Combination of mouse models and genomewide association studies highlights novel genes associated with human kidney function

Jiaojiao Jing[1,2,3], Cristian Pattaro[4], Anselm Hoppmann[1,2], Yukinori Okada[5,6]; CKDGen Consortium, Caroline S. Fox[7,8,9] and Anna Köttgen[1,3,10]

[1]Division of Genetic Epidemiology, Medical Center, University of Freiburg, Freiburg, Germany; [2]Genetics and Experimental Bioinformatics, Faculty of Biology, University of Freiburg, Freiburg, Germany; [3]Renal Division, Medical Center, University of Freiburg, Freiburg, Germany; [4]Center for Biomedicine, European Academy of Bolzano/Bozen (EURAC), Bolzano, Italy; [5]Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan; [6]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; [7]Framingham Heart Study, National Heart, Lung, and Blood Institute, Framingham, Massachusetts, USA; [8]Center for Population Studies, National Heart, Lung, and Blood Institute, Framingham, Massachusetts, USA; [9]Division of Endocrinology, Metabolism, and Hypertension, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; and [10]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

Genomewide association studies have identified numerous chronic kidney disease–associated genetic variants, but often do not pinpoint causal genes. This limitation was addressed by combining Mouse Genome Informatics with human genomewide association studies of kidney function. Genes for which mouse models showed abnormal renal physiology, morphology, glomerular filtration rate (GFR), or urinary albumin-to-creatinine ratio were identified from Mouse Genome Informatics. The corresponding human orthologs were then evaluated for GFR-associated single-nucleotide polymorphisms in 133,814 individuals and urinary albumin-to-creatinine ratio-associated SNPs in 54,451 individuals in genome-wide association studies meta-analysis of the CKDGen Consortium. After multiple testing corrections, significant associations with estimated GFR in humans were identified for single-nucleotide polymorphisms in 2, 7, and 17 genes causing abnormal GFR, abnormal physiology, and abnormal morphology in mice, respectively. Genes identified for abnormal kidney morphology showed significant enrichment for estimated GFR–associated single-nucleotide polymorphisms. In total, 19 genes contained variants associated with estimated GFR or the urinary albumin-to-creatinine ratio of which 16 mapped into previously reported genomewide significant loci. *CYP26A1* and *BMP4* emerged as novel signals subsequently validated in a large, independent study. An additional gene, *CYP24A1*, was discovered after conditioning on a published nearby association signal. Thus, our novel approach to combine comprehensive mouse phenotype information with human genomewide association studies data resulted in the identification of candidate genes for kidney disease pathogenesis.

Correspondence: *Anna Köttgen, Medical Center, University of Freiburg, Berliner Allee 29, 79110 Freiburg, Germany. E-mail: anna.koettgen@uniklinik-freiburg.de*

Received 21 January 2016; revised 20 March 2016; accepted 7 April 2016

Kidney International (2016) ■, ■–■; http://dx.doi.org/10.1016/j.kint.2016.04.004

KEYWORDS: chronic kidney disease; genomewide association studies; kidney development

Chronic kidney disease (CKD) is characterized by the sustained presence of abnormalities of kidney structure or kidney function, quantified by the estimated glomerular filtration rate (eGFR) and the urinary albumin-to-creatinine ratio (UACR).[1] CKD can progress to kidney failure and increases the risk of cardiovascular events and mortality.[2–4] CKD has a genetic component, and genomewide association studies (GWASs) of CKD and kidney function measures have identified numerous associated genetic loci.[5–11]

The unbiased identification of novel pathogenic mechanisms is the main promise of GWASs,[12] and an increasing number of experimental follow-up studies are successfully translating observed associations into an improved understanding of underlying mechanisms. However, initial discovery GWASs are limited in their ability to pinpoint causal genes and do not provide direct insights into underlying molecular mechanisms. Moreover, the stringent significance threshold necessary to account for multiple testing leads to many false-negative associations, preventing the discovery of additional loci that may provide important pathophysiological insights.

One way to address these limitations is by integration of systematic evidence from experimental animal models of the same or a closely related phenotype. Assuming conserved gene function, evidence from animal models can implicate associated loci that do not meet genomewide significance. The laboratory mouse has long been an indispensable tool to model human traits and diseases because of the high degree of homology and the possibility to precisely manipulate the

mouse genome.[13] The Mouse Genome Informatics (MGI) resource[14] is the primary database on experimental mouse models. Phenotypic information is curated and coded using the Mammalian Phenotype Ontology[15] and is available along with spontaneous, induced, and genetically engineered mutations and other knowledge through multiple avenues of MGI.[16] Here, we aimed to assess whether human orthologs of genes that, when dysfunctional, cause kidney phenotypes in mice, are enriched for variants associated with corresponding traits in humans, and can be used to prioritize association signals that do not achieve genomewide significance for further investigation.

RESULTS

The study design is outlined in Figure 1 and described in detail in the Methods section. Using the comprehensive MGI resource, we obtained the following 4 lists of genes that, when knocked out or mutated in mice, give rise to the renal phenotypes and for which a human ortholog could be identified: the list for the phenotype abnormal GFR contained 23 genes, abnormal kidney physiology contained 341 genes, abnormal kidney morphology contained 660 genes, and abnormal urine protein contained 139 genes (Figure 2, with all genes listed in Supplementary Table S1). We then investigated whether the human orthologs of these genes contained genetic variants associated with an eGFR below the listwide significance threshold (for genes on the GFR, physiology, and morphology lists) or UACR (for genes on the UACR, physiology, and morphology lists) in data from GWAS meta-analyses of the CKDGen Consortium,[11] after correcting for the number of independent single-nucleotide polymorphisms (SNPs) across the evaluated genes. The CKDGen datasets combined information from up to 133,814 individuals of European ancestry for eGFR and up to 54,451 individuals for UACR.

As shown in Table 1,[17,18] we identified multiple genes harboring genetic variants associated with the respective kidney function trait after correcting for multiple testing. Association P values for SNPs in these genes surpassed the list-specific statistical significance threshold (see Methods). Significant associations with eGFR were identified for SNPs in 2 genes from the GFR list, 7 genes from the physiology list, and 17 from the morphology list. Several of these genes were represented on more than 1 list (Table 1). Significant associations with UACR were identified for SNPs in 1 gene, *CUBN*, which was found on both the UACR and the physiology list and in no gene from the morphology list. The identified genes presented in Table 1 did not change in sensitivity analyses that corrected for the independent SNPs among the unique genes across all evaluated lists for a given human phenotype. Additional genes that contained SNPs with suggestive associations ($P < 5\times10^{-5}$) with kidney function traits are shown in Table 2.

To evaluate whether such results were likely to arise due to chance, we generated for each of the 4 lists 2000 random gene lists of equal length as the original gene list. For each randomly generated list, we queried the CKDGen data and counted the number of genes containing associated SNPs after applying the appropriate multiple testing correction (see Methods). Figure 3 shows the distributions of the number of genes containing SNPs associated with the eGFR below the listwide significance threshold from the random draws. The vertical dashed lines represent the observed number of significant genes for the respective mouse renal phenotype. Observing SNPs associated with eGFR in humans in 17 genes, as we did for the abnormal morphology list, is highly unlikely by chance alone as it was never observed for the random lists (P-enrichment < 0.0005 or $<1/2000$)) (Figure 3a). For the genes on the abnormal physiology list (Figure 3c) and the GFR list (Figure 3c), the number of observed genes with significant associations fell into the upper part of the distribution expected by chance alone, but there was no significant enrichment. Similar observations were made for the observed number of genes associated with UACR in the smaller UACR dataset from the CKDGen Consortium (Supplementary Figure S1).

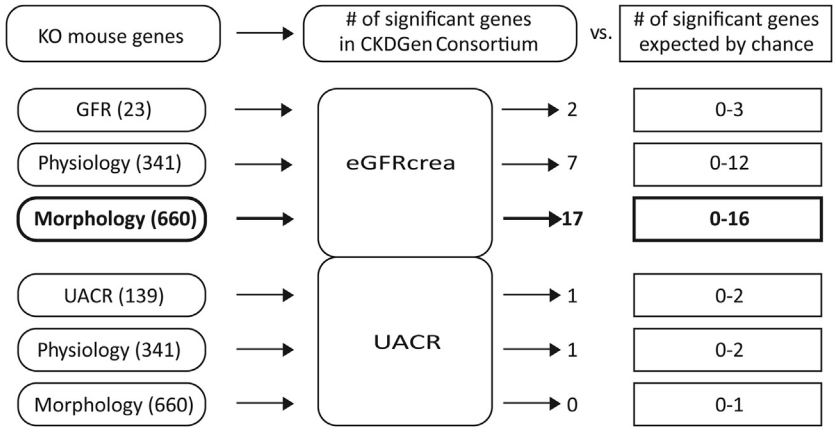

Figure 1 | Four gene lists were identified by retrieval of abnormal kidney phenotypes on the Mouse Genome Informatics website. These lists were interrogated in the CKDGen dataset for single-nucleotide polymorphisms associated with eGFRcrea or UACR. The observed number of significant genes was compared with the number expected by chance. Significant enrichment (morphology/eGFR) is indicated using bold boxes. GFR, glomerular filtration rate; eGFRcrea, estimated glomerular filtration rate; KO, knockout; UACR, urinary albumin-to-creatinine ratio.