidney disease. BMC Res. Notes. 1, 131 (2008).
15. Malas, T. B. et al. Meta-analysis of polycystic kidney disease expression profiles defines strong involvement of injury repair processes. Am. J. Physiol. Renal. Physiol. 312(4), F806–F817 (2017).
16. Song, X. et al. Systems biology of autosomal dominant polycystic kidney disease (ADPKD): computational identification of gene expression pathways and integrated regulatory networks. Hum. Mol. Genet. 18(13), 2328–2343 (2009).
17. Pandey, P., Qin, S., Ho, J., Zhou, J. & Kreidberg, J. A. Systems biology approach to identify transcriptome reprogramming and candidate microRNA targets during the progression of polycystic kidney disease. BMC Syst. Biol. 5, 56 (2011).
18. Zhang, C. et al. Cyclin-dependent kinase 1 activity is a driver of cyst growth in polycystic kidney disease. J. Am. Soc. Nephrol. 32(1), 41–51 (2021).
19. Terabayashi, T., Germino, G. G. & Menezes, L. F. Pathway identification through transcriptome analysis. Cell Signal. 74, 109701 (2020).
20. Qian, F., Watnick, T. J., Onuchic, L. F. & Germino, G. G. The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I. Cell 187(6), 979–987 (1996).
21. Formica, C. & Peters, D. J. M. Molecular pathways involved in injury-repair and ADPKD progression. Cell Signal. 72, 106484 (2020).
22. Rauchman, M. I., Nigam, S. K., Delpire, E. & Gullans, S. R. An osmotically tolerant inner medullary collecting duct cell line: expression, functional localization, and differential regulation. J. Am. Physiol. Renal. Physiol. 279(7 Pt 2), F416-424 (1993).
23. Kaplan, M. R., Plotkin, M. D., Brown, D., Hebert, S. C. & Delpire, E. Expression of the mouse Na-K-2Cl cotransporter, mBSC2, in the terminal inner medullary collecting duct, the glomerular and extraglomerular mesangium, and the glomerular afferent arteriole. J. Clin. Invest. 98(3), 723–730 (1996).
24. Amlal, H., Goel, A. & Soleimani, M. Activation of H+–ATPase by hypotonicity: a novel regulatory mechanism for H+ secretion in IMCD cells. Am. J. Physiol. 275(4), F487–501 (1998).
25. Soleimani, M., Singh, G., Bizal, G. L., Gullans, S. R. & McAteer, J. A. Na+/H+ exchanger isoforms NHE-2 and NHE-1 in inner medullary collecting duct cells: expression, functional localization, and differential regulation. J. Biol. Chem. 269(45), 27973–27978 (1994).
26. Aboudeh, K. et al. Long noncoding RNA Hoxb3os is dysregulated in autosomal dominant polycystic kidney disease and regulates mTOR signaling. J. Biol. Chem. 293(24), 9388–9398 (2018).
27. Radley, A. H. et al. Assessment of engineered cells using CellNet and RNA-seq. Nat. Protoc. 12(5), 1089–1102 (2017).
28. Hofherr, A. et al. Efficient genome editing of differentiated renal epithelial cells. Pflugers Arch. 469(2), 303–311 (2017).
29. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. Science 297(5584), 1183–1186 (2002).
30. Churchill, G. A. Fundamentals of experimental design for cDNA microarrays. Nat. Genet. 32(Suppl), 490–495 (2002).
31. Hansen, K. D., Wu, Z., Irizarry, R. A. & Leek, J. T. Sequencing technology does not eliminate biological variability. Nat. Biotechnol. 29(7), 572–573 (2011).
32. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 18(9), 1509–1517 (2008).
33. Kim, S., Kang, D., Huo, Z., Park, Y. & Tseng, G. C. Meta-analysis of polycystic kidney disease gene expression pathways and integrated regulatory networks. BMC Syst. Biol. 3(13), 289–300 (2009).
34. Costa-Silva, J., Domingues, D. & Lopes, P. F. RNA-Seq differential expression analysis: an extended review and a software tool. PLoS ONE 12(12), e0190152 (2017).
35. Consortium M et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat. Biotechnol. 24(9), 1151–1161 (2006).
36. Conesa, A. et al. A survey of best practices for RNA-seq data analysis. Genome Biol. 17, 13 (2016).
81. Felschow, D. M., McVeigh, M. L., Hoehn, G. T., Civin, C. I. & Fackler, M. J. The adapter protein CrkL associates with CD34. *Blood* 97(12), 3768–3775 (2001).

82. Fruman, D. A. et al. The PI3K pathway in human disease. *Cell* 170(4), 605–635 (2017).

83. Lin, Y. et al. PI3K/R1 negatively regulates the epithelial-mesenchymal transition and stem-like phenotype of renal cancer cells through the AKT/GSK3β/CITNBB1 signaling pathway. *Sci. Rep.* 5, 8997 (2015).

84. Boca, M. et al. Polycystin-1 induces resistance to apoptosis through the phosphatidylinositol 3-kinase/Akt signaling pathway. *J. Am. Soc. Nephrol.* 17(3), 637–647 (2006).

85. Gargaloni, A. N. et al. Polycystin-1 induces activation of the PI3K/AKT/mTOR pathway and promotes angiogenesis in renal cell carcinoma. *Cancer Lett.* 489, 135–143 (2020).

86. Boca, M. et al. Polycystin-1 induces cell migration by regulating phosphatidylinositol 3-kinase-dependent cytoskeletal rearrangements and GSK3β-dependent cell cell mechanical adhesion. *Mol. Biol. Cell.* 18(10), 4050–4061 (2007).

87. Parker, E. et al. Hyperpolarization of PKD1 cystic cells is induced by insulin-like growth factor-1 activation of the Ras/Raf signaling system. *Kidney Int.* 72(2), 157–165 (2007).

88. Margaria, J. P., Campa, C. C., De Santis, M. C., Hirsch, E. & Franco, I. The PI3K/Akt/mTOR pathway in polycystic kidney disease: a complex interaction with polycystins and primary cilia. *Cell Signal.* 66, 109468 (2020).

89. Magyar, T. A. et al. Global microRNA profiling in human urinary exosomes reveals novel disease biomarkers and cellular pathways for autosomal dominant polycystic kidney disease. *Kidney Int.* 98(2), 420–435 (2020).

90. Bill, C. A. & Vines, C. M. Phospholipase C. *Adv. Exp. Med. Biol.* 1131, 215–242 (2020).

91. Puric, S. et al. Polycystin-1 activates the calcineurin/NFAT (nuclear factor of activated T-cells) signaling pathway. *J. Biol. Chem.* 279(53), 55455–55464 (2004).

92. Tarca, A. L. et al. A novel signaling pathway impact analysis. *Bioinformatics* 25(1), 75–82 (2009).

93. Gabaldon, T. & Koonin, E. V. Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.* 14(5), 360–366 (2013).

94. Maertens, A. et al. Functionally enigmatic genes in cancer: using TCGA data to map the limitations of annotations. *Sci. Rep.* 10(1), 4106 (2020).

95. Pandey, A. K., Lu, L., Wang, X., Homayouni, R. & Williams, R. W. Functionally enigmatic genes: a case study of the brain ignorance. *PLoS ONE* 9(2), e88889 (2014).

96. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* 25(1), 25–29 (2000).

97. Gene, O. C. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* 45(1), D100–D105 (2019).

98. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* 22(10), 2008–2017 (2012).

99. Afgan, E. et al. Drift and conservation of differential exon usage across tissues in primate species. *Proc. Natl. Acad. Sci. USA.* 110(38), 15377–15382 (2013).

100. Hofherr, A., Wagner, C. J., Watanabe, T. & Kottgen, M. Targeted rescue of a polycystic kidney disease mutation by lysosomal inhibition. *Kidney Int.* 89(4), 949–955 (2016).

101. Hofherr, A., Wagner, C., Fedele, S., Sonllo, S. & Kottgen, M. N-glycosylation determines the abundance of the transient receptor potential channel TRPP2. *J. Biol. Chem.* 289(21), 14854–14867 (2014).

102. Afgan, E. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 47(D1), D100–D105 (2019).

103. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* 22(10), 2008–2017 (2012).

104. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with Salmon. *BMC Genomics* 15(12), 550 (2014).

105. Pertea, G. & Pertea, M. GFF utilities: GffRead and GffCompare. *F1000Res* 9, 215–242 (2020).

106. Viau, A. et al. Cilia-localized LKB1 regulates chemokine signaling, macrophage recruitment, and tissue homeostasis in the kidney. *EMBO J.* 37, 18 (2018).

Acknowledgements

We would like to thank Simone Diederichsen and Lukas Peintner for technical assistance and Anna Reznichenko for her input on the manuscript. The following illustrations for figure design were obtained from the DBCLS Togo TV picture gallery: mouse (Figures 1, 6, S5, S6); human (Figure 1, 6, S6); and cell culture (Figure 1, 6, S6).

Author contributions

A.H., M.K. and E.W.K. designed the research project. S.F., H.M., C.R., H.S., K.F., C.I.W., T.B. and A.V. and A.H. performed the experiments and acquired the data. A.H., S.F., H.M. and C.R. analyzed the data. S.F., H.M., C.R., M.K. and A.H. drafted and revised the manuscript. All authors approved the final version of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. M.K. is supported by the German Research Foundation (DFG): TRR 152 (project ID 239283807); SFB1453 (Project ID 431984000) and Germany’s Excellence Strategy (CIBSS—EXC-2189—Project ID 390939984). E.W.K. is supported by the German Research Foundation (DFG): KU1504/7-1 and KU1504/8-1.

Competing interests

The authors declare no competing interests.
Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-94442-8.

Correspondence and requests for materials should be addressed to M.K. or A.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021