Functional Genomic Annotation of Genetic Risk Loci in Chronic Kidney Disease

Ph.D. Thesis

Nóra Ledó, M.D.

Doctoral School of Basic Medicine
Semmelweis University

Consultants: Katalin Suszták, M.D., Ph.D.
András Tislér, M.D., Ph.D.

Official reviewers: Tamás Szelestei, M.D., Ph.D.
Kálmán Tory, M.D., Ph.D.

Head of the Complex Examination Committee:
Péter Nyirády, M.D., D.Sc.

Members of the Complex Examination Committee:
Ágnes Haris, M.D., Ph.D.
László Wagner, M.D., Ph.D.

Budapest
2017

1. Table of Contents

2. Abbreviations

ACSM2A/2B	Acyl-CoA Synthetase Medium-Chain Family Member 2A/2B
ACSM5	Acyl-CoA Synthetase Medium-Chain Family Member 5
ALDH3A2	Aldehyde dehydrogenase 3 family, member A2
ANOVA	Analysis of Variance
ANXA9	Annexin A9
bGFR	becsült glomeruláris filtrációs ráta
BMI	Body Mass Index
BUN	Blood Urea Nitrogen
cDNA	Complementary deoxyribonucleic acid
CELA2A/B	Chymotrypsin Like Elastase Family Member 2A/B
CELSR2	Cadherin EGF LAG Seven-Pass G-Type Receptor 2
CERS2	Ceramide synthase 2
ChIP-Seq	Chromatin Immunoprecipitation followed by next-generation Sequencing
CI	confidence interval
CKD	Chronic Kidney Disease
CLTB	Clathrin, light chain B
CRAT	CKD risk associated transcript
CTSS	Cathepsin S
DAB2	Disabled homolog 2
DAVID	Database for Annotation, Visualization and Integrated Discovery
DKD	Diabetic kidney disease
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
eGFR	estimated Glomerular Filtration Rate
ENCODE	Encyclopedia of DNA Elements
eQTL	Expression Quantitative Trait Loci
ERBB2	Erb-B2 Receptor Tyrosine Kinase 2
ESRD	End-stage renal disease
FAM47E	Family with sequence similarity 47, member E

FPKM	Fragments Per Kilobase of transcript per Million mapped reads
FYB	FYN binding protein
GFR	Glomerular Filtration Rate
GNAT2	G Protein Subunit Alpha Transducin 2
GP2	Glycoprotein 2
GRCh37/hg19	Genome Reference Consortium Human Build 37, synonym: Human Genome version 19 -human reference sequence, February 2009
GWAS	Genome wide association study
HGNC	Human Genome Organization Gene Nomenclature Committee
IF	Interstitial fibrosis
IPA	Ingenuity Pathway Analysis
IRB	Institutional Review Board
JAG1	Jagged1
kbp	kilobase pair
KDIGO	Kidney Disease: Improving Global Outcomes
MAGI2	Membrane-associated guanylate kinase 2
Mbp	Megabase pair
mRNA	Messenger Ribonucleic Acid
MuTHER	Multiple Tissue Human Expression Resource
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
Ph.D.	Doctor of Philosophy
Pcorr	Corrected P value after Benjamini-Hochberg-based multiple testing correction
PDILT	Protein disulfide isomerase-like, testis expressed
PLXDC1	Plexin domain containing 1
PSRC1	Proline and Serine Rich Coiled-Coil 1
QRT-PCR	Quantitative real time polymerase chain reaction
RELA	Nuclear factor NF-κB p65 subunit
RIN	RNA integrity number
RMA16	Robust Multi-Array Average
RNA	Ribonucleic acid
SD	Standard deviation

SLC7A9	Solute Carrier Family 7 Member 9
SLC34A1	Solute Carrier Family 34 Member 1
SLC47A1	Solute Carrier Family 47 Member 1
SNP	Single nucleotide polymorphism
SORBS1	Sorbin and SH3 Domain Containing 1
SORT1	Sortilin 1
TGF-β1	Transforming growth factor beta 1
TNF	Tumor necrosis factor
UCSC	University of California Santa Cruz
UMOD	Uromodulin
VEGFA	Vascular endothelial growth factor A
WDR72	WD Repeat Domain 72

3. Introduction

3.1. Chronic kidney disease, as a gene environmental disease

Chronic kidney disease (CKD) is defined as abnormalities of kidney structure or function with implications for health, which is present for more than 3 months - according to the Kidney Disease: Improving Global Outcomes (KDIGO) guideline (1). Kidney function is mostly measured by the filtration capacity of the kidneys (glomerular filtration rate - GFR), based on the plasma clearance of endogenous creatinine. CKD is classified based on cause, GFR category (G1-5), and albuminuria category (A1-3). Based on the estimated glomerular filtration rate (eGFR, calculated by GFR estimating equations), CKD is classified in five stages: stage 1 (eGFR > 90 ml/min/1.73m^2), stage 2 (eGFR between 60 and 89 ml/min/m^2), stage 3 (eGFR between 30 and 59 ml/min/1.73m^2), stage 4 (eGFR between 15 and 29 ml/min/1.73m^2) and stage 5 (eGFR < 15 ml/min/1.73m^2).

The prevalence of chronic kidney disease (stages 1-4) is as high as 15.2% (95% confidence interval (CI): 14.1%-16.1%) in the United States, based on the data of the Chronic Kidney Disease Surveillance System of the Centers for Disease Control and Prevention (www.cdc.gov/ckd). CKD is ranked to the 9th leading cause of death in the United States in 2014 according to the National Center of Health Statistics (www.cdc.gov/nchs). A recent meta-analysis found a 13.4% global prevalence of CKD of all stages (95% CI: 11.7%-15.1%) and 10.6% of CKD stages 3-5 (95% CI: 9.2%-12.2%), with a highest prevalence of CKD of all stages in Europe (18.38% (95% CI: 11.57%-25.20%) compared to other geographical regions (2). A recent community-based study in the United States found that the risk of death increases as the eGFR decreases below 60 ml/min/1.73m^2. The study revealed that the adjusted hazard ratio for death was 1.2 with an eGFR of 45 to 59 ml/min/1.73 m^2, 1.8 with an eGFR of 30 to 44 ml/min/1.73 m^2, 3.2 with an eGFR of 15 to 29 ml/min/1.73 m^2, and 5.9 with an eGFR of less than 15 ml/min/1.73 m^2. The adjusted hazard ratio of cardiovascular events and hospitalization also increased inversely with the eGFR in this population (3). These epidemiological findings indicate that chronic renal insufficiency has a great impact on both quality of life and public health financial resources.

Although there are CKD cases caused by monogenetic diseases, including polycystic kidney diseases and some glomerular diseases; chronic kidney disease is mostly a complex gene environmental disease, several environmental and genetic factors affect its development. Diabetes and hypertension are the two most important causes of chronic renal insufficiency, but CKD development clearly has a genetic component. Different studies found that heritability estimates of eGFR (based on serum creatinine levels) were between 0.41 and 0.75 in individuals with diabetes or hypertension, respectively (4,5), and 0.33 in a population-based sample (6). While previous genetic studies have identified rare genetic variants causing different forms of monogenetic kidney disease, common CKD susceptibility variants have been difficult to detect reproducibly by linkage analyses or candidate gene studies. Complex traits such as CKD often affected by multiple genetic factors, which should be examined in the general population that carry the disease, rather than by familial linkage analysis.

3.2. Different genetic methods for understanding CKD development

3.2.1. Genome-wide association studies (GWAS)

At present, one of the most powerful experiments to understand the genetics of a complex trait such as CKD is the genome-wide association study (GWAS). GWAS examines genetic variants across the human genome to identify associations between variants and phenotypes. To detect genetic variants that have small effects or appear with low frequency in complex-trait disease development requires very large study cohorts for sufficient statistical power. To avoid type I error (false positive results), a multiple-testing corrected p-value is used, most frequently the Bonferroni correction for multiple tests, where the cutoff p-value of 0.05 is corrected by the approximate one million independent tests to generate the threshold (7,8).

The GWAS divides the population into two groups of individuals: one group with a disease/parameter (cases) and another group of otherwise similar people without the parameter (controls). If a variant (e.g. single nucleotide polymorphism [SNP]) is more frequent in people with the disease, the SNP is said to be associated with the disease. In the discovery phase of the GWASs, variants that have statistically significant allele frequency differences associated with disease phenotypes are identified. These

significantly associated markers from the discovery phase are evaluated for association in additional independent study samples. Replication serves to confirm association and to detect potential bias.

Several parameters of kidney dysfunction were used as a quantitative trait in GWASs examining chronic kidney disease: most of the studies use eGFR value as a continuous trait or chose patients as cases with eGFR below 60 ml/min/1.73m^2, based on serum creatinine or cystatin C levels (9-17). However, other parameters were also used, such as the presence of end-stage renal disease (ESRD) (18-25), albuminuria (16,25-29) or proteinuria (21,22,30-33). A recently published meta-analysis of multiple cohorts with the largest sample size to date for kidney function included 175,000 individuals, and 53 loci were identified (29 known and 24 novel loci). Most of these variants are associated with eGFR (based on serum creatinine levels), one with eGFR (based on serum cystatin C levels) and four with the diagnosis of CKD (34).

GWASs became possible, because the genetic information is inherited in large genetic blocks. Linkage disequilibrium (LD) is used to describe the likeliness of the non-random association of alleles at different loci. If the coefficient value (r) of LD is 0, the variants are not inherited together, while the variants are always inherited together with r=1. In the haplotype or LD blocks, where $r^2 \geq 0.8$, there are several SNPs which are inherited together and one SNP, named leading or tagging SNP, represents that block. Therefore, we do not have to test the association with each of the 20 million genetic variations but can use fewer (about 1 million) SNPs representing the genetic variation in the entire genome. Although haplotype blocks made GWAS convenient and financially feasible, they also mean that we do not know which of the many variants within a single haplotype block is functionally relevant. To date, more than 88 million genomic variants have been cataloged in the 1000 Genomes Project.

In summary, GWAS is a very important way to reveal genetic variants in the association with CKD, however, further investigations are needed to find the functionally relevant polymorphisms.

3.2.2. Expression quantitative trait loci analysis (eQTL)

Genetic variants identified by GWASs explain only a small fraction of the heritability of CKD. To further understand the genetic basis of CKD, the variants

associated with CKD need to be tied to their target genes. Identifying quantitative phenotypes that are associated with these SNPs can facilitate the mechanistic studies for CKD development. Genomic loci which can contribute to the variation of gene expression levels are called expression quantitative trait loci. Loci located close (within 1 Megabase pair (Mbp) distance) to the transcription start site of the affected gene are called "cis-" eQTLs, while loci in a greater distance -even on other chromosomes- called "trans-" eQTLs. The examination of genetic variations and the transcriptome of the subjects simultaneously can reveal SNPs acting as eQTLs. Disease-associated genetic variants can alter binding sites for important transcription factors and influence the expression of nearby genes and act as an eQTL (35-39). Genetic variants can potentially alter steady-state expression of genes, in which case they interfere with basal transcription factor binding or can alter the amplitude of transcript changes after signal-dependent transcription factor binding. One way to prioritize regions is to combine statistical association of genetic variants with complex trait (GWAS signals) and association of genetic variants with gene expression (eQTL signals). Trait-associated GWAS SNPs found to act significantly more likely as eQTLs than expected by chance (40).

Usually the effect of the loci on the gene expression levels are examined in healthy, control subjects. For example, Musunuru et al. used the expression profiles of 960 normal healthy liver tissues to find association between the locus rs646776 (Chr1p13) associated with both plasma low-density lipoprotein cholesterol and myocardial infarction and the expression of Cadherin EGF LAG Seven-Pass G-Type Receptor 2 (*CELSR2*), Proline and Serine Rich Coiled-Coil 1 (*PSRC1*) and Sortilin 1 (*SORT1*) with microarray. The association between the locus and *PSRC1* and *SORT1* genes could be validated with quantitative real time polymerase chain reaction (QRT-PCR) in 62 normal, healthy samples. Finally, the research group demonstrated that Sort1 alters plasma low-density and very low-density lipoprotein cholesterol particle levels in mice (36). In kidney research, the association between the *UMOD* protective haplotype and the expression of the *UMOD* gene were examined in kidney samples only with normal function (eGFR> 90 ml/min/1.73m^2) (37).

Most eQTL analyses of human samples were performed in immortalized cell lines or circulating cells, because several other tissue types have been difficult to collect in large enough numbers to perform eQTL analysis (41). The transcriptome is tissue-type

specific, thus surrogate cell types cannot represent organ-specific regulation of gene expression by variants. On the other hand, there are clear examples in the literature for cross-tissue similarity when comparing results of eQTL studies conducted in large populations. Nica et al. found that 30% of the eQTLs are shared among three tissues (lymphoblastoid cell lines, skin and fat) (42). Also, major cross-tissue similarity was observed when eQTL analysis in whole blood was compared to other eQTL studies conducted in large population of B-cells, lung and liver tissue (40–70%) (43). Based on the possible cross-tissue similarity in eQTL results, there is a strong rationale for screening the SNPs of our interest in other eQTL databases to highlight potentially important genes.

Taken together, eQTL analysis is a valuable tool to understand the connection between the polymorphisms and gene expression alterations, and CKD-associated SNPs can be more accurately understood by using eQTL to link to potential target genes, and could be studied for their relevant biological functions.

3.2.3. Functional genomics

While descriptive genomics focuses on the structure of the DNA (deoxyribonucleic acid) with genetic mapping and DNA sequencing, functional genomics, part of genomics as a discipline, aims to understand the dynamic function of the genome. Functional genomics focuses on processes like transcription, translation, gene expression regulation, protein-protein interactions, etc. One of the important goals of this scientific field is to understand and find the function of the non-coding DNA regions. This so-called “junk” DNA is very important to be examined, since 83% of the disease-associated SNPs are localized to the non-coding region of the genome (35), and it is still unclear how they induce illness.

In 2003, the Encyclopedia of DNA Elements (ENCODE) project started and drew the attention to the non-coding DNA regions. The aim of the project is to identify all the functional DNA elements of the genome, both in the coding and non-coding regions. The project examines DNA and protein interactions to identify transcriptional factor binding sites, such as promoter and enhancer regions. Novel technologies were developed to unravel the functional significance of these regions, such as chromatin immunoprecipitation followed by next-generation sequencing (ChIP-Seq) or DNaseI

(deoxyribonuclease I) footprints. The ENCODE project uses cultured human cell lines of endothelial, fibroblast, myocyte, stem cell, erythroid, epithelial and lymphoid origins. Reports from the project indicate that most complex trait polymorphisms are localized to gene regulatory regions in target cell types (44-46).

Here, in this Ph.D. work several methods of functional genomics were used -described below-, mainly to study the transcriptome (e.g. microarray, RNA-sequencing, QRT-PCR) and perform gene ontology and network analysis.

In summary, GWASs can reveal the associations between a chosen parameter, such as renal function, and genetic variants, and can identify the disease-associated loci. The relationship between SNPs and gene expression can be examined by eQTL analysis, while functional genomics is applied in search for genetic basis (such as transcript level changes, gene expression regulation, etc.) of the functional changes (e.g. renal function). (*Figure 1.*)

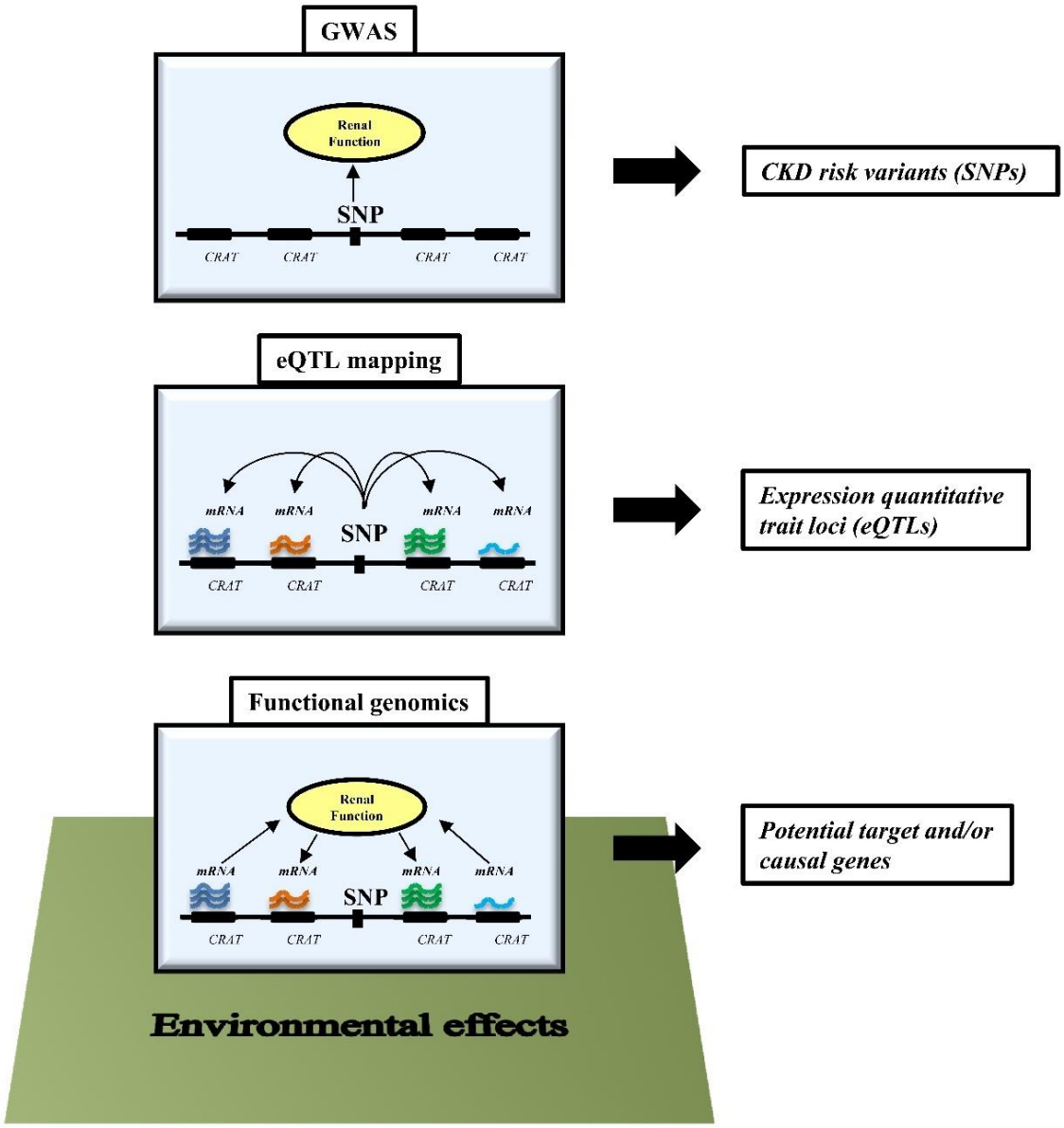


Figure 1. Schematic representation of different experimental designs to understand CKD development.

Genome-wide association studies (GWASs) examine the relationship between genetic variants (SNP, single nucleotide polymorphism) and disease state (CKD, chronic kidney disease). The eQTL (expression quantitative trait loci) analysis examines the relationship between transcript levels and genetic variations. The relationship between transcript levels around CKD risk variants and kidney function can be studied by functional genomics, by examining the contribution of genetic and environmental factors. *CRAT: CKD risk associated transcripts*

3.2.4. Other methods of CKD research

High-throughput omics datasets can be integrated to complex phenotypic disease signatures with the help of "top-down" systems biology approaches and reconstruct protein-protein interactions. Meanwhile, comprehensive molecular data from basic science ("bottom-up") are also important to understand the development of a disease (47).

In basic science, several methods can help to understand CKD development, from kidney cell cultures to animal models. For example, there are several animal models used to understand diabetic nephropathy (e.g. Streptozotocin-induced diabetic animals, Akita diabetic mice, db/db mice, Zucker diabetic fatty rats, Wistar fatty rats, etc.). The perfect animal model should exhibit progressive albuminuria and a decrease in renal function, as well as the characteristic histological changes that are observed in cases of human diabetic nephropathy. A rodent model that strongly exhibits all these features of human diabetic nephropathy has not yet been developed (48,49). Unilateral ureteral obstruction and folic acid induced nephropathy rodent models are also widely used to investigate interstitial fibrosis, beside several other animal models (50).

Recently a new and interesting field has become important part of CKD research: the epigenetics. Epigenetics is the heritable information during cell division other than the DNA sequence itself. The epigenome can be reshaped by environmental effects and as an "environmental footprint" contribute to the variation of phenotypes. The DNA in the nucleus has a highly-organized form wrapped around by proteins called histones. The state of its structure can guide transcriptional factor binding. Different stress factors from the environment can affect the epigenome through cytosine methylation (and other modification of cytosine) and histone-tail modifications. The presence of specific histone-tail modifications can identify cell-type specific gene regulatory regions, such as promoters, enhancers, silencers and insulators. As mentioned above, with the ChIP-Seq method these specific histone-tail proteins can be found, providing a map of the potential localization of the gene regulatory regions (44).

While the ENCODE project did not include kidney cell lines, there are studies examining the epigenetics of the kidney in CKD. For example, a genome-wide cytosine methylation analysis of control and diseased kidney epithelial cells was performed by Ko et al., and more than 4000 differentially methylated regions were found in CKD samples,

most of them in developmental and fibrosis-related DNA regions. These differentially methylated regions were enriched not on promoter, but on enhancer regions (51).

In summary, understanding the development of chronic kidney disease and the underlying mechanisms are challenging. Chronic kidney disease is a very complex trait; therefore, CKD research requires complexity itself. The methodology of CKD research needs to include both basic science through cell lines and animal models and high-throughput technologies with genome-, epigenome- and transcriptome-wide studies. In this Ph.D. work, I used functional genomic approaches to prioritize potentially important transcripts in CKD development.

4. Objectives

We hypothesized that polymorphisms associated with renal disease will influence the expression of nearby transcript levels in the kidney. In this Ph.D. work, I mapped the expression of these transcripts in normal and disease human kidney samples. I used functional genomics and systems biology approaches to investigate tissue-specific expression of transcripts and their correlation with kidney function.

The goals of the Ph.D. work were:

1. Providing a dataset of potential causal and/or target genes in the vicinity of the CKD risk associated loci

2. Identifying critical pathways associated with kidney function decline for further analysis

5. Methods

5.1. Human kidney samples

5.1.1. Tissue handling and microdissection

The human kidney samples were obtained from routine surgical nephrectomies. For RNA sequencing analysis, leftover portions of diagnostic kidney biopsies were used (n=2). Only the normal, non-neoplastic part of the tissue was used for further investigation. Samples were de-identified, and corresponding clinical information was collected by an individual who was not involved in the research protocol. The tissue and data collecting procedure was approved by the institutional review boards (IRBs) of the Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY, USA (IRB 2002–202) and the University of Pennsylvania, Philadelphia, PA, USA (IRB 815796).

The fresh kidney tissue was immediately placed and stored in RNAlater solution (Thermo Fisher Scientific, Ambion, Waltham, MA, USA) according to the manufacturer's instruction: the tissue was cut into pieces -smaller than 0.5 cm in any dimension- and stored at 4 ℃ overnight, allowing the solution to penetrate the whole tissue. We stored the samples in RNAlater solution at -80 ℃ until the experiments.

Before the RNA (ribonucleic acid) isolation, the kidney tissue in RNAlater solution was manually microdissected for glomerular and tubular compartment under a microscope. Using fine forceps, the glomeruli were removed from the kidney tissue and processed separately. We refer the rest of the kidney tissue as "tubules", however, it contains not only tubules but other kind of tissues, e.g. vessels and connective tissue. (*Figure 2*.)

5.1.2. Sample characteristics

To examine gene expression changes, we extracted RNA form 95 tubule samples and 51 glomeruli samples, furthermore, 41 tubule samples were used for external validation. The kidney samples were obtained from a diverse population, samples from patients of different age, gender, ethnicity with hypertensive or diabetic nephropathy were examined. Our dataset contains samples from non-Hispanic white,

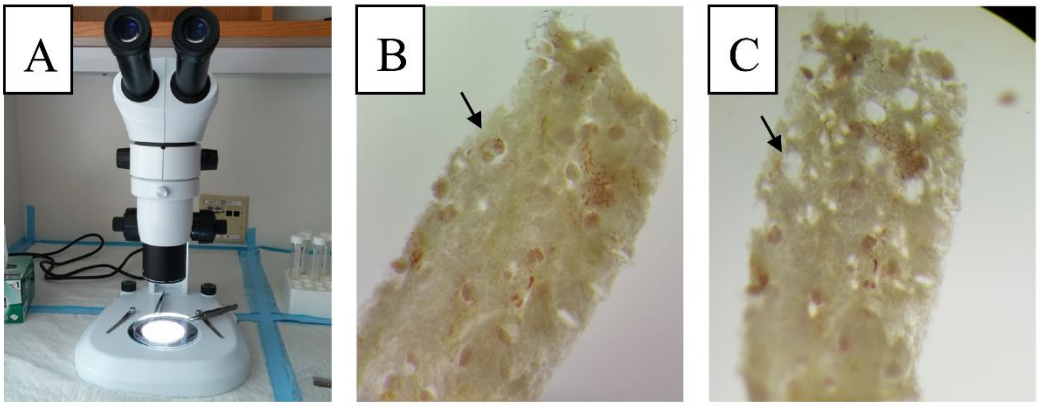


Figure 2. Microdissection of human kidney samples stabilized in RNAlater

Microscope and forceps used for microdissection (A). Intact human kidney sample -glomerulus (arrow) (B). Several glomeruli were removed (arrow) (C)

African American, Asian, Hispanic and multiracial race, so we examined our dataset to exclude any ethnicity driven gene expression changes. We performed statistical analysis (one-way ANOVA – analysis of variance) to identify gene expression differences driven by ancestry in our database. We compared gene expression profiles of kidneys obtained from non-Hispanic white, African American and other ethnicities, and were unable to identify transcripts with statistically significant differential expression in our data. (Expression profiles of 95 tubule samples and 51 glomerular samples were examined.) (*Table 1.*). Review of the literature also failed to identify ancestry specific gene expression differences. Therefore, we believe that race is not a critical driver of gene expression differences in our dataset.

The main part of our analysis was examining gene expression correlation with renal function (based on estimated glomerular filtration rate (eGFR) according to the Chronic Kidney Disease Epidemiology Collaboration [CKD-EPI] determination (52)), therefore, we analyzed the correlation between eGFR and clinical and histopathological changes, to exclude any unexpected correlations in our dataset. As expected, we found significant correlation between eGFR and serum creatinine levels, blood urea nitrogen levels (BUN), the percentage of glomerulosclerosis and interstitial fibrosis. On the other hand, we failed to detect any significant correlation between renal function (eGFR) and age, serum glucose levels, serum albumin levels and body mass index (BMI). The demographics, clinical information and histopathological analysis of the samples are summarized in *Table 2 (a-d).*

Table 1. Gene expression is not driven by ancestry in our microarray data sets

Statistical analysis (one-way ANOVA) between three ethnic groups (non-Hispanic white vs. African American vs. other ethnicity) was performed to search for differentially expressed transcripts. We failed to detect any significant gene expression changes among CRATs (CKD-risk associated transcripts) and among all entities (not shown). *eGFR: glomerular filtration rate, SD: standard deviation, P: P-values after Benjamini-Hochberg-based multiple testing correction, W: non-Hispanic white, AA: African American, O: Other ethnicity, GNAT2: G Protein Subunit Alpha Transducin 2, PSRC1: Proline and Serine Rich Coiled-Coil 1, CELA2A/B: Chymotrypsin Like Elastase Family Member 2A/B, JAG1: Jagged 1*

Data set	**Analyzed groups**	**CRATs with lowest P value**	**Gene expression values (Mean ± SD)**	**P**
Tubule samples (n=95)	W: (n=19) AA: (n=35) O: (n=41)	*GNAT2*	W: 0.380 ± 0.659 AA: -0.066 ± 0.441 O: 0.204 ± 0.531	0.93
Tubule samples eGFR>60 ml/min/1.73m^2 (n=56)	W: (n=12) AA: (n=17) O: (n=27)	*PSRC1*	W: -0.056 ± 0.511 AA: 0.330 ± 0.564 O: 0.113 ± 0.494	0.99
Glomerular samples (n=51)	W: (n=10) AA: (n=18) O: (n=23)	*CELA2A/B*	W: -0.115 ± 0.104 AA: 0.059 ± 0.204 O: 0.140 ± 0.244	0.70
Glomerular samples eGFR>60 ml/min/1.73m^2 (n=27)	W: (n=5) AA: (n=11) O: (n=11)	*JAG1*	W: -0.557 ± 0.341 AA: -0.342 ± 0.775 O: 0.458 ± 0.670	0.80

Table 2. Demographics, clinical information and histological analysis of glomerular samples (a), tubule samples (b), tubule samples of the external microarray validation (c), tubule samples of the QRT-PCR validation (d)

Data are presented as mean and standard deviation with the median values or percentage. Estimated Glomerular Filtration Rate (eGFR) was calculated according to the CKD-EPI equation. Pearson product moment correlation or Spearman correlation coefficient (R coefficient) was used to measure the strength of association between age, BMI (body mass index), serum-glucose, blood pressure (systole and diastole), serum-creatinine, BUN (blood urea nitrogen), serum-albumin, percentage of glomerulosclerosis and interstitial fibrosis and eGFR; depending on the results of the D'Agostino-Pearson normality tests. Asterisks (*) indicate when the two-tailed tests reached statistical significance ($P < 0.05$).

Table 2.a. Patient Demographics (Samples from Glomeruli)			
Total: n=51		**% or mean ± SD (median)**	**correlation with GFR (R coefficient)**
Gender	Male	47.1 %	
	Female	52.9 %	
Race	Non-Hispanic White	19.6 %	
	African American	35.3 %	
	Asian	5.9 %	
	Hispanic	15.7 %	
	Multiracial	9.8 %	
	Unknown	13.7%	
Diabetes		45.1 %	
Hypertension		80.4 %	
Age (years)		61.08 ± 12.9 (63)	-0.262
BMI (Body Mass Index) (kg/m^2)		32.18 ± 15.7 (29.2)	-0.097
Serum glucose (mg/dL)		124.8 ± 51.3 (115)	-0.254
Blood pressure - systole (mm Hg)		136.52± 20.2 (130)	-0.153
Blood pressure - diastole (mm Hg)		81.24 ± 13.4 (80)	-0.081

eGFR (ml/min/1.73m^2)	58.53 ± 28.5 (60.9)	
Serum creatinine (mg/dL)	1.66 ± 1.4 (1.2)	-0.893 *
BUN (Blood Urea Nitrogen) (mg/dL)	21.59 ± 14.2 (19)	-0.653 *
Serum albumin (g/dL)	3.75 ± 0.8 (4)	0.219
Glomerulosclerosis (%)	11.45 ± 17.4 (3.9)	-0.511 *
Interstitial Fibrosis (%)	13.91 ± 13.6 (10)	-0.586 *

Table 2.b. Patient Demographics (Samples from Tubules)

Total: n=95		**% or mean ± SD (median)**	**correlation with eGFR (R coefficient)**
Gender	Male	57.9 %	
	Female	42.1 %	
Race	Non-Hispanic White	20.0 %	
	African American	36.8 %	
	Asian	3.2 %	
	Hispanic	6.3 %	
	Multiracial	17.9 %	
	Unknown	15.8%	
Diabetes		38.9 %	
Hypertension		76.8 %	
Age (years)		63.57 ± 13.5 (65)	-0.131
BMI (Body Mass Index) (kg/m^2)		29.77 ± 9.3 (29)	0.150
Serum glucose (mg/dL)		135.4 ± 65.3 (118)	0.153
Blood pressure - systole (mm Hg)		138.97 ± 24.8 (136.5)	-0.299 *
Blood pressure - diastole (mm Hg)		78.05 ± 13.7 (78.5)	-0.174
eGFR (ml/min/1.73m^2)		60.08 ± 29.8 (64.1)	

Serum creatinine (mg/dL)	2.05 ± 2.5 (1.1)	-0.894 *
BUN (Blood Urea Nitrogen) (mg/dL)	23.2 ± 13.7 (19)	-0.696 *
Serum albumin (g/dL)	3.96 ± 0.7 (4.1)	0.228 *
Glomerulosclerosis (%)	17.97 ± 27.3 (5.5)	-0.570 *
Interstitial Fibrosis (%)	16.47 ± 21.6 (10)	-0.732 *

Table 2.c. Patient Demographics (Samples from Tubules for Replication)

Total: n=41		**% or mean ± SD (median)**	**correlation with eGFR (R coefficient)**
Gender	Male	41.5 %	
	Female	58.5 %	
Race	Non-Hispanic White	19.5 %	
	African American	41.5 %	
	Asian	2.4%	
	Hispanic	14.6 %	
	Multiracial	4.9 %	
	Unknown	17.1	
Diabetes		51.2 %	
Hypertension		78.0%	
Age (years)		60.2 ± 13.3 (60)	-0.177
BMI (Body Mass Index) (kg/m^2)		30.26 ± 6.5 (30.5)	0.042
Serum glucose (mg/dL)		140.83 ± 65.9 (129)	0.072
Blood pressure - systole (mm Hg)		142.44 ± 22.7 (151)	-0.504
Blood pressure - diastole (mm Hg)		76.22 ± 13.8 (75)	-0.246
eGFR ($ml/min/1.73m^2$)		52.7 ± 28.2 (55.7)	
Serum creatinine (mg/dL)		2.01 ± 1.8 (1.2)	-0.796 *

BUN (Blood Urea Nitrogen) (mg/dL)	25.0 ± 13.1 (22)	-0.749 *
Serum albumin (g/dL)	3.69 ± 0.9 (3.9)	0.409 *
Glomerulosclerosis (%)	17.97 ± 25.5 (14.3)	-0.641 *
Interstitial Fibrosis (%)	19.93 ± 22.0 (15)	-0.769 *

Table 2.d. Patient Demographics (Tubule samples with QRT-PCR validation)

Total: n=46		**% or mean ± SD (median)**	**correlation with eGFR (R coefficient)**
Gender	Male	54.35 %	
	Female	45.65 %	
Race	Non-Hispanic White	21.7 %	
	African American	41.3 %	
	Asian	4.35 %	
	Hispanic	4.35 %	
	Multiracial	8.7 %	
	Unknown	19.6%	
Diabetes		52.2 %	
Hypertension		73.9 %	
Age (years)		62.2 ± 13.1 (63.5)	0.162
BMI (Body Mass Index) (kg/m^2)		28.4 ± 6.3 (28.5)	0.197
Serum glucose (mg/dL)		145.8 ± 79.6 (117.5)	0.015
Blood pressure - systole (mm Hg)		139.47 ± 29.9 (135)	-0.377 *
Blood pressure - diastole (mm Hg)		77.81 ± 15.5 (76.5)	-0.291 *
eGFR ($ml/min/1.73m^2$)		54.2 ± 32.8 (58.1)	
Serum creatinine (mg/dL)		2.60 ± 3.1 (1.2)	-0.743 *
BUN (Blood Urea Nitrogen) (mg/dL)		25.93 ± 13.7 (21)	-0.712 *

Serum albumin (g/dL)	3.94 ± 0.6 (4)	0.064
Glomerulosclerosis (%)	23.7 ± 33.1 (6.2)	-0.748 *
Interstitial Fibrosis (%)	21.53 ± 25.3 (10)	-0.737 *

5.2. Sample processing and data analysis

5.2.1. Microarray process and data analysis

Dissected tissue was homogenized, and RNA was prepared using RNAeasy mini columns (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions: the tissue was placed in the lysis buffer and homogenized with an Omni Tissue Homogenizer (Omni, Kennesaw, GA, USA). DNase (deoxyribonuclease) digestion was used as an additional step to improve RNA purification. RNA quality and quantity were determined using the Laboratory-on-Chip Total RNA PicoKit Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). Only samples without evidence of degradation were further used (RNA integrity number [RIN] >6).

For microarray analysis, we prepared the first and second strand of the complementary DNA (cDNA) and after amplification, purification and cDNA fragmentation, we labelled the cDNA fragments. Purified total RNAs from 95 tubule samples were amplified using the Ovation PicoWTA SystemV2 (NuGEN Technologies, San Carlos, CA, USA) and labeled with the Encore Biotin Module (NuGEN) according to the manufacturer's protocol. The purified total RNAs from 51 glomerular samples and 41 tubule samples used for validation were amplified using the Two-Cycle Target LabelingKit (Affymetrix, Santa Clara, CA, USA) as per the manufacturer's protocol. Transcript levels were analyzed using Affymetrix U133A arrays.

After hybridization and scanning on microarray chips, raw data files were imported into GeneSpring GX software, version 12.6 (Agilent Technologies). Raw expression levels were summarized using the RMA16 (Robust Multi-Array Average) algorithm. Normalized values were generated after log transformation and baseline transformation. GeneSpring GX software then was used for statistical analysis.

5.2.2. RNA sequencing analysis

RNA sequencing was carried out on microdissected kidney tubules from kidney biopsies. Total RNA was isolated using the RNeasy mini columns (Qiagen) according to the manufacturer's protocol, as described above. An additional DNase digestion step was performed to ensure that the samples were not contaminated with genomic DNA. RNA purity was assessed using the Laboratory-on-Chip Total RNA PicoKit Agilent 2100 BioAnalyzer (Agilent Technologies). Each RNA sample had an A260:A280 ratio 1.8 and an A260:A230 ratio 2.2, with an RIN>9.0. Single-end 100-basepair RNA sequencing was carried out an Illumina HiSeq2000 machine (Illumina, San Diego, CA, USA). RNA sequencing reads were aligned to the human genome (GRCh37/hg19, University of California Santa Cruz [UCSC]) with the software TopHat (version 2.0.9) and transcriptome (hg19 RefSeq from Illumina iGenomes) using the software Cufflinks (version 2.1.1 Linux_x86_64) (53,54). We counted the number of fragments mapped to each gene annotated in the UCSC hg19. Transcript abundances were measured in Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Sequence data can be accessed at the National Center for Biotechnology Information's Gene Expression Omnibus (Accession number: GSE60119).

5.2.3. Quantitative real time polymerase chain reaction (QRT-PCR) analysis

Using reverse transcriptase, 250 ng RNA was converted to cDNA using the cDNA Archive Kit (Thermo Fisher Scientific, Applied Biosystems, Waltham, MA, US) and QRT-PCR was run in the ViiA 7 System (Applied Biosystems) machine using SYBR Green Master Mix (Applied Biosystems) and gene-specific primers. The data were normalized and analyzed using the ΔΔCT method, ubiquitin was used as a housekeeping gene for normalization.

5.2.4. Genotyping of human kidney samples

After the disruption and homogenization of the human kidney tissue as described above, DNA was extracted and purified with the DNeasy Blood and Tissue Kit (Qiagen), according to the manufacturer's protocol. Genotyping for rs881858 and rs6420094 loci was run in the ViiA 7 System (Applied Biosystems) machine using TaqMan Genotyping Master Mix (Applied Biosystems) and specific TaqMan assay probes.

5.2.5. Histology

Glomerular sclerosis and interstitial fibrosis were evaluated using periodic acid–Schiff-stained kidney sections by two independent nephropathologists.

Immunohistochemistry was performed on paraffin-embedded sections with the following antibodies: UMOD (AAH35975, Sigma Aldrich, St. Louis, MO, USA), VEGFA (Ab46154, Abcam, Cambridge, MA, USA) and ACSM2A (Ab181865, Abcam). We used the Vectastain Mouse on Mouse or anti-rabbit Elite ABC Peroxidase Kit and 3,3'diaminobenzidine (DAB) for visualizations (Vector Laboratories, Burlingame, CA, USA). Antibody specificity was evaluated separately; secondary antibodies alone showed no positive staining.

5.3. Bioinformatics

5.3.1. Gene ontology and network analyses

We performed gene ontology analysis on the CKD risk associated transcripts of interest, using the Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources, available on-line at *david.abcc.ncifcrf.gov*. (55,56)

To perform network analysis on the transcripts with expression levels showing significant linear correlation with eGFR, the transcripts were exported to the Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Qiagen). This software determines the top canonical pathways by using a ratio (calculated by dividing the number of genes in each pathway that meet cutoff criteria by the total number of genes that constitute that pathway) and then scoring the pathways using a Fisher exact test (P value < 0.05).

5.3.2. Processing publicly available datasets

We compared absolute expression levels of the transcripts of interest by processing the data of the publicly available Illumina Body Map database (The European Bioinformatics Institute, *www.ebi.ac.uk*) which provides RNA sequencing results in 16 different human organs.

For additional expression quantitative trait loci (eQTL) analysis, we examined multiple different datasets with the help of the publicly available eQTL browser at *www.ncbi.nlm.nih.gov*. These datasets included the MuTHER (Multiple Tissue Human

Expression Resource) and other studies, where transcript levels were available from liver, adipose, and lymphoblastoid samples (42,57-60).

For evaluation of the expressions of the genes of interest on protein level, additional to our own immunohistochemistry results, we reviewed the publicly available data of The Human Protein Atlas (*www.proteinatlas.org*) (61).

5.4. Overview of the used statistical methods

For statistical analysis of the demographic, clinical and histopathological parameters, Pearson product moment correlation or Spearman correlation coefficient (R coefficient) was used to measure the strength of association between age, BMI (body mass index), serum-glucose, blood pressure (systole and diastole), serum-creatinine, BUN (Blood urea nitrogen), serum-albumin, percentage of glomerulosclerosis and interstitial fibrosis and eGFR; depending on the results of the D'Agostino-Pearson normality tests. The statistical significance of the correlation was calculated with two-tailed test (alpha=0.05). To compare the expression of the genotyped samples in our eQTL analysis, one-way ANOVA and Student's t-test were used. The statistical analyses were performed using Prism 6 software (GraphPad, La Jolla, CA, USA).

GeneSpring GX software was used for statistical analysis to process microarray data. Pearson product moment correlation was used to measure the strength of association between gene expression and eGFR. We used Benjamini–Hochberg multiple testing correction with a P value of 0.05. In the case of genes with more probe set identifications, the results with the lowest P values are represented.

6. Results

6.1. Identifying CKD risk associated transcripts (CRATs)

To identify CKD risk associated transcripts, we performed manual literature search to examine all genome-wide association studies - by the time of the beginning of our study- reporting genetic association for CKD-related traits (9-28,30-33). Many of these studies used different parameters as kidney disease indicators, such as the serum creatinine or cystatin C levels, the presence of CKD or end stage renal disease (ESRD) or albuminuria/proteinuria. In our investigation, the SNPs associated with eGFR (based on serum creatinine or cystatin C calculations) or the presence of ESRD were included. Our literature analysis identified 10 publications meeting these criteria (9-15,18-20). Coding polymorphisms and SNPs that did not reach genome-wide significance ($P > 5 \times 10^{-8}$) were excluded from our study. Finally, 44 leading SNPs meeting these criteria were used for further analysis (*Table 3.*). Most publications did not differentiate cases based on disease etiology and included cases with hypertensive and diabetic kidney disease, nevertheless, three SNPs associated only with diabetic nephropathy, so they were also analyzed separately. There were only two SNPs that reached genome-wide significance in multiple studies (rs12917707 and rs9895661), these two SNPs were counted only once.

Table 3. List of single nucleotide polymorphisms (SNPs) that met our criteria

The table shows the list of the single nucleotide polymorphisms (SNPs) which reached the genome wide significance ($P < 5 \times 10^{-8}$) in the association with eGFR (estimated glomerular filtration rate, based on creatinine (crea) or cystatin C (cys) levels) and/or the presence of chronic kidney disease (CKD) or end stage renal disease (ESRD). SNPs which reached the genome-wide significance in multiple studies were counted only once (marked with "X" in the table). Genes less than 250 kbp (kilobase pair) from the leading SNPs are listed. Color-coding shows the baseline expression of the transcripts based on human kidney RNA sequencing, red: high expression, yellow: medium expression, green: low expression, blue: no expression. Genes with available probe set IDs on the microarray chip are marked bold. Gene symbols are official symbols approved by the Human Genome Organization Gene Nomenclature Committee (HGNC). *Chr: chromosome*

Table 3. List of single nucleotide polymorphisms (SNPs) that met our criteria								
	Leading SNPs	**Loca-tion (chr)**	**Position**	**Leading SNP functional location**	**Association parameter**	**Association p-value**	**Genes within 250-250kbp**	**Jour-nal**
1	rs10794720	10	1156165	Intronic	eGFRcrea	$p=2.1 \times 10^{-8}$	**LARP4B**, **GTPBP4**, **IDI2**, **IDI1**, **WDR37**, **ADARB2**	1
2	rs491567	15	53946593	Intronic	eGFRcrea	$p=1.3 \times 10^{-8}$	WDR72	1
3	rs267734	1	150951477	Upstream	eGFRcrea	$p=5.2 \times 10^{-9}$	**CTSS**, **CTSK**, **ARNT**, **SETDB1**, **CERS2**, **ANXA9**, **FAM63A**, **PRUNE**, **MLLT11**, BNIPL, **C1orf56**, GABPB2, **SEMA6C**, **CDC42SE1**, LYSMD1 **SCNM1**, TMOD4, **VPS72** **PIP5K1A**, **TNFAIP8L2**	1
4	rs347685	3	141807137	Intronic	eGFRcrea	$p=7.0 \times 10^{-9}$	**ATP1B3**, **TFDP2**, GK5, XRN1	1
5	rs4744712	9	71434707	Intronic	eGFRcrea	$p=7.2 \times 10^{-10}$	**PIP5K1B**, FAM122A, **PRKACG**, **FXN**	1
6	rs626277	13	72347696	Intronic	eGFRcrea	$p=2.9 \times 10^{-10}$	**DACH1**	1
7	rs1394125	15	76158983	Intronic	eGFRcrea	$p=3.7 \times 10^{-10}$	**SNUPN**, **IMP3**, SNX33, **CSPG4**, ODF3L1, UBE2Q2, NRG4, C15orf27	1
8	rs9895661	17	59456589	Intronic	eGFRcrea	$p=1.4 \times 10^{-8}$	**BCAS3**, **TBX2**, C17orf82, **TBX4**, **NACA2**	1
9	rs10109414	8	23751151	Intergenic	eGFRcrea	$p=1.0 \times 10^{-8}$	**NKX3-1**, NKX2-6, **STC1**	1
10	rs911119	20	23612737	Intergenic	eGFRcys	$p=2.3 \times 10^{-138}$	NAPB, CSTL1, **CST11**, **CST8**, CST9L, CST9, **CST3**, **CST4**, **CST1**, **CST2**, **CST5**	1
11	rs6465825	7	77416439	Intergenic	eGFRcrea	$p=3.5 \times 10^{-9}$	**PTPN12**, RSBN1L, TMEM60,	1

							PHTF2, MAGI2	
12	rs653178	12	112007756	Intronic	eGFRcys	$p=3.8 \times 10^{-8}$	CUX2, FAM109A, SH2B3, ATXN2, BRAP, ACAD10, ALDH2	1
13	rs6420094	5	176817636	Intronic	eGFRcrea	$p=3.8 \times 10^{-12}$	NSD1, RAB24, PRELID1, MXD3, LMAN2, RGS14, SLC34A1, PFN3, F12, GRK6, PRR7, DBN1, PDLIM7, DOK3, DDX41, FAM193B, TMED9, B4GALT7	1
14	rs11959928	5	39397132	Intronic	eGFRcrea	$p=1.8 \times 10^{-11}$	FYB, C9, DAB2	1
15	rs12917707	16	20367690	Upstream	eGFRcrea	$p=1.2 \times 10^{-20}$	GP2, UMOD, PDILT, ACSM5, ACSM2A, ACSM2B	1
16	rs2453533	15	45641225	Intergenic	eGFRcrea	$p=4.6 \times 10^{-22}$	DUOX1, DUOXA2, DUOXA1, SHF, SLC28A2, GATM, SPATA5L1, C15orf48, SLC30A4, BLOC1S6	1
17	rs17319721	4	77368847	Intronic	eGFRcrea	$p=1.1 \times 10^{-19}$	SCARB2, FAM47E, STBD1, CCDC158, SHROOM3	1
18	rs1933182	1	109999588	Intergenic	eGFRcrea	$p=1.3 \times 10^{-8}$	SARS, CELSR2, PSRC1, MYBPHL, SORT1, PSMA5, SYPL2, ATXN7L2, CYB561D1, AMIGO1, GPR61, GNAI3, AMPD2, GSTM2, GSTM4, GSTM1, GNAT2	1
19	rs16864170	2	5907880	Intergenic	CKD	$p=4.5 \times 10^{-8}$	SOX11	1
20	rs881858	6	43806609	Intergenic	eGFRcrea	$p=2.2 \times 10^{-11}$	POLH, GTPBP2, MAD2L1BP,	1

							RSPH9, **MRPS18A**, **VEGFA**, C6orf223	
21	rs7805747	7	151407801	Intronic	CKD	$p=8.6 \times 10^{-9}$	**RHEB**, **PRKAG2**	1
22	rs4014195	11	65506822	Intergenic	eGFRcrea	$p=3.3 \times 10^{-8}$	SCYL1, **LTBP3**, **SSSCA1**, **FAM89B**, **EHBP1L1**, **KCNK7**, **MAP3K11**, PCNXL3, **SIPA1**, **RELA**, **KAT5**, RNASEH2C, AP5B1, **OVOL1**, SNX32, **CFL1**, **MUS81**, **EFEMP2**, **CCDC85B**, **FOSL1**, **CTSW**, **FIBP**, **C11orf68**, TSGA10IP, **SART1**, **DRAP1**	1
23	rs12460876	19	33356891	Intronic	eGFRcrea	$p=5.5 \times 10^{-9}$	**ANKRD27**, RGS9BP, **NUDT19**, **TDRD12**, **SLC7A9**, CEP89, **C19orf40**, RHPN2, **GPATCH1**	1
24	rs2279463	6	160668389	Intronic	eGFRcrea	$p=8.7 \times 10^{-10}$	**IGF2R**, **SLC22A1**, **SLC22A2**, **SLC22A3**	1
25	rs10774021	12	349298	Intronic	eGFRcrea	$p=6.7 \times 10^{-9}$	**IQSEC3**, **SLC6A12**, **SLC6A13**, **KDM5A**, CCDC77, B4GALNT3	1
26	rs6431731	2	15863002	Intergenic	eGFRcrea	$p=4.6 \times 10^{-8}$	**DDX1**, **MYCN**	2
27	rs3925584	11	30760335	Intergenic	eGFRcrea	$p=1 \times 10^{-9}$	**MPPED2**, DCDC5, DCDC1	2
28	rs12124078	1	15869899	Intronic	eGFRcrea	$p=9.8 \times 10^{-10}$	FHAD1, **EFHD2**, **CTRC**, **CELA2A**, **CELA2B**, **CASP9**, **DNAJC16**, **AGMAT**, **DDI2**, **RSC1A1**, SLC25A34, TMEM82, FBLIM1	2

29	rs2453580	17	19438321	Intronic	eGFRcrea	p=4.6 x 10^{-8}	**EPN2**, **B9D1**, **MAPK7**, **MFAP4**, RNF112, **SLC47A1**, **ALDH3A2**, **ALDH3A1**, SLC47A2, **ULK2**	2
30	rs11078903	17	37631924	Intronic	eGFRcrea	p=2.4 x10^{-9}	FBXL20, **MED1**, **CDK12**, **NEUROD2**, PPP1R1B, **STARD3**, **PNMT**, **PGAP3**, **ERBB2**, **TCAP**	2
31	rs4293393	16	20364588	Intronic	eGFRcrea	p=2.6 x10^{-10}	**GP2**, **UMOD**, PDILT, **ACSM5**, **ACSM2A**, **ACSM2B**	3
X	rs12917707	16	20367690	Intronic	CKD	p=2.9 x 10^{-9}	**GP2**, **UMOD**, PDILT, **ACSM5**, **ACSM2A**, **ACSM2B**	4
32	rs6040055	20	10633313	Intronic	eGFRcrea	p=1 x 10^{-8}	**MKKS**, SLX4IP, **JAG1**	4
33	rs1731274	8	23766319	Intergenic	eGFRcys	p=4.6 x 10^{-8}	**STC1**, **NKX3-1**, NKX2-6	4
34	rs13038305	20	23610262	Intronic	eGFRcys	p=2.2 x 10^{-88}	NAPB, CSTL1, **CST11**, **CST8**, CST9L, CST9, **CST3**, **CST4**, **CST1**, **CST2**, **CST5**	4
35	rs10206899	2	73900900	Intronic	eGFRcrea	p=2.3 x 10^{-8}	**ALMS1**, **NAT8**, **NAT8B**, **TPRKB**, **DUSP11**, C2orf78, **STAMBP**, **ACTG2**	5
X	rs9895661	17	59456589	Intronic	eGFRcrea	p=4.8 × 10^{-11}	**BCAS3**, **TBX2**, C17orf82, **TBX4**, **NACA2**	6
36	rs11864909	16	20400839	Intronic	eGFRcrea	p=3.6 × 10^{-10}	**GP2**, **UMOD**, PDILT, **ACSM5**, **ACSM2A**, **ACSM2B**, **ACSM1**	6
37	rs13146355	4	77412140	Intronic	eGFRcrea	p=6.6 × 10^{-11}	**FAM47E**, **STBD1**, CCDC158, SHROOM3	6
38	rs10277115	7	1285195	Intergenic	eGFRcrea	p=1.0 × 10^{-10}	C7orf50, GPR146, **GPER**,	6

							ZFAND2A, UNCX, **MICALL2**, **INTS1**	
39	rs3828890	6	31440669	Unknown	eGFRcrea	$p=1.2 \times 10^{-9}$	**HLA-C**, **HLA-B**, **MICA**, **MICB**, **DDX39B**, **ATP6V1G2**, **LTA**, **NFKBIL1**, **LST1**, **NCR3**, **AIF1**, **PRRC2A**, **BAG6**, **C6orf47**, GPANK1, **CSNK2B**, LY6G5B, ABHD16A, **LY6G5C**, **APOM**, LY6G6F, **LY6G6C**, **DDAH2**, **C6orf25**, **LTB**, **TNF**	6
40	rs7208487	17	37543449	Intronic	eGFRcrea	$p=5.6 \times 10^{-9}$	**PLXDC1**, **CACNB1**, ARL5C, **RPL19**, STAC2, FBXL20, **MED1**, **CDK12**, **NEUROD2**, PPP1R1B, **STARD3**	7
41	rs4821469	22	36616445	Intergenic	ESRD	$p=1.78 \times 10^{-19}$	**APOL3**, APOL4, **APOL2**, **APOL1**, **MYH9**, **TXN2**	8
42	rs12437854	15	94141833	Intergenic	ESRD	$p=2 \times 10^{-9}$	no gene in <250 kbp distance	9
43	rs7583877	2	100460654	Intronic	ESRD	$p=1.2 \times 10^{-8}$	**AFF3**	9
44	rs1617640	7	100317298	In promoter but not missence	ESRD	$p=2.66 \times 10^{-8}$	**TSC22D4**, NYAP1, **AGFG2**, **SAP25**, **LRCH4**, **FBXO24**, **PCOLCE**, **MOSPD3**, **TFR2**, **ACTL6B**, **GNB2**, GIGYF1, **POP7**, **EPO**, ZAN, **EPHB4**, **SLC12A9**, **TRIP6**, **SRRT**, UFSP1, **ACHE**	10

List of Journals	
1	**New loci associated with kidney function and chronic kidney disease** Köttgen A et al. Nat Genet. 2010 May;42(5):376-84. doi: 10.1038/ng.568. Epub 2010 Apr 11.
2	**Genome-wide association and functional follow-up reveals new loci for kidney functio**n Pattaro C et al. PLoS Genet. 2012;8(3):e1002584. doi: 10.1371/journal.pgen.1002584. Epub 2012 Mar 29.
3	**Association of variants at UMOD with chronic kidney disease and kidney stones-role of age and comorbid diseases** Gudbjartsson DF et al. PLoS Genet. 2010 Jul 29;6(7):e1001039. doi: 10.1371/journal.pgen.1001039. Erratum in: PLoS Genet. 2010;6(11).
4	**Multiple loci associated with indices of renal function and chronic kidney disease**. Köttgen A et al. Nat Genet. 2009 Jun;41(6):712-7. doi: 10.1038/ng.377. Epub 2009 May 10
5	**Genetic loci influencing kidney function and chronic kidney disease** Chambers JC et al. Nat Genet. 2010 May;42(5):373-5. doi: 10.1038/ng.566. Epub 2010 Apr 11
6	**Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations** Okada Y et al. Nat Genet. 2012 Jul 15;44(8):904-9. doi: 10.1038/ng.2352
7	**Integration of genome-wide association studies with biological knowledge identifies six novel genes related to kidney function** Chasman DI et al. Hum Mol Genet. 2012 Dec 15;21(24):5329-43. doi: 10.1093/hmg/dds369. Epub 2012 Sep 8.
8	**Candidate genes for non-diabetic ESRD in African Americans: a genome-wide association study using pooled DNA** Bostrom MA et al. Hum Genet. 2010 Aug;128(2):195-204. doi: 10.1007/s00439-010-0842-3. Epub 2010 Jun 8
9	**New susceptibility loci associated with kidney disease in type 1 diabetes** Sandholm N et al. PLoS Genet. 2012 Sep;8(9):e1002921. doi: 10.1371/journal.pgen.1002921. Epub 2012 Sep 20
10	**Promoter polymorphism of the erythropoietin gene in severe diabetic eye and kidney complications** Tong Z et al. Proc Natl Acad Sci U S A. 2008 May 13;105(19):6998-7003. doi: 10.1073/pnas.0800454105. Epub 2008 May 5

We hypothesized that polymorphisms associated with renal disease will influence the expression of nearby transcript levels in the kidney, by altering transcriptional factor binding. Reports from the ENCODE project indicate that the majority (70%–80%) of the gene regulatory elements (promoters, enhancers, and insulators) are within 250 kbp of the gene (35). Using these criteria, we identified 306 genes within 500 kbp of 44 CKD SNPs. There was no gene within the 500-kbp window around the rs12437854 SNP; therefore, 43 loci were examined. We called these transcripts CKD risk associated transcripts (CRATs). (Original Article I.)

6.2. Kidney-specific expression of CRATs

First, we investigated whether these CRATs near the CKD risk associated loci have kidney-specific expression. We hypothesized that these CRATs play important role in kidney function, therefore they should be expressed in the kidney. We used comprehensive RNA sequencing analysis in normal, healthy kidney tubule samples (n=2) to determine the baseline expression of the 306 CRATs.

We found that 41% of the CKD risk loci-associated transcripts showed high (upper quartile), 32% showed medium, 21% showed low expression and only 6% of the transcripts were not detectable in healthy human kidney tubule samples (*Figure 3A*). Overall, we found that a large percentage of the CKD SNP neighboring transcripts (94%; 287 of 306) were expressed in the human kidney. We compared the kidney specific expression of the genes near CKD risk associated loci with the expression of other genes near 44 randomly selected loci where the nearby genes were defined similarly, with a 500-kbp window. Around the randomly selected loci, only 13% of the nearby transcripts showed high expression and 16% of the nearby transcripts were not expressed in the kidney, indicating that CRATs have a statistically significant kidney-specific enrichment (Chi-square test, $P=1.25 \times 10^{-9}$). (*Figure 3B*)

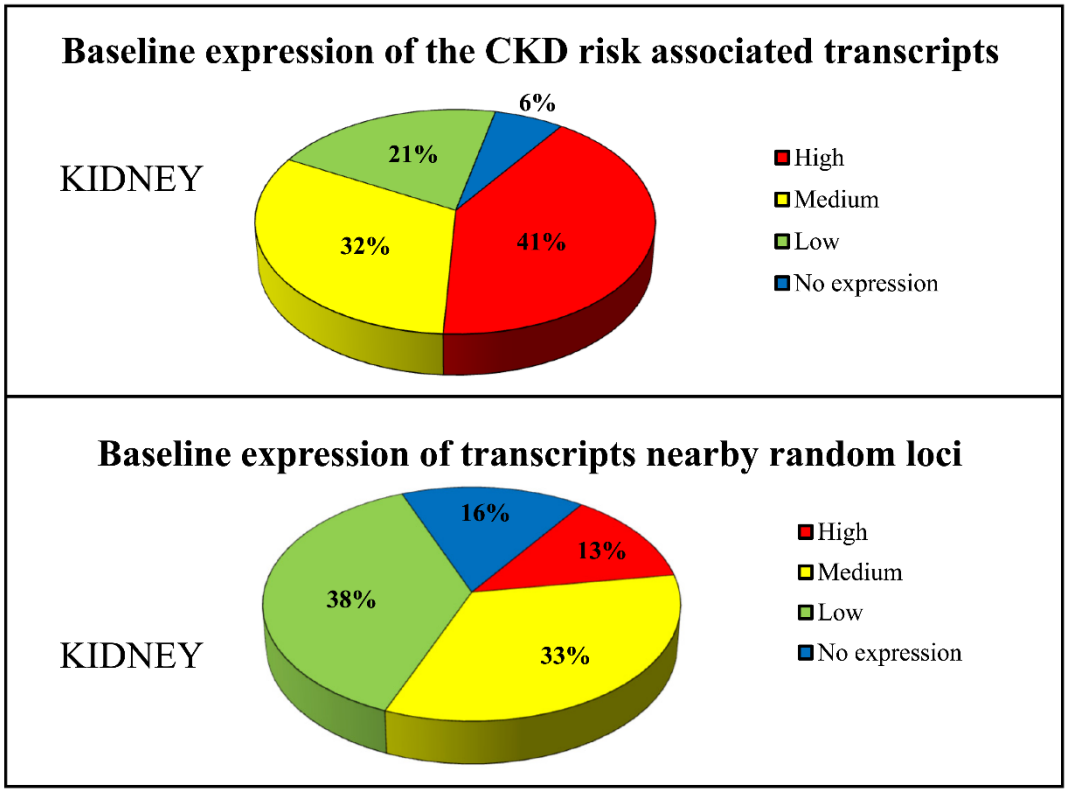
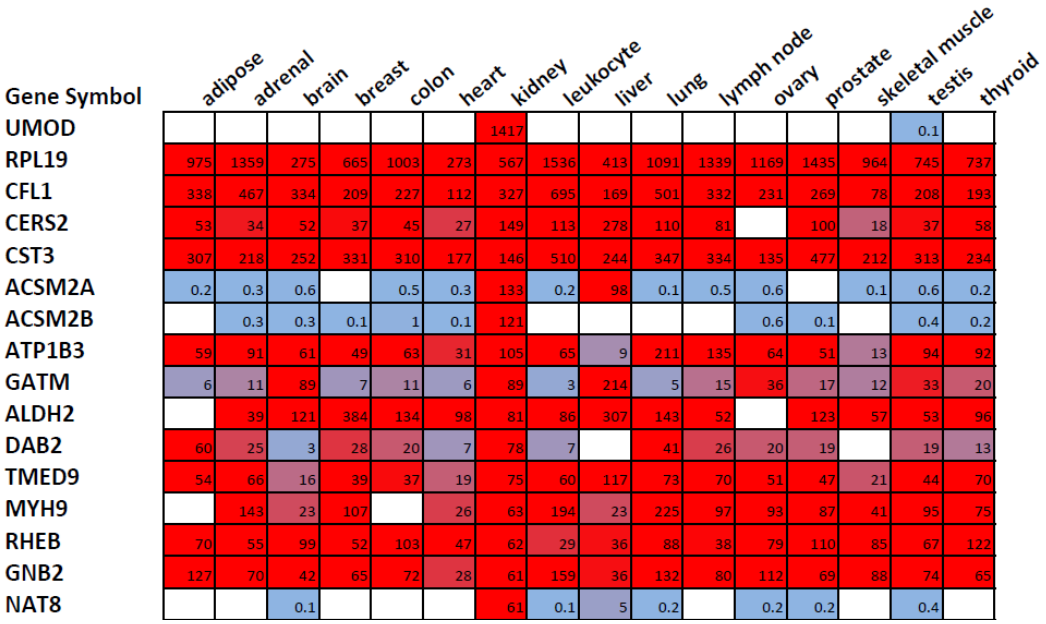


Figure 3. Baseline expression of the CKD risk loci associated transcripts in kidney compared to transcripts around random loci

Human kidney RNA sequencing data was used to examine the baseline expression of the transcripts. 41 % of the CKD risk loci neighboring transcripts are highly expressed in the kidney (red), 32 % showed medium expression (yellow), 21 % showed low expression in the kidney (green). 6% of the transcripts are not expressed (blue) (**A**).
Human kidney RNA sequencing data was used to examine the baseline expression of the transcripts around 44 randomly selected loci. 13 % of the CKD risk loci neighboring transcripts are highly expressed in the kidney (red), 33 % showed medium expression (yellow), 38 % showed low expression in the kidney (green). 16% of the transcripts are not expressed (blue) (**B**).

To confirm our results, we used publicly available datasets to examine the kidney-specific expression of CRATs. We performed gene ontology analysis (david.abcc.ncifcrf.gov), which indicated that CRATs have specific and significant enrichment in the kidney and peripheral leukocytes (P=0.0082 and P=0.0014, respectively). Additionally, we compared absolute expression levels of CRATs by RNA sequencing in 16 different human organs using the Illumina Body Map database (www.ebi.ac.uk). The atlas confirmed the statistically significant kidney-specific expression enrichment of CRATs. The atlas highlighted the high and kidney-specific

expression of Uromodulin (*UMOD*) (*Figure 4.*). In summary, expression of CRATs was enriched in the kidney and peripheral lymphocytes, potentially indicating the role of these cells in kidney disease development. (Original Article I.)



Gene Symbol	adipose	adrenal	brain	breast	colon	heart	kidney	leukocyte	liver	lung	lymph node	ovary	prostate	skeletal muscle	testis	thyroid
UMOD							1417								0.1	
RPL19	975	1359	275	665	1003	273	567	1536	413	1091	1339	1169	1435	964	745	737
CFL1	338	467	334	209	227	112	327	695	169	501	332	231	269	78	208	193
CERS2	53	34	52	37	45	27	149	113	278	110	81		100	18	37	58
CST3	307	218	252	331	310	177	146	510	244	347	334	135	477	212	313	234
ACSM2A	0.2	0.3	0.6		0.5	0.3	133	0.2	98	0.1	0.5	0.6		0.1	0.6	0.2
ACSM2B		0.3	0.3	0.1	1	0.1	121					0.6	0.1		0.4	0.2
ATP1B3	59	91	61	49	63	31	105	65	9	211	135	64	51	13	94	92
GATM	6	11	89	7	11	6	89	3	214	5	15	36	17	12	33	20
ALDH2		39	121	384	134	98	81	86	307	143	52		123	57	53	96
DAB2	60	25	3	28	20	7	78	7		41	26	20	19		19	13
TMED9	54	66	16	39	37	19	75	60	117	73	70	51	47	21	44	70
MYH9		143	23	107		26	63	194	23	225	97	93	87	41	95	75
RHEB	70	55	99	52	103	47	62	29	36	88	38	79	110	85	67	122
GNB2	127	70	42	65	72	28	61	159	36	132	80	112	69	88	74	65
NAT8			0.1				61	0.1	5	0.2		0.2	0.2		0.4	

Figure 4. Comprehensive RNA sequencing map of the CKD risk associated transcripts in 16 different human tissues

Expression levels of 306 CRATs were examined using the Illumina Body Map (www.ebi.ac.uk). The relative expression in each tissue of the first 16 genes with the highest expression level in the kidney is shown. High expression values (> 90 percentile) are marked red, low expression values (<10 percentile) marked blue, medium expression values marked as different tones of purple according to their expression value. Expression levels with FPKM values (Fragments Per Kilobase of transcript per Million mapped reads) lower than 0.1 are white. Gene symbols are official symbols approved by the Human Genome Organization Gene Nomenclature Committee (HGNC).

6.3. Expression profile of CRATs in normal and disease human kidney samples

According to our hypothesis, the functionally important CRATs are not only expressed in relevant cell types (e.g. kidney tubule cells, leukocytes) but the expression levels of these CRATs should change in CKD. To test this hypothesis, we analyzed gene expression levels in a large collection of microdissected human glomerular (n=51) and tubule (n=95) samples. Transcript profiling was performed for each individual sample

using Affymetrix U133v2 microarrays, and they contained probe set identifications for 226 transcripts from the 306 originals CRATs. (Original Article I.)

6.3.1. Expression profile of CRATs in glomerular samples

The expression of 226 CKD risk associated transcripts were analyzed in 51 human glomerular samples: 27 samples with eGFR above 60 ml/min/1.73 m^2 and 24 samples with reduced eGFR (eGFR < 60 ml/min/1.73 m^2). (*Table 2.a*) This eGFR cutoff value was used in GWASs too, to distinguish "cases" and "controls". To match our dataset to the cases of the GWASs, we included CKD samples from diabetic and hypertensive nephropathy.

First, following the method of the GWA studies, we compared the expression levels of the CKD loci neighboring transcripts in two groups: CKD cases and controls. In glomerulus samples, there were no differentially expressed CRATs in these two groups which were statistically significant.

However, linear correlation analysis identified the significant association of 34 CRATs with eGFR ($P < 0.05$) (*Table 4.*). The correlation between the expression of seven CRATs and eGFR remained significant, even after Benjamini–Hochberg-based multiple testing correction. The expression of multiple novel transcripts showed correlation with kidney function. For example, expression levels of Family with sequence similarity 47, member E (*FAM47E*), Plexin domain-containing 1 (*PLXDC1*), Vascular endothelial growth factor A (*VEGFA*) and Membrane-associated guanylate kinase 2 (*MAGI2*) correlated with eGFR. (*Figure 5.*) Immunostaining studies from the Human Protein Atlas (*www.proteinatlas.org*) showed that in non-disease human kidney tissue, proteins coded by *FAM47E* and *VEGFA* were highly expressed in glomeruli. The protein encoded by *PLXDC1* (also known as Tumor endothelial marker 7 [*TEM7*]) has glomerular endothelial-specific expression in normal human kidney tissue, while *MAGI2* seems to have a podocyte-specific expression pattern, potentially indicating its role in this cell type. Interestingly, *FAM47E*, *PLXDC1*, and *MAGI2* have not been identified in GWASs as potential causal or target genes in the vicinity of CKD risk loci. In summary, the analysis highlighted that the expression of several CRATs in glomeruli correlates with renal function.

Table 4. Expression levels of 34 transcripts (CRATs) in glomeruli showed significant correlation with eGFR

Pearson product moment correlation coefficient (Pearson R) was used to measure the strength of association between gene expression and eGFR. Two-tailed test was used to determine the statistical significance. With Benjamini-Hochberg multiple-testing correction, 7 transcripts showed significant correlation with eGFR (P corrected<0.05). Gene symbols are official symbols approved by the Human Genome Organization Gene Nomenclature Committee (HGNC).

Table 4. eGFR correlating CRATs in Glomeruli				
Gene symbol	**Pearson R**	**95% confidence interval**	**P (two-tailed)**	**P (corrected)**
CTSS	-0.501	-0.682 to -0.261	1.8×10^{-4}	***0.0426***
FAM47E///STBD1	0.496	0.256 to 0.679	2.1×10^{-4}	***0.0426***
FYB	-0.478	-0.666 to -0.233	3.9×10^{-4}	***0.0427***
LTB	-0.471	-0.661 to -0.225	4.8×10^{-4}	***0.0427***
EHBP1L1	-0.466	-0.657 to -0.219	5.7×10^{-4}	***0.0427***
MFAP4	-0.462	-0.654 to -0.214	6.4×10^{-4}	***0.0427***
MICALL2	-0.457	-0.651 to -0.208	7.5×10^{-4}	***0.0428***
CTSK	-0.428	-0.629 to -0.173	1.7×10^{-3}	0.077
PLXDC1	-0.417	-0.621 to -0.160	2.3×10^{-3}	0.085
F12	-0.390	-0.601 to -0.129	4.6×10^{-3}	0.154
VEGFA	0.372	0.107 to 0.587	7.2×10^{-3}	0.222
PCOLCE	-0.369	-0.585 to -0.104	7.8×10^{-3}	0.222
MYCN	0.364	0.098 to 0.581	8.6×10^{-3}	0.231
GP2	0.356	0.090 to 0.575	***0.010***	0.235
SLC34A1	0.356	0.089 to 0.575	***0.010***	0.235
EFHD2	-0.355	-0.574 to -0.088	***0.011***	0.235
LST1	-0.352	-0.572 to -0.085	***0.011***	0.236
MICB	-0.345	-0.566 to -0.076	***0.013***	0.25
LRCH4	-0.336	-0.560 to -0.067	***0.016***	0.275
ANXA9	0.334	0.065 to 0.558	***0.017***	0.275
MAGI2	0.321	0.049 to 0.548	***0.022***	0.336
ACSM5	0.313	0.041 to 0.54	***0.025***	0.375
SLC22A3	-0.308	-0.538 to -0.035	***0.028***	0.387
PSRC1	-0.304	-0.535 to -0.032	***0.030***	0.398
UMOD	0.301	0.027 to 0.532	***0.032***	0.401
ALDH3A2	0.298	0.024 to 0.530	***0.034***	0.401
GATM	0.297	0.023 to 0.529	***0.035***	0.401
SORT1	0.294	0.020 to 0.527	***0.036***	0.401

AGMAT	0.293	0.019 to 0.526	***0.037***	0.401
DRAP1	0.293	0.018 to 0.526	***0.037***	0.401
CASP9	0.289	0.014 to 0.523	***0.040***	0.401
KDM5A	-0.287	-0.521 to -0.012	***0.041***	0.401
SLC6A13	0.286	0.011 to 0.520	***0.042***	0.401
HLA-C	-0.284	-0.519 to -0.009	***0.044***	0.406

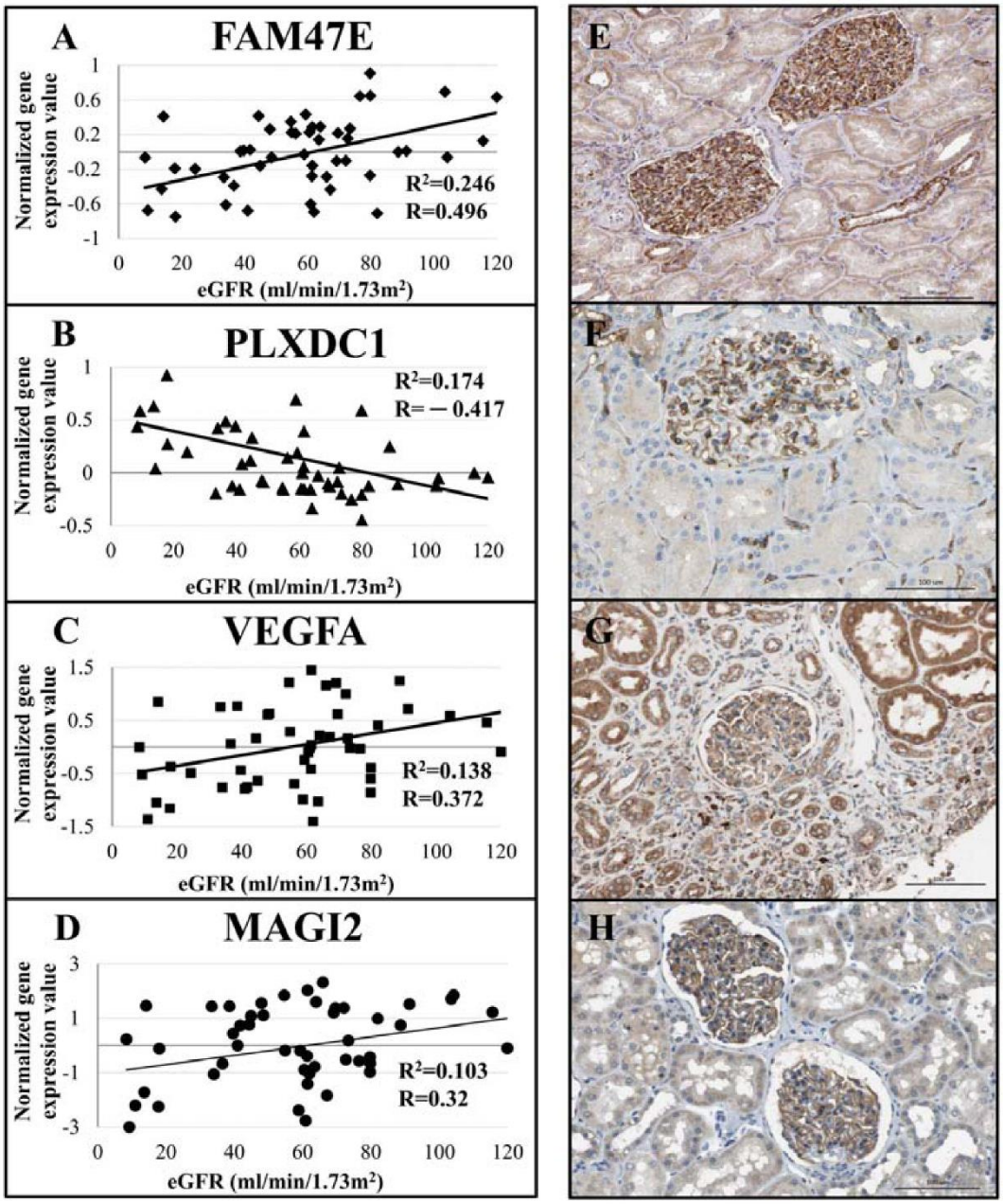


Figure 5. Correlation between CRAT expression in glomeruli and renal function

The y-axis shows the relative normalized glomerular expressions of *FAM47E* (family with sequence similarity 47, member E) (A), *PLXDC1* (plexin domain containing 1) (B), *VEGFA* (vascular endothelial growth factor A) (C) and *MAGI2* (membrane associated guanylate kinase) (D). The x-axis shows the eGFR (estimated glomerular filtration rate) for each sample. Each dot represents one individual miscrodissected glomerular sample. Immunohistochemistry shows the protein expression in human glomeruli. (FAM47E (E), PLXDC1 (F), VEGFA (G), MAGI2 (H)) Scale bar: 100 μm. (Source: www.proteinatlas.org)

6.3.2. Expression profile of CRATs in tubule samples

We also examined the expression of the 226 transcripts in 95 tubule samples from patients with normal and diseased renal function (*Table 2.b*). Samples were obtained from patients with a wide range of kidney function: 56 samples with normal eGFR (eGFR > 60 ml/min/1.73m^2) and 39 samples with kidney disease (eGFR < 60 ml/min/1.73 m^2).

Again, we performed a binary analysis by comparing the expression levels of CRATs in control versus CKD samples (eGFR cutoff value: 60 ml/min/1.73m^2). Using statistical correction for multiple testing (Benjamini–Hochberg corrected P value < 0.05), 73 CRATs showed differential expression when CKD tubule samples were compared to controls. (*Table 5*.)

Table 5. List of CRATs showing differential expression in control vs. CKD tubule samples

In tubules, 73 transcripts in the neighborhood of the CKD risk loci showed significant differences when CKD samples are compared to controls. Benjamini-Hochberg multiple-testing correction was used with a P value < 0.05. Gene symbols are official symbols approved by the Human Genome Organization Gene Nomenclature Committee (HGNC).

Table 5. Differentially expressed CRATs in chronic kidney disease					
Gene Symbol	**P value (corrected)**	**Regulation**	**Gene Symbol**	**P value (corrected)**	**Regulation**
LST1	6.40×10^{-8}	up	***PLXDC1***	3.63×10^{-3}	up
SLC7A9	4.56×10^{-7}	down	***SLC30A4***	4.33×10^{-3}	down
ALDH3A2	1.89×10^{-6}	down	***GATM***	5.84×10^{-3}	down
SLC34A1	2.04×10^{-6}	down	***PGAP3***	6.17×10^{-3}	down
CTSS	4.19×10^{-6}	up	***SLC6A12***	6.81×10^{-3}	down
FYB	4.19×10^{-6}	up	***IGF2R***	7.18×10^{-3}	down
ACSM5	1.85×10^{-5}	down	***MICALL2***	8.09×10^{-3}	up
LTB	2.55×10^{-5}	up	***CTSK***	0.011	up
UMOD	2.70×10^{-5}	down	***DDI2///RSC1A1***	0.012	down
ACSM2A/ ACSM2B	3.81×10^{-5}	down	***ATXN2***	0.013	down
SLC47A1	4.26×10^{-5}	down	***CCDC85B***	0.014	up
ANXA9	4.54×10^{-5}	down	***HLA-C***	0.014	up
DNAJC16	1.24×10^{-4}	down	***TFDP2***	0.015	down
NAT8B	1.63×10^{-4}	down	***AIF1***	0.018	up

ACAD10	1.97 x 10^{-4}	down	***DDX1***	0.018	down
GSTM4	2.55 x 10^{-4}	down	***PRUNE***	0.018	down
VEGFA	4.05 x 10^{-4}	down	***MFAP4***	0.018	up
CTSW	5.04 x 10^{-4}	up	***DBN1***	0.021	up
NAT8/ NAT8B	5.04 x 10^{-4}	down	***CELA2A/// CELA2B***	0.024	up
FAM89B	6.09 x 10^{-4}	up	***DACH1***	0.024	down
AFF3	6.28 x 10^{-4}	up	***TBX2***	0.024	down
MYCN	6.28 x 10^{-4}	down	***ERBB2***	0.027	down
ALDH2	6.64 x 10^{-4}	down	***GP2***	0.030	down
FAM47E/ STBD1	6.70 x 10^{-4}	down	***F12***	0.031	up
GNAI3	6.87 x 10^{-4}	up	***PHTF2***	0.031	up
SLC22A2	8.54 x 10^{-4}	down	***CDC42SE1***	0.033	up
DAB2	1.21 x 10^{-3}	down	***LARP4B***	0.033	down
STC1	1.59 x 10^{-3}	down	***PTPN12***	0.035	up
APOM	1.83 x 10^{-3}	down	***PDLIM7***	0.036	up
GPER	1.93 x 10^{-3}	down	***IDI1***	0.037	down
SLC22A1	1.96 x 10^{-3}	down	***BRAP***	0.039	down
AGMAT	2.18 x 10^{-3}	down	***DUOX1***	0.045	up
EHBP1L1	2.47 x 10^{-3}	up	***FIBP***	0.047	down
SLC6A13	2.47 x 10^{-3}	down	***MPPED2***	0.048	down
FAM193B	2.66 x 10^{-3}	up	***MYH9***	0.048	up
CERS2	3.14 x 10^{-3}	down	***WDR37***	0.049	down
LRCH4	3.38 x 10^{-3}	up			

Besides comparing the expression levels in the two groups, we were searching for stronger associations between renal function and the expressions of the transcripts nearby CKD risk loci. We performed linear correlation tests between the gene expression arrays and eGFR in tubule samples. Pearson correlation identified 92 transcripts with statistically significant ($P < 0.05$) linear correlation with kidney function (*Table 6.*). The correlation between the expression of 70 CRATs and eGFR remained significant, even after Benjamini–Hochberg-based multiple testing correction.

Table 6. In tubules, expression levels of 92 transcripts (CRATs) showed significant correlation with eGFR

Pearson product moment correlation coefficient (Pearson R) was used to measure the strength of association between gene expression and eGFR. Two-tailed test was used to

determine the statistical significance. With Benjamini-Hochberg multiple-testing correction, 70 transcripts showed significant correlation with eGFR (P corrected<0.05). Asterisks (*) indicate when the gene expression significantly correlated with eGFR in the external validation microarray dataset containing 41 tubule samples, double asterisks (**) indicate when this correlation in the external data set remained significant after the multiple-testing correction. Gene symbols are official symbols approved by the Human Genome Organization Gene Nomenclature Committee (HGNC).

Table 6. eGFR correlating CRATs in Tubules

Gene Symbol	Pearson R	95% confidence interval	P (two-tailed)	P (corrected)
SLC34A1 **	0.610	0.466 to 0.723	5.3×10^{-11}	$\mathbf{2.1 \times 10^{-8}}$
SLC7A9 **	0.588	0.439 to 0.706	3.6×10^{-10}	$\mathbf{7 \times 10^{-8}}$
ACSM5 **	0.551	0.393 to 0.677	7.3×10^{-9}	$\mathbf{9.6 \times 10^{-7}}$
FYB **	-0.531	-0.662 to -0.370	3×10^{-8}	$\mathbf{2.4 \times 10^{-6}}$
ACSM2A/ACSM2B *	0.526	0.363 to 0.658	4.3×10^{-8}	$\mathbf{2.4 \times 10^{-6}}$
NAT8B *	0.518	0.354 to 0.652	7.4×10^{-8}	$\mathbf{3.3 \times 10^{-6}}$
ALDH3A2 **	0.517	0.352 to 0.651	8.3×10^{-8}	$\mathbf{3.3 \times 10^{-6}}$
LTB **	-0.517	-0.651 to -0.352	8.3×10^{-8}	$\mathbf{3.3 \times 10^{-6}}$
LST1 **	-0.514	-0.646 to -0.351	1×10^{-7}	$\mathbf{3.6 \times 10^{-6}}$
UMOD **	0.504	0.337 to 0.641	1.9×10^{-7}	$\mathbf{6.1 \times 10^{-6}}$
ACAD10 *	0.485	0.314 to 0.625	6.5×10^{-7}	$\mathbf{1.8 \times 10^{-5}}$
DNAJC16	0.478	0.306 to 0.620	9.7×10^{-7}	$\mathbf{2.5 \times 10^{-5}}$
GSTM4 **	0.474	0.301 to 0.617	1.2×10^{-6}	$\mathbf{3 \times 10^{-5}}$
SLC6A13 *	0.469	0.295 to 0.613	1.7×10^{-6}	$\mathbf{3.3 \times 10^{-5}}$
VEGFA **	0.468	0.294 to 0.612	1.7×10^{-6}	$\mathbf{3.3 \times 10^{-5}}$
CTSS **	-0.468	-0.612 to -0.294	1.8×10^{-6}	$\mathbf{3.3 \times 10^{-5}}$
ANXA9 **	0.464	0.289 to 0.609	2.2×10^{-6}	$\mathbf{3.9 \times 10^{-5}}$
SLC6A12	0.454	0.278 to 0.601	3.8×10^{-6}	$\mathbf{6 \times 10^{-5}}$
FAM47E/ STBD1 **	0.444	0.267 to 0.593	6.5×10^{-6}	$\mathbf{9.5 \times 10^{-5}}$
SLC47A1 **	0.444	0.266 to 0.593	6.5×10^{-6}	$\mathbf{9.5 \times 10^{-5}}$
NAT8/ NAT8B *	0.441	0.263 to 0.5901	7.7×10^{-6}	$\mathbf{1.1 \times 10^{-4}}$
ALDH2 *	0.440	0.262 to 0.589	8.1×10^{-6}	$\mathbf{1.1 \times 10^{-4}}$
CERS2	0.436	0.257 to 0.586	1×10^{-5}	$\mathbf{1.2 \times 10^{-4}}$
STC1 *	0.431	0.251 to 0.582	1.3×10^{-5}	$\mathbf{1.6 \times 10^{-4}}$
APOM **	0.428	0.248 to 0.579	1.5×10^{-5}	$\mathbf{1.8 \times 10^{-4}}$
DAB2 **	0.410	0.227 to 0.565	3.7×10^{-5}	$\mathbf{3.9 \times 10^{-4}}$
AGMAT **	0.397	0.213 to 0.555	6.7×10^{-5}	$\mathbf{6.9 \times 10^{-4}}$
GATM **	0.393	0.209 to 0.551	8×10^{-5}	$\mathbf{7.8 \times 10^{-4}}$

SLC22A2	0.393	0.208 to 0.551	8.3×10^{-5}	$\mathbf{7.9 \times 10^{-4}}$
FAM89B *	-0.391	-0.549 to -0.206	8.9×10^{-5}	$\mathbf{8.1 \times 10^{-4}}$
SLC30A4	0.382	0.196 to 0.542	1×10^{-4}	$\mathbf{1.1 \times 10^{-3}}$
MYCN **	0.376	0.188 to 0.537	2×10^{-4}	$\mathbf{1.4 \times 10^{-3}}$
AIF1 *	-0.352	-0.517 to -0.163	5×10^{-4}	$\mathbf{3.4 \times 10^{-3}}$
TRIP6 *	0.352	0.162 to 0.517	5×10^{-4}	$\mathbf{3.5 \times 10^{-3}}$
LARP4B	0.347	0.157 to 0.513	6×10^{-4}	$\mathbf{4 \times 10^{-3}}$
GPER **	0.347	0.157 to 0.513	6×10^{-4}	$\mathbf{4 \times 10^{-3}}$
LRCH4	-0.342	-0.509 to -0.151	7×10^{-4}	$\mathbf{4.7 \times 10^{-3}}$
SLC22A1	0.342	0.151 to 0.508	7×10^{-4}	$\mathbf{4.7 \times 10^{-3}}$
FAM193B	-0.338	-0.505 to -0.146	8×10^{-4}	$\mathbf{5.4 \times 10^{-3}}$
CTSK **	-0.335	-0.502 to -0.143	9×10^{-4}	$\mathbf{5.9 \times 10^{-3}}$
IGF2R	0.321	0.128 to 0.491	1.5×10^{-3}	$\mathbf{9.5 \times 10^{-3}}$
DDI2/ RSC1A1	0.320	0.126 to 0.490	1.6×10^{-3}	$\mathbf{9.9 \times 10^{-3}}$
PLXDC1 **	-0.315	-0.486 to -0.121	1.9×10^{-3}	***0.012***
TFDP2 *	0.313	0.119 to 0.484	2×10^{-3}	***0.012***
BAG6	0.311	0.117 to 0.483	2.1×10^{-3}	***0.013***
EHBP1L1 **	-0.311	-0.483 to -0.117	2.1×10^{-3}	***0.013***
CTSW	-0.309	-0.481 to -0.115	2.3×10^{-3}	***0.013***
ATXN2	0.309	0.115 to 0.481	2.3×10^{-3}	***0.013***
ERBB2 *	0.308	0.113 to 0.479	2.4×10^{-3}	***0.013***
PHTF2 *	-0.307	-0.479 to -0.112	2.5×10^{-3}	***0.013***
TBX2	0.305	0.110 to 0.477	2.7×10^{-3}	***0.014***
MICB **	-0.303	-0.476 to -0.108	2.8×10^{-3}	***0.014***
SIPA1	-0.298	-0.472 to -0.103	3.3×10^{-3}	***0.017***
PRUNE	0.298	0.102 to 0.471	3.4×10^{-3}	***0.017***
CCDC85B *	-0.287	-0.460 to -0.094	4.9×10^{-3}	***0.023***
GRK6	-0.287	-0.461 to -0.090	4.9×10^{-3}	***0.023***
PIP5K1A	0.278	0.082 to 0.455	6.3×10^{-3}	***0.029***
LTBP3 *	-0.276	-0.452 to -0.079	6.9×10^{-3}	***0.032***
CASP9 *	0.273	0.075 to 0.449	7.5×10^{-3}	***0.034***
NSD1	0.272	0.074 to 0.449	7.7×10^{-3}	***0.034***
SOX11	-0.271	-0.448 to -0.074	7.8×10^{-3}	***0.034***
DUOX1	-0.269	-0.447 to -0.072	8.4×10^{-3}	***0.036***
AFF3	-0.267	-0.445 to -0.070	8.8×10^{-3}	***0.036***
NCR3	-0.267	-0.445 to -0.070	8.8×10^{-3}	***0.036***
FIBP	0.265	0.067 to 0.443	9.6×10^{-3}	***0.039***
GNAI3	-0.263	-0.441 to -0.065	*0.010*	***0.040***
FOSL1	0.261	0.063 to 0.439	*0.011*	***0.042***
MPPED2	0.259	0.061 to 0.438	*0.011*	***0.044***
MFAP4 **	-0.258	-0.437 to -0.059	*0.012*	***0.045***
B9D1	-0.254	-0.434 to -0.056	*0.013*	***0.049***
SLC12A9	0.254	0.055 to 0.433	*0.013*	0.050

CFL1	-0.250	-0.430 to -0.051	*0.015*	0.055
CST8	-0.248	-0.428 to -0.050	*0.015*	0.057
DBN1 **	-0.242	-0.423 to -0.042	*0.018*	0.067
CDC42SE1 **	-0.242	-0.423 to -0.042	*0.018*	0.067
GP2 **	0.233	0.033 to 0.415	*0.023*	0.079
MYH9 *	-0.233	-0.415 to -0.033	*0.023*	0.079
IDI1	0.233	0.033 to 0.415	*0.023*	0.079
MLLT11 *	-0.233	-0.415 to -0.033	*0.023*	0.079
DACH1	0.229	0.029 to 0.412	*0.026*	0.087
DDX1	0.226	0.025 to 0.409	*0.028*	0.093
ACSM1	-0.225	-0.408 to -0.025	*0.028*	0.094
PDLIM7 *	-0.225	-0.408 to -0.024	*0.029*	0.094
RELA	-0.224	-0.407 to -0.023	*0.029*	0.096
SORT1	0.221	0.021 to 0.405	*0.031*	0.10
SLC28A2 **	0.221	0.020 to 0.405	*0.031*	0.10
EPN2 *	-0.213	-0.398 to -0.012	*0.038*	0.119
CELA2A/CELA2B **	-0.213	-0.397 to -0.012	*0.039*	0.120
CUX2	-0.212	-0.396 to -0.010	*0.040*	0.122
PTPN12 *	-0.211	-0.396 to -0.010	*0.040*	0.122
MICALL2 **	-0.206	-0.391 to -0.004	*0.046*	0.137
MKKS	0.204	0.003 to 0.389	*0.048*	0.141

More transcripts (58%) showed a positive correlation with renal function (their expression was decreased in samples with lower GFR), and 42% showed an inverse correlation. Renal function correlated with the expression of 25 CRATs both in glomeruli and tubules. Tubule-specific expression of solute carriers had the strongest correlation with renal function. For example, the levels of the Solute carrier family 34, member 1 (*SLC34A1*), which codes a type II sodium/phosphate cotransporter, and *SLC7A9* (Solute Carrier Family 7 Member 9), which codes the light chain of an amino acid transporter (*Figure 6. A, B*), correlated strongly with eGFR (with R values of 0.61 and 0.59, respectively). Based on the data of the Human Protein Atlas, both transcripts encode proteins that are highly and specifically expressed in renal tubule epithelial cells (*Figure 6. D, E*). In addition to solute carriers, the expression of a metabolic enzyme, the Acyl-CoA synthetase medium chain family member 5 (*ACSM5*), also highly correlated with renal function and showed high protein expression in tubule epithelial cells (*Figure 6. C, F*).

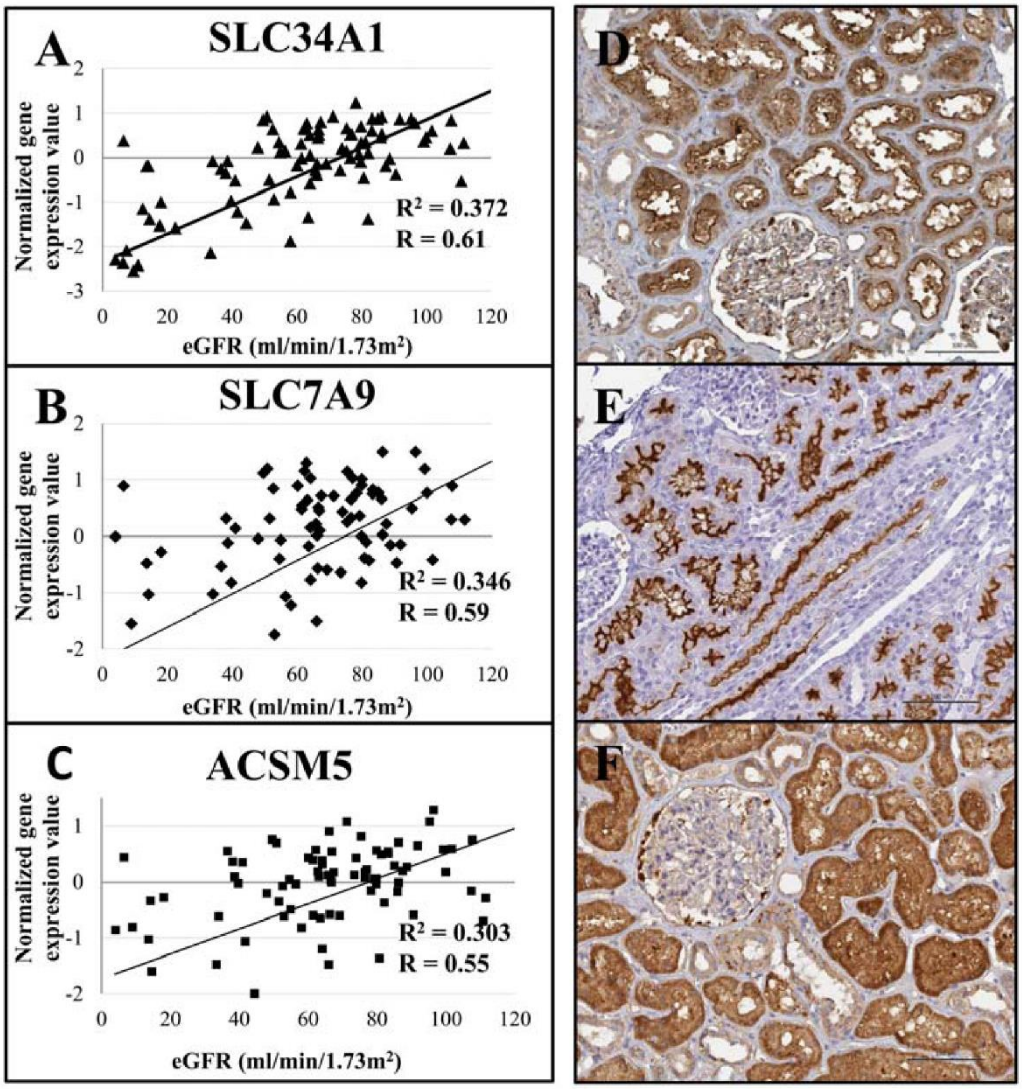


Figure 6. Correlation between CRAT expression in tubules and renal function

Expressions of *SLC34A1* (solute carrier family 34, member 1) (A), *SLC7A9* (solute carrier family 7, member 9) (B) and *ACSM5* (acyl-CoA synthetase medium chain family member 5) (C) correlate with eGFR (estimated glomerular filtration rate) in tubule samples. The x-axis represents eGFR (ml/min/1.73m^2), while the y-axis represents the normalized gene expression values of the transcript. Each dot represents transcript levels and eGFR values from a single kidney sample. The line is the fitted correlation value. Immunohistochemistry shows tubular specific expression of SLC34A1 (D), SLC7A9 (E) and ACSM5 (F). Scale bar: 100 μm. (Source: www.proteinatlas.org)

For external validation, we used a gene expression dataset containing genome-wide transcription profiling from 41 microdissected tubule samples (*Table 2.c*). The samples in this dataset were different from the primary dataset, and a slightly different method was used for microarray probe labeling (as described above). Although this

dataset was much smaller with a narrower eGFR range, we confirmed the significant linear correlation of 51 transcripts, highlighting the importance of these CRATs.

In summary, the gene expression and kidney function correlation analysis revealed CRATs for future prioritization.

6.4. Transcript levels around CKD risk associated loci

6.4.1. Transcript levels around UMOD locus

We specifically investigated expression changes of the *UMOD* transcript, because it is a potential causal or target gene underlying the polymorphism of some of the best characterized CKD associated loci on chromosome 16 (rs12917707, rs4293393 and rs11864909). This gene encodes one of the most abundant proteins in human urine: Uromodulin or Tamm–Horsfall protein. Furthermore, functional studies seem to link *UMOD* expression both as a biomarker and a causal gene for CKD development (37). We found that UMOD transcript levels showed a highly significant linear correlation with renal function (Pcorr=6.09 x 10^{-6}) in tubule samples (*Figure 7A*). We also performed immunohistochemistry staining from samples used for the transcriptomic analysis: samples with low *UMOD* mRNA (messenger Ribonucleic Acid) expression and with high expression. The results of the staining indicate excellent correlation between uromodulin protein expression and its transcript levels (*Figure 7. B-E*).

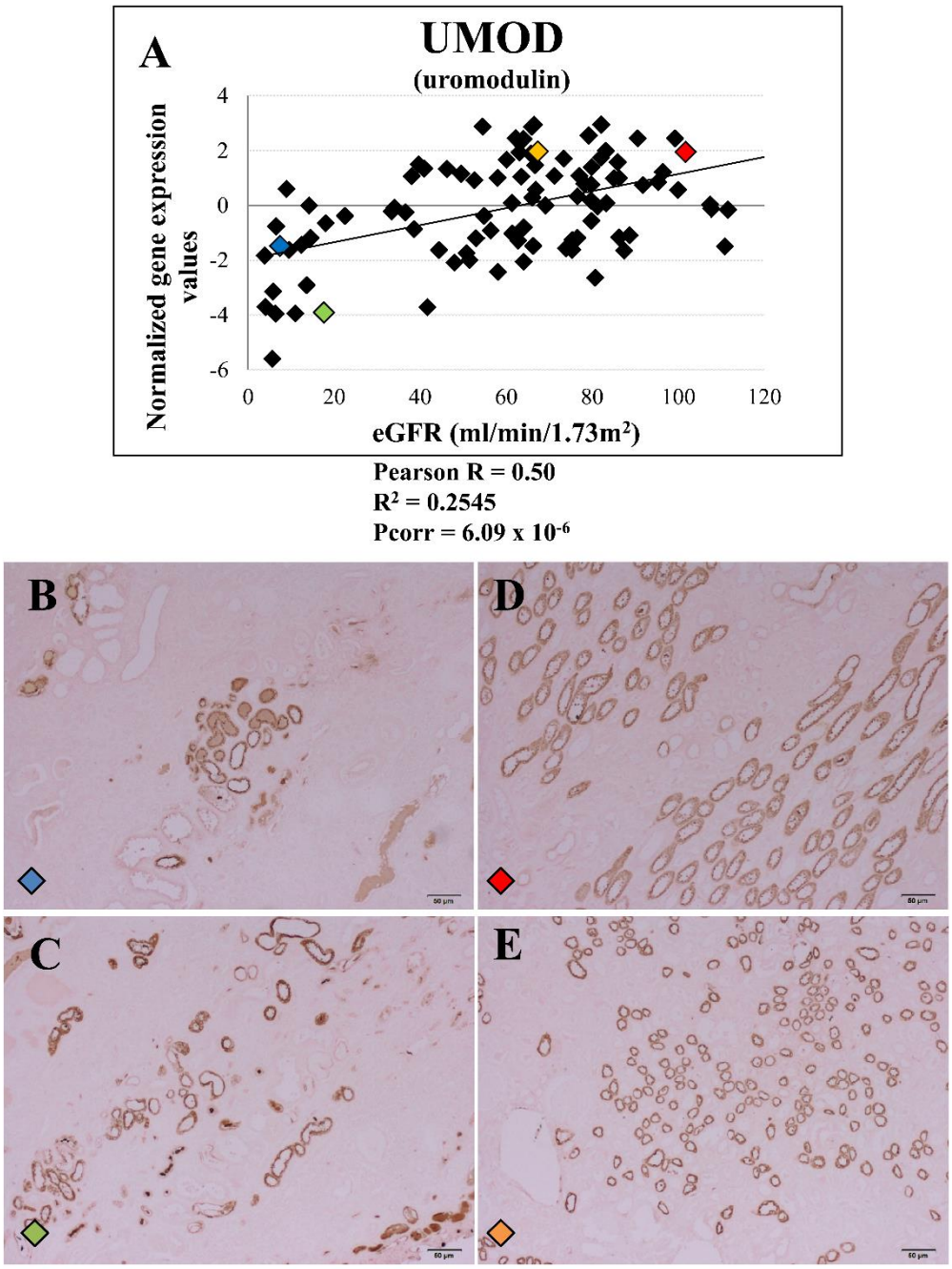


Figure 7. *UMOD* expression correlates with renal function

The expression of *UMOD* (uromodulin) (A) correlates with eGFR (estimated glomerular filtration rate) in tubule samples. The x-axis represents eGFR (ml/min/1.73m^2), while y-axis represents the normalized gene expression values of the transcript. Each dot represents transcript levels and GFR values from a single kidney sample. The line is the fitted correlation value. Immunohistochemistry of the samples with low and high mRNA expression showed differences of the uromodulin (B-E) expression on protein level. Scale bar: 50 μm.

Although *UMOD* has emerged as an important causal gene for CKD, unexpectedly, we found that three other nearby genes were also highly expressed in renal tubules, and their expression strongly correlated with eGFR. To illustrate this observation, *Figure 8* shows the "chromosome 16 locus", including three leading SNPs (rs12917707, rs4293393 and rs11864909) that best correlate with CKD. Closest genes to these polymorphisms are *UMOD* and *PDILT* (Protein disulfide isomerase-like, testis expressed). The expression level of *PDILT* was nearly undetectable in normal kidney samples by RNA sequencing, but the analysis showed high *UMOD* transcript levels in human kidney samples. We observed that *ACSM5* and *ACSM2A/B* were also highly expressed in human kidney tubule samples (*Figure 8A*). RNA sequencing data from the Illumina Body Map also confirmed the expression levels of these transcripts (*Figure 8B*). By linear correlation analysis, we found highly significant correlation between renal function and *UMOD*, *ACSM5* and *ACSM2A/2B* transcript levels (*Figures 7A*, *6C*, *9A*). Unfortunately, *PDILT* probes were absent from the human U133 chips, and therefore, the correlation between *PDILT* and renal function could not be analyzed by microarray. However, we also validated the transcript expression of *UMOD*, Glycoprotein 2 (*GP2*), *ACSM5*, *ACSM2A/B*, *ACSM1*, and *PDILT* by QRT-PCR (*Table 2.d*) to confirm the microarray results, and *PDILT* did not show correlation with renal function (*Figure 8C*).

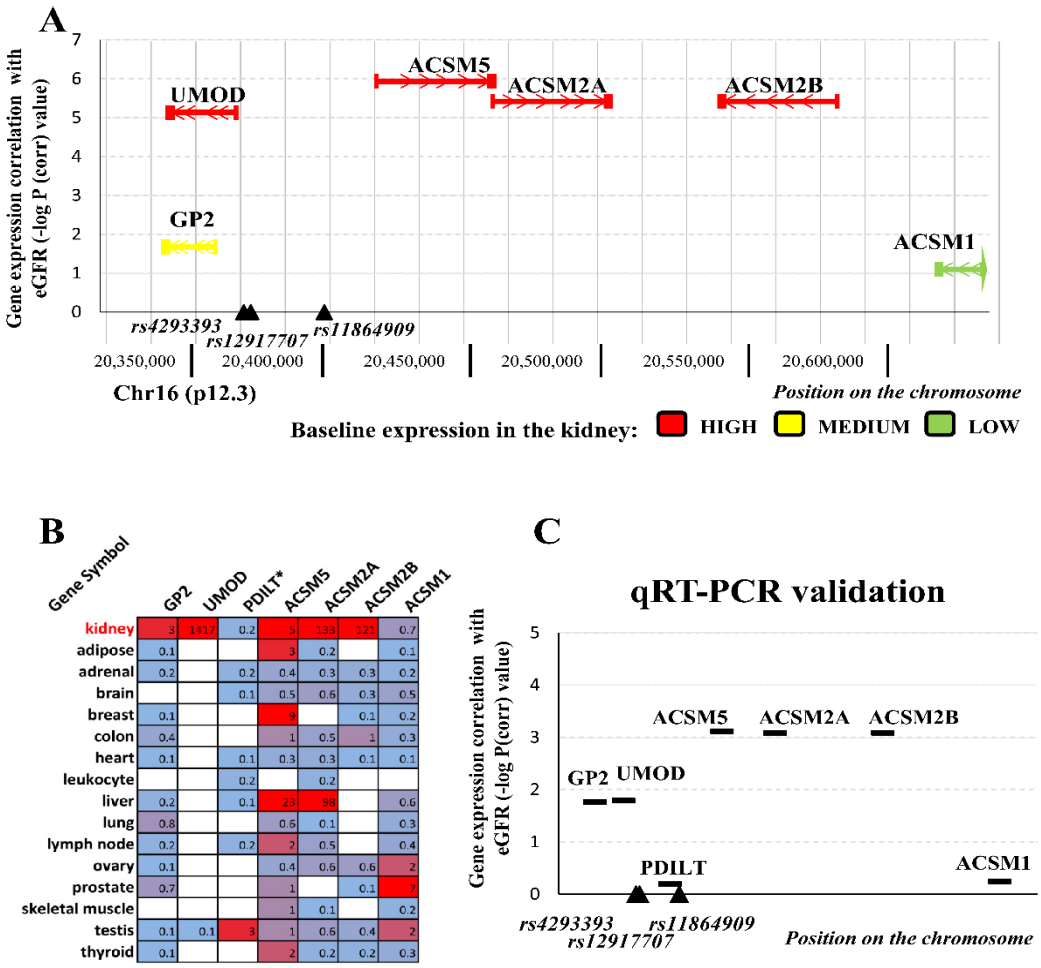


Figure 8. Correlation of tubule specific transcript levels with renal function near the UMOD locus (rs4293393, rs12917707 and rs11864909 polymorphisms)

The x-axis represents the genomic position of each gene on chromosome 16 (A, C). The y-axis represents the negative logarithm of the corrected p-value (significance) between the expression of each gene and eGFR (estimated glomerular filtration rate, ml/min/1.73m^2). Color coding represents the baseline expression of the transcripts in human kidney based on the RNA sequencing data. Red: high, yellow: medium, green: low expression in the kidney (A). Based on the results of the Illumina Body Map (www.ebi.ac.uk) a heat map was generated from the FPKM values of the CKD risk associated transcripts near these SNPs. High expression values (> 90 percentile) are marked red, low expression values (<10 percentile) marked blue. Expressions with FPKM values lower than 0.1 are marked white. Asterisks indicate genes without probe set IDs on the Affymetrix arrays (B). Quantitative real-time PCR (qRT-PCR) validation confirmed the significant correlation with eGFR of the following transcripts: *GP2* (glycoprotein 2), *UMOD* (uromodulin), *ACSM5, ACSM2A, ACSM2B* (acyl-CoA synthetase medium chain family members) (C). The chart on panel A shows a strong correlation between *UMOD* expression and eGFR, while the expressions of *ACSM5, ACSM2A/2B* also highly correlate with renal function.

We performed immunohistochemistry staining on samples with high and low *ACSM2A/2B* transcript expression and found the expression level changes on protein level as well (*Figure 9. B-E*). *ACSM5* and *ACSM2A/2B* genes (ACSM family members) encode three genes in fatty acid oxidation pathways. However, these transcripts showed high expression in the kidney and their expression strongly correlated with renal function, they were not mentioned before in the GWASs as potential causal or target genes. Our results indicate a potential functional role of these transcripts. (Original Article I.)

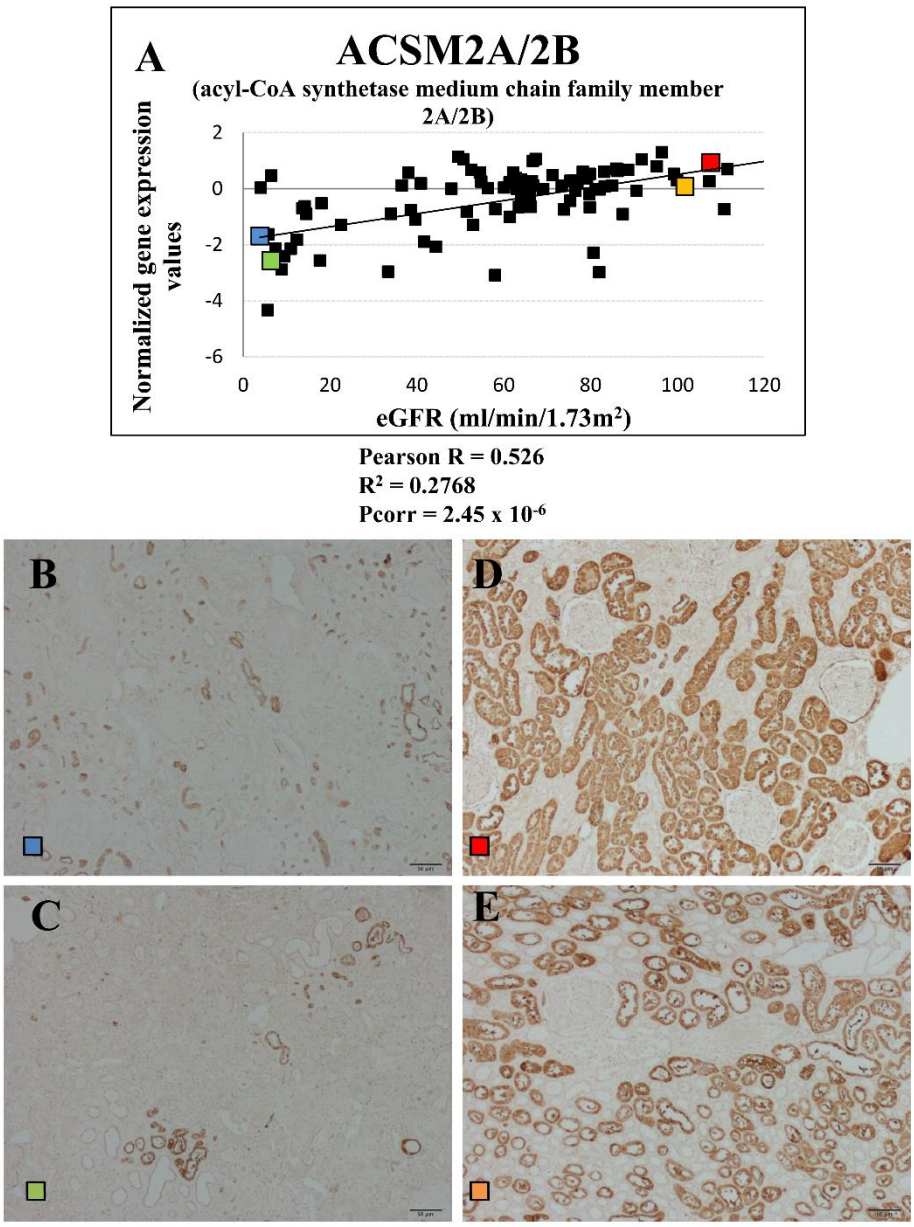


Figure 9. *ACSM2A* expression correlates with renal function

The expression of *ACSM2A* (acyl-CoA synthetase medium chain family member 2A) (A) correlates with eGFR (estimated glomerular filtration rate) in tubule samples. The x-axis represents eGFR (ml/min/1.73m^2), while y-axis represents the normalized gene expression values of the transcript. Each dot represents transcript levels and GFR values from a single kidney sample. The line is the fitted correlation value. Immunohistochemistry of the samples with low and high mRNA expression showed differences of the ACSM2A (B-E) expression on protein level. Scale bar: 50 μm.

6.4.2. Transcript levels around other CKD risk associated loci

We examined whether in the proximity of a single SNP changes in expression of a single gene or multiple genes could be observed. We found that, on 23 of 43 examined CKD risk loci, multiple neighboring transcripts correlated with renal function. For example, at the *SLC47A1* locus (rs2453580), not only the *SLC47A1* (Solute Carrier Family 47 Member 1, a multidrug extrusion protein) but also, the Aldehyde dehydrogenase 3 family, member A2 (*ALDH3A2*) correlated with eGFR ($P=8 \times 10^{-8}$). (*Figure 10.*)

We found that, around the rs267734 polymorphism on chromosome 1, both Ceramide synthase 2 (*CERS2*; $P=1.01 \times 10^{-5}$) and Annexin A9 (*ANXA9*) ($P=2.2 \times 10^{-6}$) transcripts correlated with eGFR in tubule samples. In addition, transcript level of Cathepsin S (*CTSS*) correlated with renal function both in tubules and glomeruli ($P=1.77 \times 10^{-6}$ and $P=1.8 \times 10^{-4}$, respectively) (*Figure 11.*)

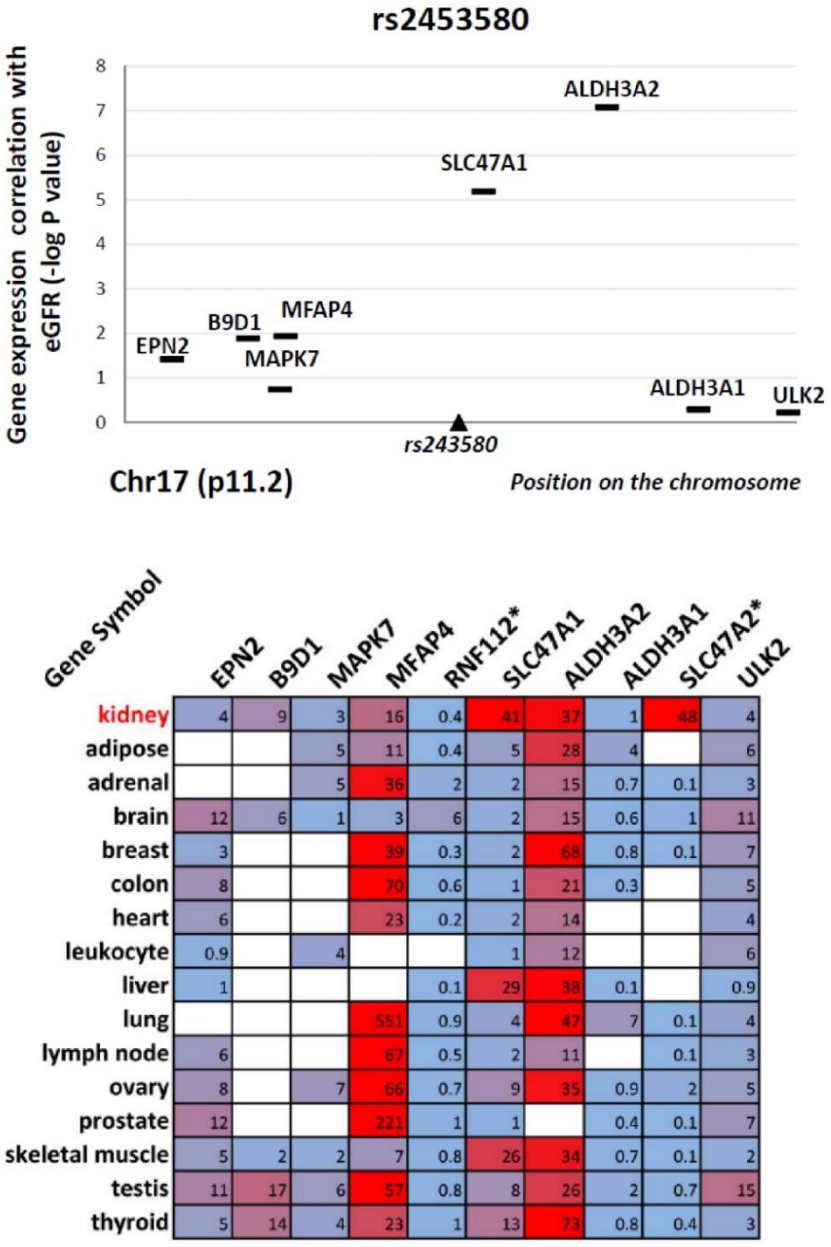


Figure 10. Transcript level expression and correlation of tubule specific transcript levels with renal function around rs2453580 locus

The x-axis represents the genomic position of each gene on chromosome 17. The y-axis represents the negative logarithm of the p-value (significance) between the expression of each gene and eGFR (estimated glomerular filtration rate, ml/min/1.73m^2). The lower panel of the chart represents the expression of transcripts within the 250 kbp vicinity of the CKD SNP in 16 human organs. Asterisks indicate genes without probe set IDs on the Affymetrix arrays. Gene symbols are official symbols approved by the Human Genome Organization Gene Nomenclature Committee (HGNC).

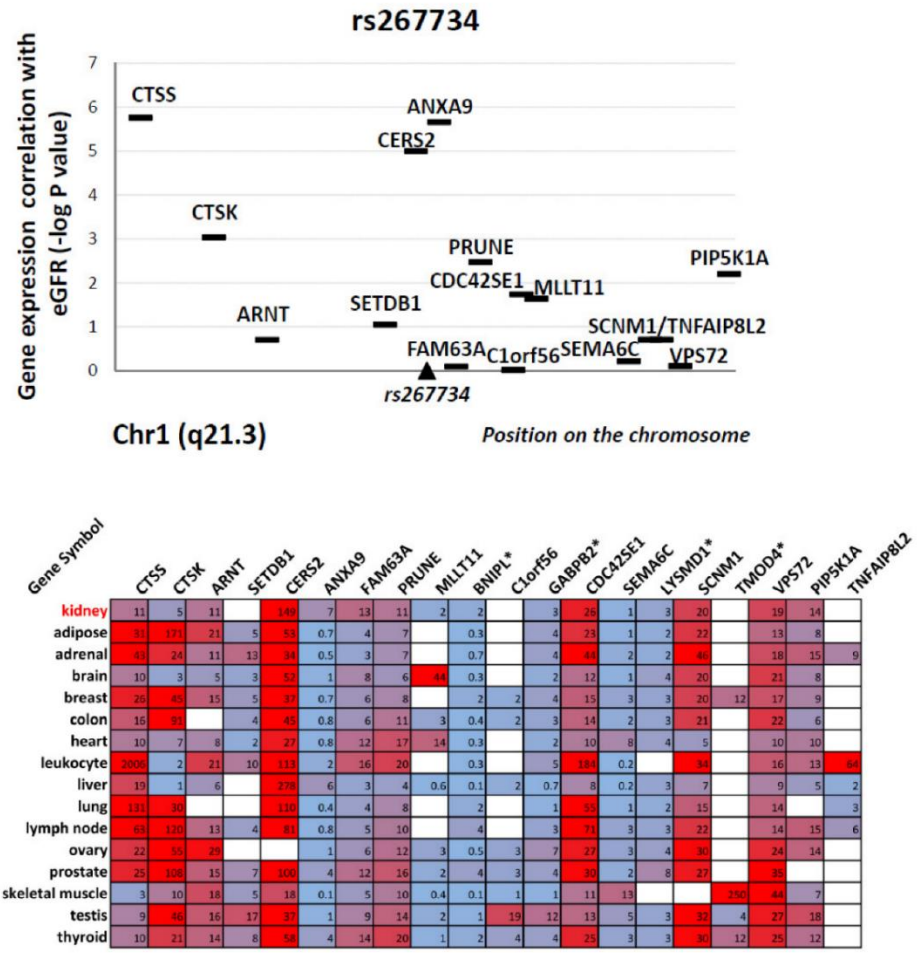


Figure 11. Transcript level expression and correlation of tubule specific transcript levels with renal function around rs267734 locus

The x-axis represents the genomic position of each gene on chromosome 1. The y-axis represents the negative logarithm of the p-value (significance) between the expression of each gene and eGFR (estimated glomerular filtration rate, ml/min/1.73m^2). The lower panel of the chart represents the expression of transcripts within the 250 kbp vicinity of the CKD SNP in 16 human organs. Asterisks indicate genes without probe set IDs on the Affymetrix arrays. Gene symbols are official symbols approved by the Human Genome Organization Gene Nomenclature Committee (HGNC).

On chromosome 5 at the rs11959928 polymorphism, both Disabled homolog 2 (*DAB2*, a putative mitogen-responsive phosphoprotein) and FYN binding protein (*FYB*) showed strong correlation with renal function (P=3.68 x 10^{-5} and P=3 x 10^{-8}, respectively) (*Figure 12A*). The correlation of these genes with eGFR were also validated by QRT-PCR (*Figure 12C*).

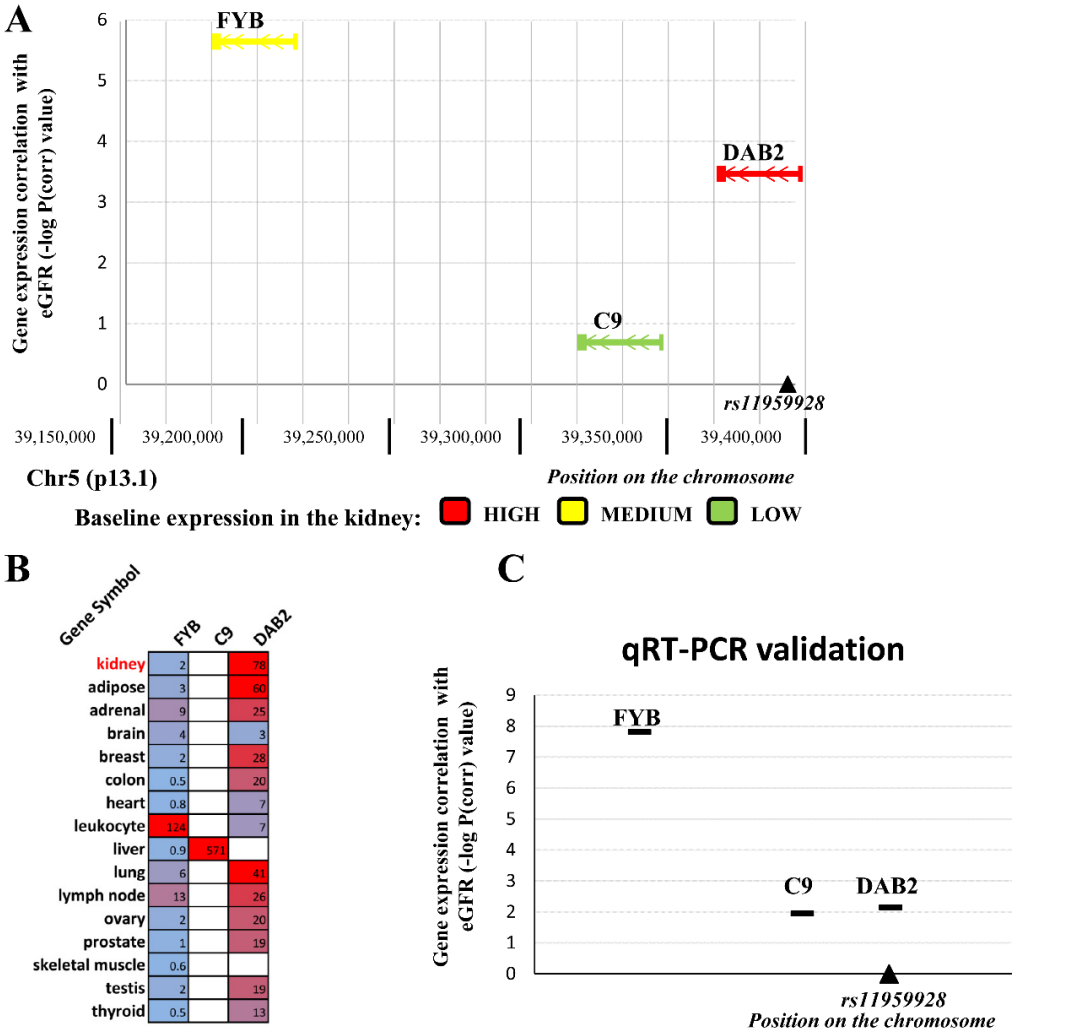


Figure 12. Correlation of tubule specific transcript levels with renal function near the *DAB2* locus (rs11959928)

The x-axis represents the genomic position of each gene on chromosome 5 (A, C). The y-axis represents the negative logarithm of the corrected p-value (significance) between the expression of each gene and eGFR (estimated glomerular filtration rate, ml/min/1.73m^2). Color coding represents the baseline expression of the transcripts in human kidney based on the RNA sequencing data. Red: high, yellow: medium, green: low expression in the kidney (A). Based on the results of the Illumina Body Map (www.ebi.ac.uk) a heat map was generated from the FPKM values of the CKD risk associated transcripts near these SNPs. High expression values (> 90 percentile) are marked red, low expression values (<10 percentile) marked blue. Expressions with FPKM values lower than 0.1 are marked white (B). Quantitative real-time PCR (QRT-PCR) validation confirmed the significant correlation with eGFR of the following transcripts: *FYB* (FYN binding protein) and *DAB2* (Dab, mitogen-responsive phosphoprotein, homolog 2) (C). At the rs11959928 locus, not only the transcript *DAB2* but also the *FYB* shows high correlation with eGFR (A).

Based on renal expression and renal function association, we could prioritize potential target and/or causal genes for CKD development for 39 of 44 (89%) examined loci. As mentioned earlier, there was no gene around rs12437854, and the only nearby gene (*WDR72*, WD Repeat Domain 72) around rs491567 had no probe on the human U133 chips. On the other hand, *WDR72* is highly expressed in human kidney by RNA sequencing analysis, so it still can be an interesting gene for further investigation. No nearby transcript showed association with renal function for three SNPs (rs1394125, rs7805747 and rs4744712). The correlation between these loci and kidney function would need to be re-evaluated.

In summary, our observations showed that the expression of multiple genes around a single locus correlated with kidney function, potentially indicating that the regulation of these genes could be linked.

6.5. Expression quantitative trait loci (eQTL) analysis

6.5.1. eQTL analysis from published gene expression datasets

Polymorphisms associated with kidney function can also directly control baseline transcript levels in disease-relevant types. Based on the possible cross-tissue similarity in eQTL results (42,43), we examined multiple different datasets where genotype and gene expression correlation data were available. These datasets included the MuTHER and other studies (42,57-60), where transcript levels were available from liver, adipose, and lymphoblastoid samples. We examined whether CKD risk SNPs influence local transcript levels (in cis; within 1-Mbp distance) in these datasets. We found that 4 SNPs from the previously identified 44 leading SNPs and 16 SNPs in their linkage disequilibrium ($r^2 \geq 0.8$) acted as cis-eQTLs for 11 different transcripts ($P < 0.05$) (*Table 7.*). Four of these transcripts (33%) were outside of the 500-kbp window that we used to identify CRATs. All 11 transcripts were at least moderately expressed in human kidney tissue. Only one transcript *CLTB* (Clathrin, light chain B) showed significant linear correlation with eGFR in glomerulus samples (P=0.016). Another transcript, *CERS2* showed variation in gene expression in lymphoblastoid tissue based on the rs267734 and rs267738 genotypes. Furthermore, *CERS2* was differentially expressed in CKD and highly correlated with eGFR in tubule samples (*Tables 5. and 6., Figure 11.*), making it a potential candidate gene for CKD development.

Table 7. eQTL analysis of the CKD risk loci in external datasets

SNPs associated with chronic kidney disease (CKD) acts as cis expression quantitative trait loci (eQTL) in other tissues. P-values indicate the strength of the association between the SNP (as eQTL) and the nearby gene. The sources of the analyses are marked and detailed on the bottom of the table (analysis ID: 1-4). The table shows the results of our study in the case of these genes: baseline expression in the kidney based on our RNA sequencing data indicated in the last column; transcripts which are differentially expressed in CKD and/or transcripts whose expression levels correlate with estimated glomerular filtration rate (eGFR) are marked in red. Gene symbols are official symbols approved by the Human Genome Organization Gene Nomenclature Committee (HGNC). *Chr: chromosome*

Table 7. eQTL analysis of the CKD risk loci in external datasets							
Gene Symbol	**SNP**	**Chr**	**Dis-tance from the SNP (kbp)**	**Gene in the same LD block**	**P-value (eQTL)**	**Ana-lysis ID**	**Baseline expression in the kidney**
SYPL2	rs10857787	1	0	Yes	2.9×10^{-35}	1	High
ATXN7L2	rs10857787	1	16	Yes	1.5×10^{-3}	1	Medium
CERS2	rs267734	1	4	Yes	3.70×10^{-5}	2	High
CERS2	rs267738	1	0	Yes	7.51×10^{-5}	2	
SHROOM3	rs17253722	4	243	Yes	7.32×10^{-6}	3	Medium
CLTB	rs3812035	5	974	No	1.5×10^{-4}	2	High
CLTB	rs6420094	5	974	No	1.5×10^{-4}	2	
CLTB	rs6862195	5	979	No	1.5×10^{-4}	2	
RMND5B	rs3812035	5	748	No	4.06×10^{-5}	2	High
RMND5B	rs6420094	5	747	No	4.06×10^{-5}	2	
RMND5B	rs6862195	5	743	No	4.06×10^{-5}	2	
SLC25A37	rs17786744	8	347	No	4.68×10^{-5}	3	Medium
AP5B1	rs11227299	11	0	Yes	2.39×10^{-6}	4	Medium
AP5B1	rs4014195	11	35	Yes	2.96×10^{-6}	4	
AP5B1	rs9666878	11	66	Yes	4.03×10^{-6}	4	
THUMPD1	rs13333226	16	379	No	8.99×10^{-5}	2	High
THUMPD1	rs13335818	16	385	No	8.99×10^{-5}	2	
THUMPD1	rs4293393	16	380	No	8.99×10^{-5}	2	
CDK12	rs11078895	17	217	Yes	2.0×10^{-11}	1	Medium

PGAP3	rs8076494	17	311	Yes	7.87×10^{-7}	1	High

Type of the eQTL analysis				
ID	**Title**	**Tissue**	**Expression profiling**	**Samples (n)**
1	Mapping the genetic architecture of gene expression in human liver	Liver	Array	427
2	Transcriptome genetics using second generation sequencing in a Caucasian population	Lymphoblastoid	RNAseq	60
3	Mapping cis- and trans-regulatory effects across multiple tissues in twins	Adipose	Array	856
4	Population genomics of human gene expression	Lymphoblastoid	Array	210

6.5.2. eQTL analysis of kidney samples

To search for expression changes driven by CKD associated loci in our samples, we genotyped the kidney samples used for the QRT-PCR validation cohort for rs881858 and rs6420094 SNPs.

Tubule-specific *VEGFA* transcript levels were lower in patients who were homozygous for the major allele on the rs881858 locus compared to heterozygous or minor allele homozygous samples (*Figure 13A*). However, this association was only true in healthy kidney samples (n=21, eGFR > 85 ml/min/1.73m^2), but as mentioned above, eQTL analysis is mostly performed on tissues or cells from healthy, control samples. Glomerular or tubule-specific *VEGFA* transcript and protein expression levels also highly correlated with eGFR (*Figures 5C, 13B-D*). These results indicate that the rs881858 polymorphism likely influences *VEGFA* transcript levels, however, this association cannot be observed in samples with kidney disease, because of the high influence of disease conditions on *VEGFA* expression. These results indicate that *VEGFA* could be an important CKD candidate gene.

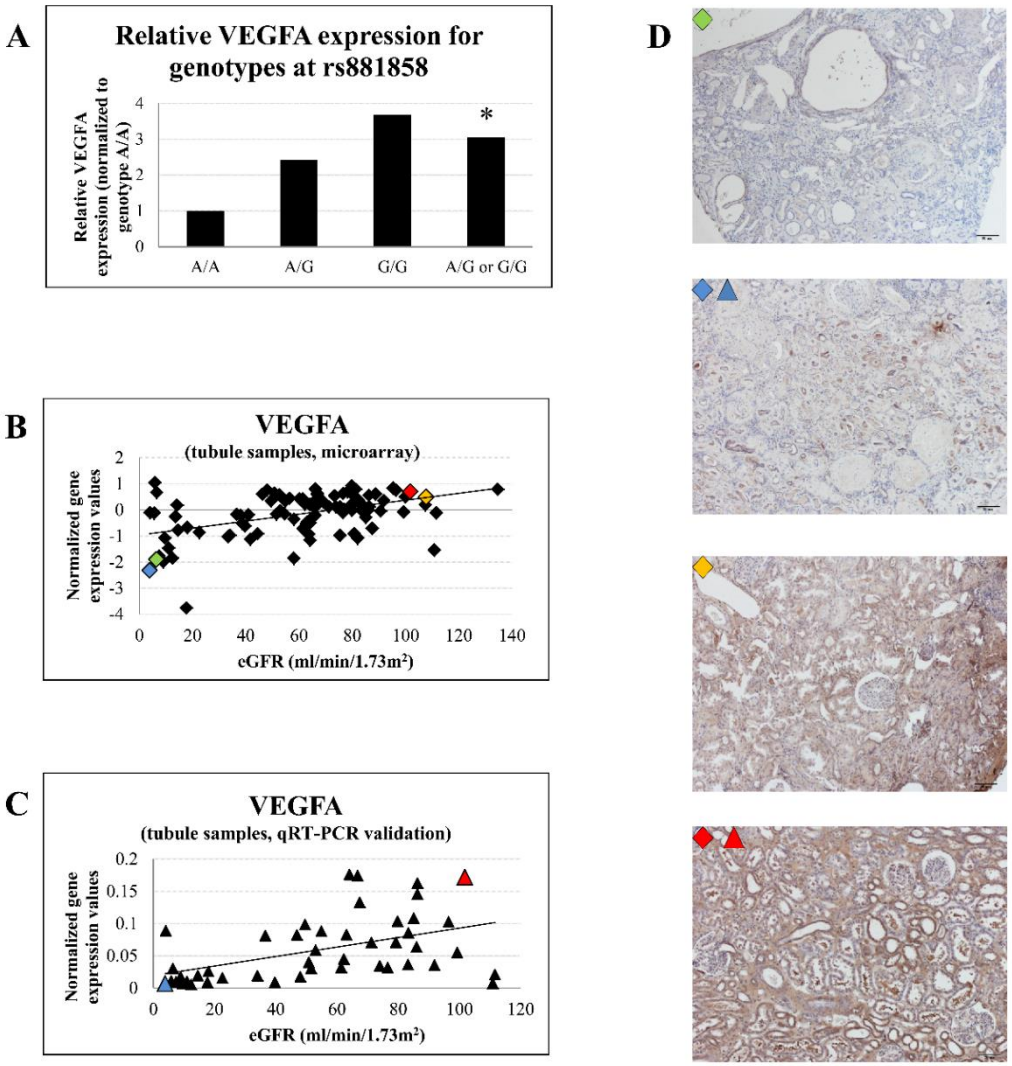


Figure 13. The expression of VEGFA correlates with renal function

The expression of *VEGFA* (vascular endothelial growth factor A) is significantly lower (P=0.025) in samples homozygous for A alleles (A/A, n=7) at the rs881585 locus compering to samples with minor alleles (A/G, n=7 or G/G, n=7) at this locus. Only control samples (eGFR> 85 ml/min/1.73m2) were used for the analysis (A). Microarray based transcript levels of *VEGFA* correlate with renal function in tubule samples (R^2=0.219, P=1.7 x 10^{-6}) (B). QRT-PCR-based *VEGFA* transcript levels (R^2=0.228, P=7.8 x 10^{-4}) confirm its correlation with kidney function (C). VEGFA protein expression (by immunohistochemistry) correlates with transcript levels (Scale bar: 50 µm, counterstained with Hematoxylin) (D).

Additionally, we examined whether genetic polymorphism (rs6420094) on chromosome 5 around *SLC34A1* will influence transcript expression. We found that tubule-specific *SLC34A1* expression was significantly higher in patients who were homozygous for the major allele on the rs6420094 locus compared with heterozygous or

minor allele homozygous samples (n=18, eGFR > 85 ml/min/1.73m^2, P=0.0305) (*Figure 14.*). (Original Article I.)

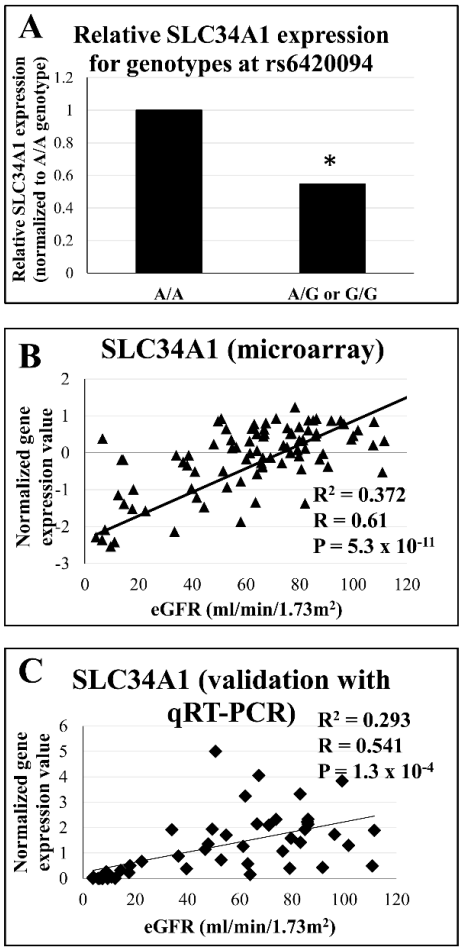


Figure 14. Transcript level of *SLC34A1* is different by rs6420094 genotype

The expression of *SLC34A1* (solute carrier family 34, member 1) is significantly higher (P=0.0305) in samples homozygous for A alleles (A/A, n=9) at the rs6420094 locus compering to samples with minor alleles (A/G, n=3 or G/G, n=6) at this locus. Only control samples (eGFR> 85 ml/min/1.73m^2) were used for the analysis (A). Microarray based transcript levels of *SCL34A1* correlate with renal function in tubule samples (R^2=0.372, P=5.3 x 10^{-11}) (B). QRT-PCR-based *SLC34A1* transcript levels (R^2=0.293, P=1.3 x 10^{-4}) confirm its correlation with kidney function (C).

6.6. Network analysis of CRATs

Taken together, we identified 104 transcripts of 226 CRATs showing significant correlation with eGFR at 39 out of the 44 loci. We examined whether the 104 renal function-correlating CRATs (either in tubule or glomerular samples) in the neighborhood of 39 CKD risk loci show relatedness and can form a network. The network analysis was performed separately on genes that showed positive or negative correlation with kidney function.

Genes showing negative correlation with kidney function (higher expression in CKD) clustered at the TNF (tumor necrosis factor), TGF-β1 (transforming growth factor beta) and NF-κB/RELA (nuclear factor kappa B with p65 subunit) regulatory nodes (*Figure 15A*.). Most members of this cluster are known to play a role in immune function and regulation of inflammation. The second cluster (transcripts with expression that positively correlated with kidney function) centered at VEGFA and ERBB2 (Erb-B2 Receptor Tyrosine Kinase 2) molecules (*Figure 15B*.). These molecules play important roles in maintaining epithelial and endothelial functions. In summary, network analysis highlighted the relatedness of the regulated genes and the potential role of epithelial cell biology and inflammation in CKD. (Original Article I.)

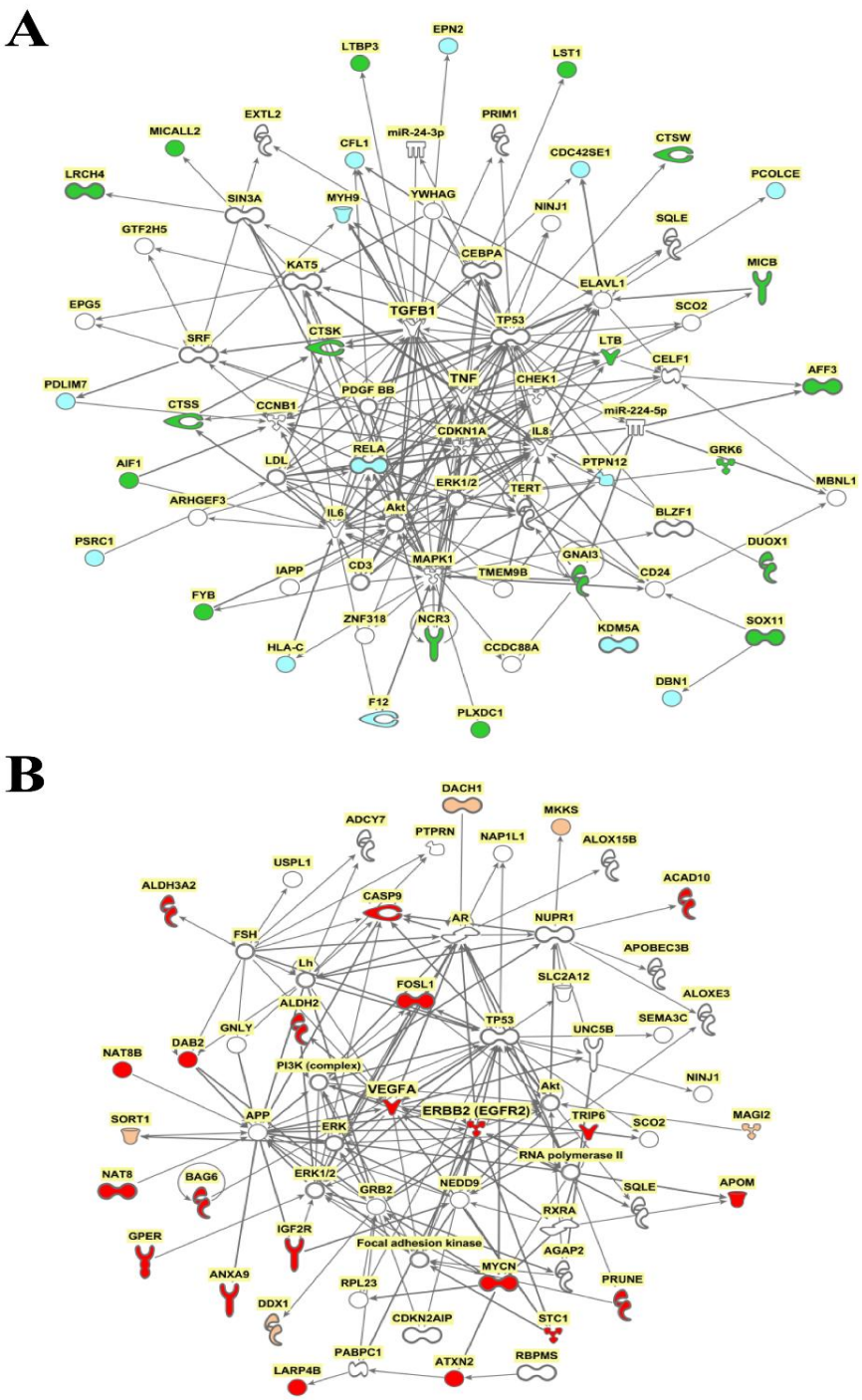


Figure 15. Kidney function correlating CKD risk associated transcripts form tight networks

CKD risk associated transcripts (CRATs) showing negative correlation with estimated glomerular filtration rate (eGFR) (green with P corrected<0.05) clustered around TNF and TGFβ (A). CRATs showing positive correlation with eGFR (red with P corrected<0.05) centered around VEGFA and ERBB2 (B). (Ingenuity Systems, Inc.)

6.7. Transcript levels around loci associated with diabetic nephropathy

We also specifically examined the correlation of the diabetic CKD-associated polymorphisms (rs12437854, rs7583877 and rs1617640) and transcript changes in glomerular and tubular samples only with normal renal function and diabetic kidney disease. After excluding the samples with hypertensive CKD, the analysis was performed on 42 glomerular and 76 tubule samples. The analysis highlighted that the expression of Procollagen C-endopeptidase enhancer (*PCOLCE*) and Thyroid hormone receptor interactor 6 (*TRIP6*) in the vicinity of diabetic CKD SNPs correlate with kidney function. (*Table 8*.) (Original Article I.)

Table 8. The correlation between levels of diabetic CKD risk associated transcripts (D-CRATs) and kidney function in glomeruli (a) and tubules (b)

We identified 18 D-CRATs in the neighborhood of three loci associating with diabetic kidney disease development (rs12437854, rs7583877 and rs1617640). Pearson product moment correlation coefficient (Pearson R) was used to measure the strength of association between gene expression and eGFR. Two-tailed test was used to determine the statistical significance. Four transcripts showed significant correlation with eGFR (P corrected<0.05) after Benjamini Hochberg based multiple testing correction and 10 transcripts showed correlation with GFR with uncorrected p values. Gene symbols are official symbols approved by the Human Genome Organization Gene Nomenclature Committee (HGNC).

Table 8.a eGFR correlation of DKD specific CRATs in Glomeruli				
Gene Symbol	**Pearson R**	**95% confidence interval**	**P (two-tailed)**	**P corrected**
PCOLCE	-0.4555	-0.667 to -0.176	***0.0024***	0.091
LRCH4	-0.3602	-0.597 to -0.063	***0.0191***	0.363
TFR2	0.2841	-0.022 to 0.541	0.068	0.864
MOSPD3	0.1742	-0.137 to 0.454	0.270	0.993
TSC22D4	0.1661	-0.145 to 0.448	0.293	0.993
AGFG2	0.1561	-0.155 to 0.439	0.323	0.993
TRIP6	0.1498	-0.161 to 0.434	0.343	0.993
SRRT	0.1461	-0.165 to 0.431	0.356	0.993
EPO	0.1256	-0.185 to 0.414	0.428	0.993

EPHB4	-0.1240	-0.412 to 0.187	0.434	0.993
GNB2	-0.0958	-0.388 to 0.214	0.546	0.993
ACHE	-0.0864	-0.380 to 0.223	0.586	0.993
LRCH4///SAP25	0.0805	-0.229 to 0.375	0.612	0.993
AFF3	-0.0439	-0.343 to 0.264	0.783	0.993
SLC12A9	-0.0265	-0.328 to 0.280	0.868	0.993
POP7	0.0186	-0.287 to 0.321	0.907	0.993
ACTL6B	0.0156	-0.290 to 0.318	0.922	0.993
FBXO24	0.0036	-0.301 to 0.307	0.982	0.993

Table 8.b eGFR correlation of DKD specific CRATs in Tubules

Gene Symbol	Pearson R	95% confidence interval	P (two-tailed)	P corrected
TRIP6	0.5121	0.3241 to 0.6612	***2.26 x 10^{-6}***	***8.59 x 10^{-5}***
LRCH4	-0.3761	-0.5545 to -0.1646	***8.14 x 10^{-4}***	***0.0155***
SLC12A9	0.3564	0.1423 to 0.5385	***1.58 x 10^{-3}***	***0.0200***
MOSPD3	0.320	0.1019 to 0.5087	***4.84 x 10^{-3}***	***0.0459***
AFF3	-0.2731	-0.4696 to -0.0507	***0.017***	0.100
SRRT	0.2666	0.0438 to 0.4642	***0.020***	0.100
ACHE	-0.2642	-0.4621 to -0.0412	***0.021***	0.100
AGFG2	0.2518	0.0279 to 0.4516	***0.028***	0.119
TFR2	-0.2316	-0.4344 to -0.0065	***0.044***	0.167
EPO	-0.2036	-0.4103 to 0.0229	0.078	0.246
EPHB4	0.1858	-0.0413 to 0.3948	0.108	0.316
FBXO24	-0.1379	-0.3524 to 0.0904	0.235	0.558
ACTL6B	-0.0941	-0.3129 to 0.1342	0.419	0.728
LRCH4///SAP25	-0.0763	-0.2966 to 0.1518	0.512	0.847
POP7	-0.0682	-0.2892 to 0.1597	0.558	0.850
PCOLCE	0.0429	-0.1844 to 0.2658	0.713	0.913
TSC22D4	0.0284	-0.1983 to 0.2523	0.807	0.913
GNB2	0.0140	-0.2121 to 0.2387	0.904	0.913

Furthermore, genes in the vicinity of rs1326934 locus were examined. A multi-stage based GWAS found associations between this locus and diabetic nephropathy in patients with type 1 diabetes (Original Article II.). However, none of the examined SNPs reached genome-wide significance in the discovery cohort, additional analysis drew the attention to this locus. In the discovery GWAS, the minor rs1326934-C allele was less frequent in cases than in controls (0.34 vs 0.43) and was associated with a decreased risk for diabetic nephropathy (OR 0.70 [95% CI 0.60, 0.82]; p=7.87 x 10^{-6}). In an independent cohort, three SNPs demonstrated significant association with diabetic nephropathy:

rs11188343 ($p=9.06 \times 10^{-5}$) and rs4917695 ($p=1.27 \times 10^{-4}$) in addition to rs1326934 ($p=9.85 \times 10^{-5}$). These other two SNPs on chromosome 10 are in near complete linkage disequilibrium with rs1326934 ($r^2 \sim 1$) and mapped to the *SORBS1* gene (Sorbin and SH3 Domain Containing 1).

We examined the expression of the *SORBS1* gene in normal, non-diabetic samples and in samples with diabetic kidney disease (17 glomerular and 39 tubule samples). First, we performed binary analysis comparing cases versus controls, and found significantly higher expression of *SORBS1* in tubule samples of diabetic nephropathy ($P=6 \times 10^{-4}$) (*Figure 16A*). Additional linear correlation analysis revealed a significant inverse correlation between *SORBS1* expression values and eGFR in tubule samples ($R=-0.493$; $p=1.44 \times 10^{-3}$) (*Figure 16B*). According to the data of the Human Protein Atlas, SORBS1 is also expressed in kidney tubule cells on the protein level.

Furthermore, we examined the expression and correlation with eGFR of other nearby genes in the vicinity of rs1326934, where not only *SORBS1* but other genes also correlated with eGFR. (*Figure 16C*)

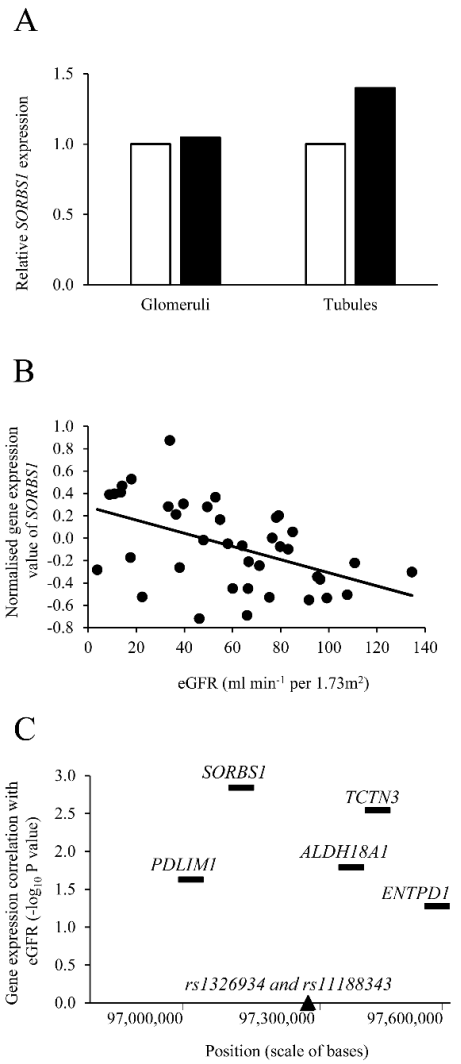


Figure 16. Expression profile of SORBS1 in diabetic nephropathy

The y-axis represents the relative expression of the *SORBS1* (sorbin and SH3 domain containing 1) transcript. Control samples: white bars, diabetic kidney disease (DKD): black bars. The tubular expression of *SORBS1* is significantly up-regulated in DKD compared to control samples in tubules ($P = 6 \times 10^{-4}$) (A). The y-axis shows the relative normalized tubular expressions of *SORBS1*, while the x-axis represents eGFR (estimated glomerular filtration rate, ml/min/1.73m^2) for each sample. Each dot represents transcript levels and eGFR values from a single kidney sample. The line represents the fitted linear correlation values (B). The x-axis represents the genomic position of each gene on chromosome 10 (q24.1) in the 500 kbp vicinity of rs1326934 and rs11188343 loci (triangle). The y-axis represents the negative logarithm of the p-value (significance) between the expression of each gene and eGFR (ml/min/1.73m^2). Not only the *SORBS1* transcript, but also other transcripts correlate with renal function in the vicinity of rs1326934 and rs11188343 loci. Gene symbols are official symbols approved by the Human Genome Organization Gene Nomenclature Committee (HGNC) (C). (Original article II.)

7. Discussion

Understanding complex trait development, such as chronic kidney disease, is a formidable challenge. As discussed above, CKD is a gene environmental disease with several genetic and environmental effects on its development. The first step to understand the development of CKD is to interpret the genetic architecture of the disease. Initial GWASs have provided the first impression of critical regions in the genome with variations that are associated with kidney function. The second step is to identify transcripts that are regulated by SNPs. The working hypothesis of the field is that causal polymorphisms alter transcription factor binding, causing changes in transcript levels in target cell types and inducing disease in specific organs. Because there are hundreds of genetic variants associated with disease development, analyzing variants individually is a daunting task.

Recently, two complementary methods have been developed and successfully applied to identify genes that are targets of the polymorphisms. The first method uses the transcript levels as quantitative traits to identify polymorphisms that influence their levels (eQTL) (62). To perform such analysis, a large human tissue bank from target cell types is necessary where both genetic polymorphisms and transcript levels are analyzed. The second (newer) method uses the cell type specific cellular epigenome for regulatory element annotation and identifies target transcripts that are associated with genetic variants (51). A critical limitation of these methods is that they only identify transcripts that are influenced by a basal transcription factor, because these datasets are generated from control healthy samples. However, it is possible that polymorphisms control transcription factor binding sites for signal-dependent transcription factors. This would mean that the expression of a CKD causing gene is not altered at baseline but shows differences under stress conditions.

In this Ph.D. work, I performed the initial level of such analysis by examining the correlation between transcripts in the vicinity of CKD SNPs and eGFR. Based on recent observations that close to 90% of target transcripts are within 250 kbp of the polymorphism, we defined 306 CRATs. Most prior studies focused only on the two flanking genes, ignoring transcripts that are farther away (12,13). These 306 CRATs could be important for future studies as potential candidates for CKD development. We

determined their baseline expression patterns using highly accurate RNA sequencing methods. Their strong enrichment in the kidney supports their functional role, although it also highlighted that two separate cell types are likely important for CKD development: the kidney and peripheral leukocytes. This finding is supported by both network analysis and tissue-specific gene expression analysis. Mechanistic studies shall determine the role of these cells in CKD development. While diabetic and hypertensive renal disease are considered non-immune-mediated renal diseases, this dogma might need to be revisited.

The highlight of our work is the identification of novel genes in the vicinity of CKD-associated SNPs that show strong correlation with kidney function; thereby, they are potential candidates for CKD development (for example, *FAM47E*, *PLXDC1*, *ACSM2A/B*, *ACSM5*, and *MAGI2*). GWASs of Parkinson's disease showed significant associations with loci at *FAM47E*, but the function of the protein coded by this gene is still unknown. *PLXDC1* (previously known as tumor endothelial marker 7) is primarily associated with angiogenesis in the cancer field, including kidney cancers (63). Also, its increased expression in diabetic retinopathy has been reported (64). We found that the *MAGI2* expression correlates with renal function in glomeruli. Although *MAGI2* is expressed in the brain, *MAGI2* expression is enriched in podocytes (65). Given the critical role of podocytes in kidney disease development, this gene could be an important candidate. *ACSM2A/B* and *-5* are part of the fatty acid oxidation pathway, which emphasizes the importance of metabolic pathways in CKD development. This likely reflects the fact that the kidney is an organ with high-energy demand and fluctuations in energy levels are likely responsible for variations in function (66). Our results are in line with multiple recent publications indicating the importance of energy supply, including those highlighting the role of fatty acid metabolism and mitochondrial function in acute and chronic kidney disease (67,68). Overall, metabolic gene signature can have a critical role in kidney function alterations.

The expression of *CERS2* not only correlates with kidney function, but in other tissues, *CERS2* levels are strongly influenced by a nearby polymorphism, making this gene a very attractive CKD candidate. Ceramide is a common precursor of sphingomyelin and glycosphingolipids in mammalian cells. *CERS2* is responsible for the synthesis of very long chain fatty acid (C20-26 fatty acids)-containing ceramides (69). A recent study showed that there are strain-specific changes in sphingolipid acylation, closely related to

ceramide synthase 2 protein content and activity, with reduced *CERS2* levels/activity observed in glucose intolerant strains and increased content in BALB/c mice which were protected from high fat diet induced glucose intolerance. Overexpression of *CERS2* in primary mouse hepatocytes induced a specific elevation in very long-chain ceramide, but despite the overall increase in ceramide abundance, there was a substantial improvement in insulin signal transduction, as well as decreased ER stress and gluconeogenic markers. These findings suggest that very long-chain sphingolipid species exhibit a protective role against the development of glucose intolerance and hepatic insulin resistance (70).

A critically important observation of this work is that the expression of more than one gene correlated with eGFR on a single genetic locus. We illustrated this observation on the chromosome 16 locus, where not only *UMOD* but also, a cluster of ACSM genes (*ACSM2A/B* and *-5*) showed association with eGFR. This interesting coregulatory pattern was present for most of the CKD GWAS SNPs, potentially indicating that a single polymorphism can control the expression of multiple genes. These observations suggest that a SNP may not only influence a single gene but may cause the differential regulation of an entire gene cluster.

We specifically examined the top hits of GWAS conducted in diabetic kidney disease. Our first analysis revealed that the expression of *PCOLCE* and *TRIP6* in the vicinity of diabetic CKD SNPs correlates with kidney function. Glomerular *PCOLCE* mRNA expression showed negative correlation with eGFR (higher mRNA levels in diseased kidney samples). This gene encodes a glycoprotein which binds and drives the enzymatic cleavage of type I procollagen and increases C-proteinase activity. While it has not been linked to kidney fibrosis in functional studies, its importance in organ fibrosis has been revealed for example in liver fibrosis induced by toxins (71). The protein encoded by *TRIP6* localizes to focal adhesion sites and along actin stress fibers. It regulates lysophosphatidic acid-induced cell migration and it has been implicated in cancer progression (72). We also examined the genes in the vicinity of rs1326934 locus associated with diabetic nephropathy in patients with type 1 diabetes. Around this locus, *SORBS1* showed high negative correlation with eGFR. The sorbin protein, coded by *SORBS1*, was found to be differentially upregulated in glomeruli of rats with diabetic nephropathy compared with rats without diabetic nephropathy (73). In our study, gene expression changes of *SORBS1* were easier to detect in tubules, as *SORBS1* has a higher

tubular expression. Although *SORBS1* expression was significantly upregulated only in tubules, we cannot exclude the importance of glomerular *SORBS1*. Sorbin functions in the signaling and stimulation of insulin. While renal actions of sorbin are not fully established, we speculate that it plays a key role in several processes involved in diabetic nephropathy, including insulin resistance and cytoskeleton architecture. Sorbin acts in the genesis of stress fibers and might, therefore, be involved in podocyte alterations of the slit diaphragm barrier.

Our analysis emphasized the importance of small expression differences in many genes in CKD, but these genes do not seem to be independent but instead, form organized clusters and pathways. We identified two major clusters. One of them centered at epithelial and VEGF signaling. These genes show a linear correlation with kidney function, likely indicating the relationship between epithelial and vascular integrity in progressive nephropathy. The second cluster highlighted TNF and TGF-b1; these genes are known to play important roles in inflammation and fibrosis. Expressions of these transcripts showed an inverse correlation with renal function, indicating an increased expression of these genes in CKD. Interstitial fibrosis (IF) is one of the main histological manifestations of CKD. IF correlates well with CKD and predicts its progression (74,75). It has also been suggested that in diabetic nephropathy, IF predicts eGFR decline better than proteinuria alone or baseline eGFR (76). In recent years the appreciation for fibrosis has increased and several large companies have launched programs aiming to selectively target fibrosis (77). Immune system activation has been consistently observed in non-immune mediated kidney diseases such as hypertension and diabetic CKD (78). However, the enrichment for immune system genes in fibrosis likely represents the influx of inflammatory cells rather than increased expression of inflammatory genes by resident cells (79). This inflammatory cell influx is synonymous with the fibrotic stroma. Taken together, along with functional experiments from the literature, our findings also suggest that increased inflammation and destruction of functioning epithelial cells are cornerstones of fibrosis development.

In summary, we performed a comprehensive functional genomic analysis of CKD-associated GWAS hits in a large set of microdissected human kidney samples. Our results highlighted several novel candidate genes which can have important role in CKD development.

8. Conclusions

Recent GWASs have identified several SNPs associated with chronic kidney disease, however the functional role of these loci has not revealed yet.

In this Ph.D. work:

- We identified 306 CKD-risk associated transcripts (CRATs) in the vicinity of 44 CKD-risk associated loci.
- We examined the expression of these CRATs in a large set of human normal and diseased kidney samples and described the gene expression correlation with kidney function.
- We could highlight genes for further prioritization for 39 of 44 loci (89%).
- Using *UMOD*, *ACSM2A*, and *VEGFA* genes as examples, we showed that these expression changes likely correlate with protein levels.
- Our results also suggest that not only the closest gene but also, several genes in the close vicinity correlate highly with renal function, indicating their potential importance and their potential co-regulation.
- Network analysis of eGFR-correlating CRATs highlighted two major clusters; a positive correlation with epithelial and vascular functions and an inverse correlation with inflammatory gene cluster.

Limitations of the work include the use of nephrectomies instead of kidney biopsies. However, as all samples were obtained from nephrectomies, this would represent a systemic bias and not a disease state-specific bias, and therefore would not change the differential expression analysis between the different groups. Unfortunately, comparable size expression data obtained from kidney biopsies are not available for diabetic and hypertensive CKD subjects. Furthermore, as with most human studies, the work mostly highlights associations that cannot fully establish causality. Changes of transcript levels do not fully indicate that they are functionally relevant. However, even if some of the identified genes are not causally linked to CKD development, the expression levels of these transcripts correlate with kidney function in a large collection

of human kidney samples. Therefore, these genes could be important potential candidate biomarkers for renal function decline.

In summary, we performed a comprehensive functional genomic analysis of CKD-associated GWAS hits. These results highlight multiple new CKD risk associated candidate genes, that were not originally considered by GWAS experiments. Future molecular and cell biology experiments will be needed to understand the functional role of these CRATs. Our findings can direct the renal community toward identification of genes and pathways that may serve as disease biomarkers or causal pathways.

9. Summary

Recent genome-wide association studies (GWAS) have identified multiple loci associated with the risk of chronic kidney disease (CKD) development. Most of these disease-associated single nucleotide polymorphisms are localized to the non-coding region of the genome; therefore, the functional role of these variants in CKD development is largely unknown.

We hypothesized that polymorphisms alter transcription factor binding, thereby influencing the expression of nearby genes. The aim of the Ph.D. work was to map the expression of these transcripts in normal and disease human kidney samples using system biology approaches to identify potential causal and/or target genes for prioritization.

We interrogated the expression and regulation of transcripts in the vicinity of the CKD risk loci using RNA sequencing and gene expression arrays from 95 microdissected control and diseased tubule samples and 51 glomerular samples. Gene expression analysis from 41 tubule samples served for external validation.

In the main part of the work, we examined the expression of 226 transcripts in the vicinity of 44 single nucleotide polymorphisms. 92 transcripts in the tubule compartment and 34 transcripts in glomeruli showed statistically significant correlation with estimated glomerular filtration rate (eGFR). We observed that the expression of multiple genes in the vicinity of any single CKD risk allele correlated with renal function, potentially indicating that genetic variants influence multiple transcripts. Network analysis of eGFR-correlating transcripts highlighted two major clusters; a positive correlation with epithelial and vascular functions and an inverse correlation with inflammatory gene cluster. We separately examined the correlation between gene expression and renal function near loci associated with diabetic nephropathy.

In summary, our functional genomics analysis highlighted novel genes and critical pathways associated with kidney function for future analysis.

10. Összefoglaló

Teljes genomra kiterjedő asszociációs vizsgálatok (GWAS) feltártak több, krónikus vesebetegség kockázatával járó egypontos nukleotid-polimorfizmust (SNP). Ezen SNP-k jelentős része a nem kódoló genetikai állományban helyezkedik el, ezért a krónikus vesebetegségek kialakulásában betöltött szerepükről még keveset tudunk.

Feltételeztük, hogy ezek a polimorfizmusok módosítják a transzkripciós faktorok kötődését, ezzel befolyásolva a környező gének expresszióját. A Ph.D. munka célja az volt, hogy feltérképezzük ezen gének expresszióját egészséges és beteg humán vesemintákban. Rendszerbiológiai megközelítéssel vizsgáltuk ezen SNP-k körüli gének expresszióját, hogy azonosíthassunk a krónikus vesebetegség kialakulásában potenciálisan szerepet játszó géneket.

A krónikus vesebetegséggel összefüggésbe hozható polimorfizmusok körüli gének expresszióját egészséges és beteg mikrodisszektált mintákban (95 tubulus és 51 glomerulus) RNS-szekvenálás és génexpressziós array-k segítségével vizsgáltuk meg. Az eredmények megerősítésére további 41 tubulus minta génexpressziós profilja szolgált.

A munka fő részében a 44 SNP körül elhelyezkedő 226 gén expresszióját vizsgáltuk meg. A tubulusokban 92, míg a glomerulusokban 34 gén expressziója mutatott szignifikáns összefüggést a becsült glomeruláris filtrációs rátával (bGFR). Egy-egy SNP körül több olyan gént is találtunk, amelyek expressziója korrelált a bGFR-val, jelezve, hogy egy polimorfizmus egyszerre akár több gén kifejeződését is befolyásolhatja. A vesefunkcióval korreláló gének hálózati elemzése gyulladás- (inverz korreláció) és epiteliális-vaszkuláris (pozitív korreláció) biológiai hálózatok jelentőségére mutatott rá. Külön vizsgáltuk a diabeteses vesebetegség kockázatával összefüggő lókuszok körüli gének expressziójának összefügéseit a vesefunkcióval.

Összefoglalva, funkcionális genomikai elemzésünk új, vesefunkcióval összefüggő gének és jelátviteli pályák fontosságára hívja fel a figyelmet, ami további vizsgálatok alapja lehet.

11. Bibliography

1. Stevens, P. E., Levin, A., and Kidney Disease: Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group, M. (2013) Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. *Ann Intern Med,* 158: 825-830
2. Hill, N. R., Fatoba, S. T., Oke, J. L., Hirst, J. A., O'Callaghan, C. A., Lasserson, D. S., and Hobbs, F. D. (2016) Global Prevalence of Chronic Kidney Disease - A Systematic Review and Meta-Analysis. *PLoS One,* 11: e0158765
3. Go, A. S., Chertow, G. M., Fan, D., McCulloch, C. E., and Hsu, C. Y. (2004) Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med,* 351: 1296-1305
4. Bochud, M., Elston, R. C., Maillard, M., Bovet, P., Schild, L., Shamlaye, C., and Burnier, M. (2005) Heritability of renal function in hypertensive families of African descent in the Seychelles (Indian Ocean). *Kidney Int,* 67: 61-69
5. Langefeld, C. D., Beck, S. R., Bowden, D. W., Rich, S. S., Wagenknecht, L. E., and Freedman, B. I. (2004) Heritability of GFR and albuminuria in Caucasians with type 2 diabetes mellitus. *Am J Kidney Dis,* 43: 796-800
6. Fox, C. S., Yang, Q., Cupples, L. A., Guo, C. Y., Larson, M. G., Leip, E. P., Wilson, P. W., and Levy, D. (2004) Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: the Framingham Heart Study. *J Am Soc Nephrol,* 15: 2457-2461
7. Hoggart, C. J., Clark, T. G., De Iorio, M., Whittaker, J. C., and Balding, D. J. (2008) Genome-wide significance for dense SNP and resequencing data. *Genet Epidemiol,* 32: 179-185
8. Pe'er, I., Yelensky, R., Altshuler, D., and Daly, M. J. (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol,* 32: 381-385
9. Chambers, J. C., Zhang, W., Lord, G. M., van der Harst, P., Lawlor, D. A., Sehmi, J. S., Gale, D. P., Wass, M. N., Ahmadi, K. R., Bakker, S. J., Beckmann, J., Bilo, H. J., Bochud, M., Brown, M. J., Caulfield, M. J., Connell, J. M., Cook, H. T.,

Cotlarciuc, I., Davey Smith, G., de Silva, R., Deng, G., Devuyst, O., Dikkeschei, L. D., Dimkovic, N., Dockrell, M., Dominiczak, A., Ebrahim, S., Eggermann, T., Farrall, M., Ferrucci, L., Floege, J., Forouhi, N. G., Gansevoort, R. T., Han, X., Hedblad, B., Homan van der Heide, J. J., Hepkema, B. G., Hernandez-Fuentes, M., Hypponen, E., Johnson, T., de Jong, P. E., Kleefstra, N., Lagou, V., Lapsley, M., Li, Y., Loos, R. J., Luan, J., Luttropp, K., Marechal, C., Melander, O., Munroe, P. B., Nordfors, L., Parsa, A., Peltonen, L., Penninx, B. W., Perucha, E., Pouta, A., Prokopenko, I., Roderick, P. J., Ruokonen, A., Samani, N. J., Sanna, S., Schalling, M., Schlessinger, D., Schlieper, G., Seelen, M. A., Shuldiner, A. R., Sjogren, M., Smit, J. H., Snieder, H., Soranzo, N., Spector, T. D., Stenvinkel, P., Sternberg, M. J., Swaminathan, R., Tanaka, T., Ubink-Veltmaat, L. J., Uda, M., Vollenweider, P., Wallace, C., Waterworth, D., Zerres, K., Waeber, G., Wareham, N. J., Maxwell, P. H., McCarthy, M. I., Jarvelin, M. R., Mooser, V., Abecasis, G. R., Lightstone, L., Scott, J., Navis, G., Elliott, P., and Kooner, J. S. (2010) Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet*, 42: 373-375

10. Chasman, D. I., Fuchsberger, C., Pattaro, C., Teumer, A., Boger, C. A., Endlich, K., Olden, M., Chen, M. H., Tin, A., Taliun, D., Li, M., Gao, X., Gorski, M., Yang, Q., Hundertmark, C., Foster, M. C., O'Seaghdha, C. M., Glazer, N., Isaacs, A., Liu, C. T., Smith, A. V., O'Connell, J. R., Struchalin, M., Tanaka, T., Li, G., Johnson, A. D., Gierman, H. J., Feitosa, M. F., Hwang, S. J., Atkinson, E. J., Lohman, K., Cornelis, M. C., Johansson, A., Tonjes, A., Dehghan, A., Lambert, J. C., Holliday, E. G., Sorice, R., Kutalik, Z., Lehtimaki, T., Esko, T., Deshmukh, H., Ulivi, S., Chu, A. Y., Murgia, F., Trompet, S., Imboden, M., Coassin, S., Pistis, G., Consortium, C. A., Consortium, I., Consortium, C. A., Wtccc, Harris, T. B., Launer, L. J., Aspelund, T., Eiriksdottir, G., Mitchell, B. D., Boerwinkle, E., Schmidt, H., Cavalieri, M., Rao, M., Hu, F., Demirkan, A., Oostra, B. A., de Andrade, M., Turner, S. T., Ding, J., Andrews, J. S., Freedman, B. I., Giulianini, F., Koenig, W., Illig, T., Meisinger, C., Gieger, C., Zgaga, L., Zemunik, T., Boban, M., Minelli, C., Wheeler, H. E., Igl, W., Zaboli, G., Wild, S. H., Wright, A. F., Campbell, H., Ellinghaus, D., Nothlings, U., Jacobs, G., Biffar, R., Ernst, F., Homuth, G., Kroemer, H. K., Nauck, M., Stracke, S., Volker, U., Volzke, H.,

Kovacs, P., Stumvoll, M., Magi, R., Hofman, A., Uitterlinden, A. G., Rivadeneira, F., Aulchenko, Y. S., Polasek, O., Hastie, N., Vitart, V., Helmer, C., Wang, J. J., Stengel, B., Ruggiero, D., Bergmann, S., Kahonen, M., Viikari, J., Nikopensius, T., Province, M., Ketkar, S., Colhoun, H., Doney, A., Robino, A., Kramer, B. K., Portas, L., Ford, I., Buckley, B. M., Adam, M., Thun, G. A., Paulweber, B., Haun, M., Sala, C., Mitchell, P., Ciullo, M., Kim, S. K., Vollenweider, P., Raitakari, O., Metspalu, A., Palmer, C., Gasparini, P., Pirastu, M., Jukema, J. W., Probst-Hensch, N. M., Kronenberg, F., Toniolo, D., Gudnason, V., Shuldiner, A. R., Coresh, J., Schmidt, R., Ferrucci, L., Siscovick, D. S., van Duijn, C. M., Borecki, I. B., Kardia, S. L., Liu, Y., Curhan, G. C., Rudan, I., Gyllensten, U., Wilson, J. F., Franke, A., Pramstaller, P. P., Rettig, R., Prokopenko, I., Witteman, J., Hayward, C., Ridker, P. M., Parsa, A., Bochud, M., Heid, I. M., Kao, W. H., Fox, C. S., and Kottgen, A. (2012) Integration of genome-wide association studies with biological knowledge identifies six novel genes related to kidney function. *Hum Mol Genet,* 21: 5329-5343

11. Gudbjartsson, D. F., Holm, H., Indridason, O. S., Thorleifsson, G., Edvardsson, V., Sulem, P., de Vegt, F., d'Ancona, F. C., den Heijer, M., Wetzels, J. F., Franzson, L., Rafnar, T., Kristjansson, K., Bjornsdottir, U. S., Eyjolfsson, G. I., Kiemeney, L. A., Kong, A., Palsson, R., Thorsteinsdottir, U., and Stefansson, K. (2010) Association of variants at UMOD with chronic kidney disease and kidney stones-role of age and comorbid diseases. *PLoS Genet,* 6: e1001039

12. Kottgen, A., Glazer, N. L., Dehghan, A., Hwang, S. J., Katz, R., Li, M., Yang, Q., Gudnason, V., Launer, L. J., Harris, T. B., Smith, A. V., Arking, D. E., Astor, B. C., Boerwinkle, E., Ehret, G. B., Ruczinski, I., Scharpf, R. B., Chen, Y. D., de Boer, I. H., Haritunians, T., Lumley, T., Sarnak, M., Siscovick, D., Benjamin, E. J., Levy, D., Upadhyay, A., Aulchenko, Y. S., Hofman, A., Rivadeneira, F., Uitterlinden, A. G., van Duijn, C. M., Chasman, D. I., Pare, G., Ridker, P. M., Kao, W. H., Witteman, J. C., Coresh, J., Shlipak, M. G., and Fox, C. S. (2009) Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet,* 41: 712-717

13. Kottgen, A., Pattaro, C., Boger, C. A., Fuchsberger, C., Olden, M., Glazer, N. L., Parsa, A., Gao, X., Yang, Q., Smith, A. V., O'Connell, J. R., Li, M., Schmidt, H.,

Tanaka, T., Isaacs, A., Ketkar, S., Hwang, S. J., Johnson, A. D., Dehghan, A., Teumer, A., Pare, G., Atkinson, E. J., Zeller, T., Lohman, K., Cornelis, M. C., Probst-Hensch, N. M., Kronenberg, F., Tonjes, A., Hayward, C., Aspelund, T., Eiriksdottir, G., Launer, L. J., Harris, T. B., Rampersaud, E., Mitchell, B. D., Arking, D. E., Boerwinkle, E., Struchalin, M., Cavalieri, M., Singleton, A., Giallauria, F., Metter, J., de Boer, I. H., Haritunians, T., Lumley, T., Siscovick, D., Psaty, B. M., Zillikens, M. C., Oostra, B. A., Feitosa, M., Province, M., de Andrade, M., Turner, S. T., Schillert, A., Ziegler, A., Wild, P. S., Schnabel, R. B., Wilde, S., Munzel, T. F., Leak, T. S., Illig, T., Klopp, N., Meisinger, C., Wichmann, H. E., Koenig, W., Zgaga, L., Zemunik, T., Kolcic, I., Minelli, C., Hu, F. B., Johansson, A., Igl, W., Zaboli, G., Wild, S. H., Wright, A. F., Campbell, H., Ellinghaus, D., Schreiber, S., Aulchenko, Y. S., Felix, J. F., Rivadeneira, F., Uitterlinden, A. G., Hofman, A., Imboden, M., Nitsch, D., Brandstatter, A., Kollerits, B., Kedenko, L., Magi, R., Stumvoll, M., Kovacs, P., Boban, M., Campbell, S., Endlich, K., Volzke, H., Kroemer, H. K., Nauck, M., Volker, U., Polasek, O., Vitart, V., Badola, S., Parker, A. N., Ridker, P. M., Kardia, S. L., Blankenberg, S., Liu, Y., Curhan, G. C., Franke, A., Rochat, T., Paulweber, B., Prokopenko, I., Wang, W., Gudnason, V., Shuldiner, A. R., Coresh, J., Schmidt, R., Ferrucci, L., Shlipak, M. G., van Duijn, C. M., Borecki, I., Kramer, B. K., Rudan, I., Gyllensten, U., Wilson, J. F., Witteman, J. C., Pramstaller, P. P., Rettig, R., Hastie, N., Chasman, D. I., Kao, W. H., Heid, I. M., and Fox, C. S. (2010) New loci associated with kidney function and chronic kidney disease. *Nat Genet*, 42: 376-384

14. Okada, Y., Sim, X., Go, M. J., Wu, J. Y., Gu, D., Takeuchi, F., Takahashi, A., Maeda, S., Tsunoda, T., Chen, P., Lim, S. C., Wong, T. Y., Liu, J., Young, T. L., Aung, T., Seielstad, M., Teo, Y. Y., Kim, Y. J., Lee, J. Y., Han, B. G., Kang, D., Chen, C. H., Tsai, F. J., Chang, L. C., Fann, S. J., Mei, H., Rao, D. C., Hixson, J. E., Chen, S., Katsuya, T., Isono, M., Ogihara, T., Chambers, J. C., Zhang, W., Kooner, J. S., KidneyGen, C., Consortium, C. K., Albrecht, E., consortium, G., Yamamoto, K., Kubo, M., Nakamura, Y., Kamatani, N., Kato, N., He, J., Chen, Y. T., Cho, Y. S., Tai, E. S., and Tanaka, T. (2012) Meta-analysis identifies

multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet,* 44: 904-909

15. Pattaro, C., Kottgen, A., Teumer, A., Garnaas, M., Boger, C. A., Fuchsberger, C., Olden, M., Chen, M. H., Tin, A., Taliun, D., Li, M., Gao, X., Gorski, M., Yang, Q., Hundertmark, C., Foster, M. C., O'Seaghdha, C. M., Glazer, N., Isaacs, A., Liu, C. T., Smith, A. V., O'Connell, J. R., Struchalin, M., Tanaka, T., Li, G., Johnson, A. D., Gierman, H. J., Feitosa, M., Hwang, S. J., Atkinson, E. J., Lohman, K., Cornelis, M. C., Johansson, A., Tonjes, A., Dehghan, A., Chouraki, V., Holliday, E. G., Sorice, R., Kutalik, Z., Lehtimaki, T., Esko, T., Deshmukh, H., Ulivi, S., Chu, A. Y., Murgia, F., Trompet, S., Imboden, M., Kollerits, B., Pistis, G., Consortium, C. A., Consortium, I., Consortium, C. A., Wellcome Trust Case Control, C., Harris, T. B., Launer, L. J., Aspelund, T., Eiriksdottir, G., Mitchell, B. D., Boerwinkle, E., Schmidt, H., Cavalieri, M., Rao, M., Hu, F. B., Demirkan, A., Oostra, B. A., de Andrade, M., Turner, S. T., Ding, J., Andrews, J. S., Freedman, B. I., Koenig, W., Illig, T., Doring, A., Wichmann, H. E., Kolcic, I., Zemunik, T., Boban, M., Minelli, C., Wheeler, H. E., Igl, W., Zaboli, G., Wild, S. H., Wright, A. F., Campbell, H., Ellinghaus, D., Nothlings, U., Jacobs, G., Biffar, R., Endlich, K., Ernst, F., Homuth, G., Kroemer, H. K., Nauck, M., Stracke, S., Volker, U., Volzke, H., Kovacs, P., Stumvoll, M., Magi, R., Hofman, A., Uitterlinden, A. G., Rivadeneira, F., Aulchenko, Y. S., Polasek, O., Hastie, N., Vitart, V., Helmer, C., Wang, J. J., Ruggiero, D., Bergmann, S., Kahonen, M., Viikari, J., Nikopensius, T., Province, M., Ketkar, S., Colhoun, H., Doney, A., Robino, A., Giulianini, F., Kramer, B. K., Portas, L., Ford, I., Buckley, B. M., Adam, M., Thun, G. A., Paulweber, B., Haun, M., Sala, C., Metzger, M., Mitchell, P., Ciullo, M., Kim, S. K., Vollenweider, P., Raitakari, O., Metspalu, A., Palmer, C., Gasparini, P., Pirastu, M., Jukema, J. W., Probst-Hensch, N. M., Kronenberg, F., Toniolo, D., Gudnason, V., Shuldiner, A. R., Coresh, J., Schmidt, R., Ferrucci, L., Siscovick, D. S., van Duijn, C. M., Borecki, I., Kardia, S. L., Liu, Y., Curhan, G. C., Rudan, I., Gyllensten, U., Wilson, J. F., Franke, A., Pramstaller, P. P., Rettig, R., Prokopenko, I., Witteman, J. C., Hayward, C., Ridker, P., Parsa, A., Bochud, M., Heid, I. M., Goessling, W., Chasman, D. I., Kao, W. H., and Fox, C.

S. (2012) Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet*, 8: e1002584

16. Liu, C. T., Garnaas, M. K., Tin, A., Kottgen, A., Franceschini, N., Peralta, C. A., de Boer, I. H., Lu, X., Atkinson, E., Ding, J., Nalls, M., Shriner, D., Coresh, J., Kutlar, A., Bibbins-Domingo, K., Siscovick, D., Akylbekova, E., Wyatt, S., Astor, B., Mychaleckjy, J., Li, M., Reilly, M. P., Townsend, R. R., Adeyemo, A., Zonderman, A. B., de Andrade, M., Turner, S. T., Mosley, T. H., Harris, T. B., Consortium, C. K., Rotimi, C. N., Liu, Y., Kardia, S. L., Evans, M. K., Shlipak, M. G., Kramer, H., Flessner, M. F., Dreisbach, A. W., Goessling, W., Cupples, L. A., Kao, W. L., and Fox, C. S. (2011) Genetic association for renal traits among participants of African ancestry reveals new loci for renal function. *PLoS Genet*, 7: e1002264
17. Kottgen, A., Kao, W. H., Hwang, S. J., Boerwinkle, E., Yang, Q., Levy, D., Benjamin, E. J., Larson, M. G., Astor, B. C., Coresh, J., and Fox, C. S. (2008) Genome-wide association study for renal traits in the Framingham Heart and Atherosclerosis Risk in Communities Studies. *BMC Med Genet*, 9: 49
18. Bostrom, M. A., Lu, L., Chou, J., Hicks, P. J., Xu, J., Langefeld, C. D., Bowden, D. W., and Freedman, B. I. (2010) Candidate genes for non-diabetic ESRD in African Americans: a genome-wide association study using pooled DNA. *Hum Genet*, 128: 195-204
19. Sandholm, N., Salem, R. M., McKnight, A. J., Brennan, E. P., Forsblom, C., Isakova, T., McKay, G. J., Williams, W. W., Sadlier, D. M., Makinen, V. P., Swan, E. J., Palmer, C., Boright, A. P., Ahlqvist, E., Deshmukh, H. A., Keller, B. J., Huang, H., Ahola, A. J., Fagerholm, E., Gordin, D., Harjutsalo, V., He, B., Heikkila, O., Hietala, K., Kyto, J., Lahermo, P., Lehto, M., Lithovius, R., Osterholm, A. M., Parkkonen, M., Pitkaniemi, J., Rosengard-Barlund, M., Saraheimo, M., Sarti, C., Soderlund, J., Soro-Paavonen, A., Syreeni, A., Thorn, L. M., Tikkanen, H., Tolonen, N., Tryggvason, K., Tuomilehto, J., Waden, J., Gill, G. V., Prior, S., Guiducci, C., Mirel, D. B., Taylor, A., Hosseini, S. M., Group, D. E. R., Parving, H. H., Rossing, P., Tarnow, L., Ladenvall, C., Alhenc-Gelas, F., Lefebvre, P., Rigalleau, V., Roussel, R., Tregouet, D. A., Maestroni, A., Maestroni, S., Falhammar, H., Gu, T., Mollsten, A., Cimponeriu, D., Ioana, M.,

Mota, M., Mota, E., Serafinceanu, C., Stavarachi, M., Hanson, R. L., Nelson, R. G., Kretzler, M., Colhoun, H. M., Panduru, N. M., Gu, H. F., Brismar, K., Zerbini, G., Hadjadj, S., Marre, M., Groop, L., Lajer, M., Bull, S. B., Waggott, D., Paterson, A. D., Savage, D. A., Bain, S. C., Martin, F., Hirschhorn, J. N., Godson, C., Florez, J. C., Groop, P. H., and Maxwell, A. P. (2012) New susceptibility loci associated with kidney disease in type 1 diabetes. *PLoS Genet,* 8: e1002921

20. Tong, Z., Yang, Z., Patel, S., Chen, H., Gibbs, D., Yang, X., Hau, V. S., Kaminoh, Y., Harmon, J., Pearson, E., Buehler, J., Chen, Y., Yu, B., Tinkham, N. H., Zabriskie, N. A., Zeng, J., Luo, L., Sun, J. K., Prakash, M., Hamam, R. N., Tonna, S., Constantine, R., Ronquillo, C. C., Sadda, S., Avery, R. L., Brand, J. M., London, N., Anduze, A. L., King, G. L., Bernstein, P. S., Watkins, S., Genetics of, D., Diabetic Complication Study, G., Jorde, L. B., Li, D. Y., Aiello, L. P., Pollak, M. R., and Zhang, K. (2008) Promoter polymorphism of the erythropoietin gene in severe diabetic eye and kidney complications. *Proc Natl Acad Sci U S A,* 105: 6998-7003

21. Pezzolesi, M. G., Poznik, G. D., Mychaleckyj, J. C., Paterson, A. D., Barati, M. T., Klein, J. B., Ng, D. P., Placha, G., Canani, L. H., Bochenski, J., Waggott, D., Merchant, M. L., Krolewski, B., Mirea, L., Wanic, K., Katavetin, P., Kure, M., Wolkow, P., Dunn, J. S., Smiles, A., Walker, W. H., Boright, A. P., Bull, S. B., Group, D. E. R., Doria, A., Rogus, J. J., Rich, S. S., Warram, J. H., and Krolewski, A. S. (2009) Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. *Diabetes,* 58: 1403-1410

22. Zhang, D., Efendic, S., Brismar, K., and Gu, H. F. (2010) Effects of MCF2L2, ADIPOQ and SOX2 genetic polymorphisms on the development of nephropathy in type 1 Diabetes Mellitus. *BMC Med Genet,* 11: 116

23. Hanson, R. L., Craig, D. W., Millis, M. P., Yeatts, K. A., Kobes, S., Pearson, J. V., Lee, A. M., Knowler, W. C., Nelson, R. G., and Wolford, J. K. (2007) Identification of PVT1 as a candidate gene for end-stage renal disease in type 2 diabetes using a pooling-based genome-wide single nucleotide polymorphism association study. *Diabetes,* 56: 975-983

24. McDonough, C. W., Palmer, N. D., Hicks, P. J., Roh, B. H., An, S. S., Cooke, J. N., Hester, J. M., Wing, M. R., Bostrom, M. A., Rudock, M. E., Lewis, J. P.,

Talbert, M. E., Blevins, R. A., Lu, L., Ng, M. C., Sale, M. M., Divers, J., Langefeld, C. D., Freedman, B. I., and Bowden, D. W. (2011) A genome-wide association study for diabetic nephropathy genes in African Americans. *Kidney Int*, 79: 563-572

25. Shimazaki, A., Kawamura, Y., Kanazawa, A., Sekine, A., Saito, S., Tsunoda, T., Koya, D., Babazono, T., Tanaka, Y., Matsuda, M., Kawai, K., Iiizumi, T., Imanishi, M., Shinosaki, T., Yanagimoto, T., Ikeda, M., Omachi, S., Kashiwagi, A., Kaku, K., Iwamoto, Y., Kawamori, R., Kikkawa, R., Nakajima, M., Nakamura, Y., and Maeda, S. (2005) Genetic variations in the gene encoding ELMO1 are associated with susceptibility to diabetic nephropathy. *Diabetes*, 54: 1171-1178

26. Boger, C. A., Chen, M. H., Tin, A., Olden, M., Kottgen, A., de Boer, I. H., Fuchsberger, C., O'Seaghdha, C. M., Pattaro, C., Teumer, A., Liu, C. T., Glazer, N. L., Li, M., O'Connell, J. R., Tanaka, T., Peralta, C. A., Kutalik, Z., Luan, J., Zhao, J. H., Hwang, S. J., Akylbekova, E., Kramer, H., van der Harst, P., Smith, A. V., Lohman, K., de Andrade, M., Hayward, C., Kollerits, B., Tonjes, A., Aspelund, T., Ingelsson, E., Eiriksdottir, G., Launer, L. J., Harris, T. B., Shuldiner, A. R., Mitchell, B. D., Arking, D. E., Franceschini, N., Boerwinkle, E., Egan, J., Hernandez, D., Reilly, M., Townsend, R. R., Lumley, T., Siscovick, D. S., Psaty, B. M., Kestenbaum, B., Haritunians, T., Bergmann, S., Vollenweider, P., Waeber, G., Mooser, V., Waterworth, D., Johnson, A. D., Florez, J. C., Meigs, J. B., Lu, X., Turner, S. T., Atkinson, E. J., Leak, T. S., Aasarod, K., Skorpen, F., Syvanen, A. C., Illig, T., Baumert, J., Koenig, W., Kramer, B. K., Devuyst, O., Mychaleckyj, J. C., Minelli, C., Bakker, S. J., Kedenko, L., Paulweber, B., Coassin, S., Endlich, K., Kroemer, H. K., Biffar, R., Stracke, S., Volzke, H., Stumvoll, M., Magi, R., Campbell, H., Vitart, V., Hastie, N. D., Gudnason, V., Kardia, S. L., Liu, Y., Polasek, O., Curhan, G., Kronenberg, F., Prokopenko, I., Rudan, I., Arnlov, J., Hallan, S., Navis, G., Consortium, C. K., Parsa, A., Ferrucci, L., Coresh, J., Shlipak, M. G., Bull, S. B., Paterson, N. J., Wichmann, H. E., Wareham, N. J., Loos, R. J., Rotter, J. I., Pramstaller, P. P., Cupples, L. A., Beckmann, J. S., Yang, Q., Heid, I. M., Rettig, R., Dreisbach, A. W., Bochud, M.,

Fox, C. S., and Kao, W. H. (2011) CUBN is a gene locus for albuminuria. *J Am Soc Nephrol,* 22: 555-570

27. Pezzolesi, M. G., Katavetin, P., Kure, M., Poznik, G. D., Skupien, J., Mychaleckyj, J. C., Rich, S. S., Warram, J. H., and Krolewski, A. S. (2009) Confirmation of genetic associations at ELMO1 in the GoKinD collection supports its role as a susceptibility gene in diabetic nephropathy. *Diabetes,* 58: 2698-2702

28. Pezzolesi, M. G., Poznik, G. D., Skupien, J., Smiles, A. M., Mychaleckyj, J. C., Rich, S. S., Warram, J. H., and Krolewski, A. S. (2011) An intergenic region on chromosome 13q33.3 is associated with the susceptibility to kidney disease in type 1 and 2 diabetes. *Kidney Int,* 80: 105-111

29. Teumer, A., Tin, A., Sorice, R., Gorski, M., Yeo, N. C., Chu, A. Y., Li, M., Li, Y., Mijatovic, V., Ko, Y. A., Taliun, D., Luciani, A., Chen, M. H., Yang, Q., Foster, M. C., Olden, M., Hiraki, L. T., Tayo, B. O., Fuchsberger, C., Dieffenbach, A. K., Shuldiner, A. R., Smith, A. V., Zappa, A. M., Lupo, A., Kollerits, B., Ponte, B., Stengel, B., Kramer, B. K., Paulweber, B., Mitchell, B. D., Hayward, C., Helmer, C., Meisinger, C., Gieger, C., Shaffer, C. M., Muller, C., Langenberg, C., Ackermann, D., Siscovick, D., Dcct/Edic, Boerwinkle, E., Kronenberg, F., Ehret, G. B., Homuth, G., Waeber, G., Navis, G., Gambaro, G., Malerba, G., Eiriksdottir, G., Li, G., Wichmann, H. E., Grallert, H., Wallaschofski, H., Volzke, H., Brenner, H., Kramer, H., Mateo Leach, I., Rudan, I., Hillege, H. L., Beckmann, J. S., Lambert, J. C., Luan, J., Zhao, J. H., Chalmers, J., Coresh, J., Denny, J. C., Butterbach, K., Launer, L. J., Ferrucci, L., Kedenko, L., Haun, M., Metzger, M., Woodward, M., Hoffman, M. J., Nauck, M., Waldenberger, M., Pruijm, M., Bochud, M., Rheinberger, M., Verweij, N., Wareham, N. J., Endlich, N., Soranzo, N., Polasek, O., van der Harst, P., Pramstaller, P. P., Vollenweider, P., Wild, P. S., Gansevoort, R. T., Rettig, R., Biffar, R., Carroll, R. J., Katz, R., Loos, R. J., Hwang, S. J., Coassin, S., Bergmann, S., Rosas, S. E., Stracke, S., Harris, T. B., Corre, T., Zeller, T., Illig, T., Aspelund, T., Tanaka, T., Lendeckel, U., Volker, U., Gudnason, V., Chouraki, V., Koenig, W., Kutalik, Z., O'Connell, J. R., Parsa, A., Heid, I. M., Paterson, A. D., de Boer, I. H., Devuyst, O., Lazar, J., Endlich, K., Susztak, K., Tremblay, J., Hamet, P., Jacob, H. J., Boger, C. A., Fox, C. S.,

Pattaro, C., and Kottgen, A. (2016) Genome-wide Association Studies Identify Genetic Loci Associated With Albuminuria in Diabetes. *Diabetes,* 65: 803-817

30. Maeda, S., Kobayashi, M. A., Araki, S., Babazono, T., Freedman, B. I., Bostrom, M. A., Cooke, J. N., Toyoda, M., Umezono, T., Tarnow, L., Hansen, T., Gaede, P., Jorsal, A., Ng, D. P., Ikeda, M., Yanagimoto, T., Tsunoda, T., Unoki, H., Kawai, K., Imanishi, M., Suzuki, D., Shin, H. D., Park, K. S., Kashiwagi, A., Iwamoto, Y., Kaku, K., Kawamori, R., Parving, H. H., Bowden, D. W., Pedersen, O., and Nakamura, Y. (2010) A single nucleotide polymorphism within the acetyl-coenzyme A carboxylase beta gene is associated with proteinuria in patients with type 2 diabetes. *PLoS Genet,* 6: e1000842
31. McKnight, A. J., Currie, D., Patterson, C. C., Maxwell, A. P., Fogarty, D. G., and Warren, U. K. G. S. G. (2009) Targeted genome-wide investigation identifies novel SNPs associated with diabetic nephropathy. *Hugo J,* 3: 77-82
32. McKnight, A. J., Maxwell, A. P., Fogarty, D. G., Sadlier, D., Savage, D. A., and Warren, U. K. G. S. G. (2009) Genetic analysis of coronary artery disease single-nucleotide polymorphisms in diabetic nephropathy. *Nephrol Dial Transplant,* 24: 2473-2476
33. McKnight, A. J., Patterson, C. C., Pettigrew, K. A., Savage, D. A., Kilner, J., Murphy, M., Sadlier, D., Maxwell, A. P., and Warren, U. K. G. o. K. i. D. S. G. (2010) A GREM1 gene variant associates with diabetic nephropathy. *J Am Soc Nephrol,* 21: 773-781
34. Pattaro, C., Teumer, A., Gorski, M., Chu, A. Y., Li, M., Mijatovic, V., Garnaas, M., Tin, A., Sorice, R., Li, Y., Taliun, D., Olden, M., Foster, M., Yang, Q., Chen, M. H., Pers, T. H., Johnson, A. D., Ko, Y. A., Fuchsberger, C., Tayo, B., Nalls, M., Feitosa, M. F., Isaacs, A., Dehghan, A., d'Adamo, P., Adeyemo, A., Dieffenbach, A. K., Zonderman, A. B., Nolte, I. M., van der Most, P. J., Wright, A. F., Shuldiner, A. R., Morrison, A. C., Hofman, A., Smith, A. V., Dreisbach, A. W., Franke, A., Uitterlinden, A. G., Metspalu, A., Tonjes, A., Lupo, A., Robino, A., Johansson, A., Demirkan, A., Kollerits, B., Freedman, B. I., Ponte, B., Oostra, B. A., Paulweber, B., Kramer, B. K., Mitchell, B. D., Buckley, B. M., Peralta, C. A., Hayward, C., Helmer, C., Rotimi, C. N., Shaffer, C. M., Muller, C., Sala, C., van Duijn, C. M., Saint-Pierre, A., Ackermann, D., Shriner, D., Ruggiero, D.,

Toniolo, D., Lu, Y., Cusi, D., Czamara, D., Ellinghaus, D., Siscovick, D. S., Ruderfer, D., Gieger, C., Grallert, H., Rochtchina, E., Atkinson, E. J., Holliday, E. G., Boerwinkle, E., Salvi, E., Bottinger, E. P., Murgia, F., Rivadeneira, F., Ernst, F., Kronenberg, F., Hu, F. B., Navis, G. J., Curhan, G. C., Ehret, G. B., Homuth, G., Coassin, S., Thun, G. A., Pistis, G., Gambaro, G., Malerba, G., Montgomery, G. W., Eiriksdottir, G., Jacobs, G., Li, G., Wichmann, H. E., Campbell, H., Schmidt, H., Wallaschofski, H., Volzke, H., Brenner, H., Kroemer, H. K., Kramer, H., Lin, H., Leach, I. M., Ford, I., Guessous, I., Rudan, I., Prokopenko, I., Borecki, I., Heid, I. M., Kolcic, I., Persico, I., Jukema, J. W., Wilson, J. F., Felix, J. F., Divers, J., Lambert, J. C., Stafford, J. M., Gaspoz, J. M., Smith, J. A., Faul, J. D., Wang, J. J., Ding, J., Hirschhorn, J. N., Attia, J., Whitfield, J. B., Chalmers, J., Viikari, J., Coresh, J., Denny, J. C., Karjalainen, J., Fernandes, J. K., Endlich, K., Butterbach, K., Keene, K. L., Lohman, K., Portas, L., Launer, L. J., Lyytikainen, L. P., Yengo, L., Franke, L., Ferrucci, L., Rose, L. M., Kedenko, L., Rao, M., Struchalin, M., Kleber, M. E., Cavalieri, M., Haun, M., Cornelis, M. C., Ciullo, M., Pirastu, M., de Andrade, M., McEvoy, M. A., Woodward, M., Adam, M., Cocca, M., Nauck, M., Imboden, M., Waldenberger, M., Pruijm, M., Metzger, M., Stumvoll, M., Evans, M. K., Sale, M. M., Kahonen, M., Boban, M., Bochud, M., Rheinberger, M., Verweij, N., Bouatia-Naji, N., Martin, N. G., Hastie, N., Probst-Hensch, N., Soranzo, N., Devuyst, O., Raitakari, O., Gottesman, O., Franco, O. H., Polasek, O., Gasparini, P., Munroe, P. B., Ridker, P. M., Mitchell, P., Muntner, P., Meisinger, C., Smit, J. H., Consortium, I., Consortium, A., Cardiogram, Group, C. H.-H. F., Consortium, E. C., Kovacs, P., Wild, P. S., Froguel, P., Rettig, R., Magi, R., Biffar, R., Schmidt, R., Middelberg, R. P., Carroll, R. J., Penninx, B. W., Scott, R. J., Katz, R., Sedaghat, S., Wild, S. H., Kardia, S. L., Ulivi, S., Hwang, S. J., Enroth, S., Kloiber, S., Trompet, S., Stengel, B., Hancock, S. J., Turner, S. T., Rosas, S. E., Stracke, S., Harris, T. B., Zeller, T., Zemunik, T., Lehtimaki, T., Illig, T., Aspelund, T., Nikopensius, T., Esko, T., Tanaka, T., Gyllensten, U., Volker, U., Emilsson, V., Vitart, V., Aalto, V., Gudnason, V., Chouraki, V., Chen, W. M., Igl, W., Marz, W., Koenig, W., Lieb, W., Loos, R. J., Liu, Y., Snieder, H., Pramstaller, P. P., Parsa, A., O'Connell, J. R., Susztak, K., Hamet, P., Tremblay, J., de Boer, I. H.,

Boger, C. A., Goessling, W., Chasman, D. I., Kottgen, A., Kao, W. H., and Fox, C. S. (2016) Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun,* 7: 10023

35. Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A. K., Canfield, T. K., Giste, E., Diegel, M., Bates, D., Hansen, R. S., Neph, S., Sabo, P. J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S. R., Kaul, R., and Stamatoyannopoulos, J. A. (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science,* 337: 1190-1195
36. Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N. E., Ahfeldt, T., Sachs, K. V., Li, X., Li, H., Kuperwasser, N., Ruda, V. M., Pirruccello, J. P., Muchmore, B., Prokunina-Olsson, L., Hall, J. L., Schadt, E. E., Morales, C. R., Lund-Katz, S., Phillips, M. C., Wong, J., Cantley, W., Racie, T., Ejebe, K. G., Orho-Melander, M., Melander, O., Koteliansky, V., Fitzgerald, K., Krauss, R. M., Cowan, C. A., Kathiresan, S., and Rader, D. J. (2010) From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature,* 466: 714-719
37. Trudu, M., Janas, S., Lanzani, C., Debaix, H., Schaeffer, C., Ikehata, M., Citterio, L., Demaretz, S., Trevisani, F., Ristagno, G., Glaudemans, B., Laghmani, K., Dell'Antonio, G., Swiss Kidney Project on Genes in Hypertension, t., Loffing, J., Rastaldi, M. P., Manunta, P., Devuyst, O., and Rampoldi, L. (2013) Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression. *Nat Med,* 19: 1655-1660
38. Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M., and Bernstein, B. E. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature,* 473: 43-49
39. Harismendy, O., Notani, D., Song, X., Rahim, N. G., Tanasa, B., Heintzman, N., Ren, B., Fu, X. D., Topol, E. J., Rosenfeld, M. G., and Frazer, K. A. (2011) 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature,* 470: 264-268

40. Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., and Cox, N. J. (2010) Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet,* 6: e1000888

41. Goring, H. H., Curran, J. E., Johnson, M. P., Dyer, T. D., Charlesworth, J., Cole, S. A., Jowett, J. B., Abraham, L. J., Rainwater, D. L., Comuzzie, A. G., Mahaney, M. C., Almasy, L., MacCluer, J. W., Kissebah, A. H., Collier, G. R., Moses, E. K., and Blangero, J. (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet,* 39: 1208-1216

42. Nica, A. C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M., Travers, M., Potter, S., Grundberg, E., Small, K., Hedman, A. K., Bataille, V., Tzenova Bell, J., Surdulescu, G., Dimas, A. S., Ingle, C., Nestle, F. O., di Meglio, P., Min, J. L., Wilk, A., Hammond, C. J., Hassanali, N., Yang, T. P., Montgomery, S. B., O'Rahilly, S., Lindgren, C. M., Zondervan, K. T., Soranzo, N., Barroso, I., Durbin, R., Ahmadi, K., Deloukas, P., McCarthy, M. I., Dermitzakis, E. T., Spector, T. D., and Mu, T. C. (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet,* 7: e1002003

43. Schramm, K., Marzi, C., Schurmann, C., Carstensen, M., Reinmaa, E., Biffar, R., Eckstein, G., Gieger, C., Grabe, H. J., Homuth, G., Kastenmuller, G., Magi, R., Metspalu, A., Mihailov, E., Peters, A., Petersmann, A., Roden, M., Strauch, K., Suhre, K., Teumer, A., Volker, U., Volzke, H., Wang-Sattler, R., Waldenberger, M., Meitinger, T., Illig, T., Herder, C., Grallert, H., and Prokisch, H. (2014) Mapping the genetic architecture of gene regulation in whole blood. *PLoS One,* 9: e93844

44. Susztak, K. (2014) Understanding the epigenetic syntax for the genetic alphabet in the kidney. *J Am Soc Nephrol,* 25: 10-17

45. Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., Kellis, M., and Ren, B. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature,* 459: 108-112

46. Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*, 39: 311-318

47. Cisek, K., Krochmal, M., Klein, J., and Mischak, H. (2016) The application of multi-omics and systems biology to identify therapeutic targets in chronic kidney disease. *Nephrol Dial Transplant*, 31: 2003-2011

48. Brosius, F. C., 3rd, Alpers, C. E., Bottinger, E. P., Breyer, M. D., Coffman, T. M., Gurley, S. B., Harris, R. C., Kakoki, M., Kretzler, M., Leiter, E. H., Levi, M., McIndoe, R. A., Sharma, K., Smithies, O., Susztak, K., Takahashi, N., Takahashi, T., and Animal Models of Diabetic Complications, C. (2009) Mouse models of diabetic nephropathy. *J Am Soc Nephrol*, 20: 2503-2512

49. Kitada, M., Ogura, Y., and Koya, D. (2016) Rodent models of diabetic nephropathy: their utility and limitations. *Int J Nephrol Renovasc Dis* 9: 279-290

50. Yang, H. C., Zuo, Y., and Fogo, A. B. (2010) Models of chronic kidney disease. *Drug Discov Today Dis Models*, 7: 13-19

51. Ko, Y. A., Mohtat, D., Suzuki, M., Park, A. S., Izquierdo, M. C., Han, S. Y., Kang, H. M., Si, H., Hostetter, T., Pullman, J. M., Fazzari, M., Verma, A., Zheng, D., Greally, J. M., and Susztak, K. (2013) Cytosine methylation changes in enhancer regions of core pro-fibrotic genes characterize kidney fibrosis development. *Genome Biol*, 14: R108

52. Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y. L., Castro, A. F., 3rd, Feldman, H. I., Kusek, J. W., Eggers, P., Van Lente, F., Greene, T., Coresh, J., and Ckd, E. P. I. (2009) A new equation to estimate glomerular filtration rate. *Ann Intern Med*, 150: 604-612

53. Trapnell, C., Pachter, L., and Salzberg, S. L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25: 1105-1111

54. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 7: 562-578

55. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4: 44-57

56. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res,* 37: 113

57. Grundberg, E., Small, K. S., Hedman, A. K., Nica, A. C., Buil, A., Keildson, S., Bell, J. T., Yang, T. P., Meduri, E., Barrett, A., Nisbett, J., Sekowska, M., Wilk, A., Shin, S. Y., Glass, D., Travers, M., Min, J. L., Ring, S., Ho, K., Thorleifsson, G., Kong, A., Thorsteindottir, U., Ainali, C., Dimas, A. S., Hassanali, N., Ingle, C., Knowles, D., Krestyaninova, M., Lowe, C. E., Di Meglio, P., Montgomery, S. B., Parts, L., Potter, S., Surdulescu, G., Tsaprouni, L., Tsoka, S., Bataille, V., Durbin, R., Nestle, F. O., O'Rahilly, S., Soranzo, N., Lindgren, C. M., Zondervan, K. T., Ahmadi, K. R., Schadt, E. E., Stefansson, K., Smith, G. D., McCarthy, M. I., Deloukas, P., Dermitzakis, E. T., Spector, T. D., and Multiple Tissue Human Expression Resource, C. (2012) Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet,* 44: 1084-1089

58. Montgomery, S. B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R. P., Ingle, C., Nisbett, J., Guigo, R., and Dermitzakis, E. T. (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature,* 464: 773-777

59. Schadt, E. E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P. Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C., Zhu, J., Millstein, J., Sieberts, S., Lamb, J., GuhaThakurta, D., Derry, J., Storey, J. D., Avila-Campillo, I., Kruger, M. J., Johnson, J. M., Rohl, C. A., van Nas, A., Mehrabian, M., Drake, T. A., Lusis, A. J., Smith, R. C., Guengerich, F. P., Strom, S. C., Schuetz, E., Rushmore, T. H., and Ulrich, R. (2008) Mapping the genetic architecture of gene expression in human liver. *PLoS Biol,* 6: e107

60. Stranger, B. E., Nica, A. C., Forrest, M. S., Dimas, A., Bird, C. P., Beazley, C., Ingle, C. E., Dunning, M., Flicek, P., Koller, D., Montgomery, S., Tavare, S., Deloukas, P., and Dermitzakis, E. T. (2007) Population genomics of human gene expression. *Nat Genet,* 39: 1217-1224

61. Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P. H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., and Ponten, F. (2015) Proteomics. Tissue-based map of the human proteome. *Science*, 347: 1260419

62. Flutre, T., Wen, X., Pritchard, J., and Stephens, M. (2013) A statistical framework for joint eQTL analysis in multiple tissues. *PLoS Genet*, 9: e1003486

63. Bagley, R. G., Rouleau, C., Weber, W., Mehraein, K., Smale, R., Curiel, M., Callahan, M., Roy, A., Boutin, P., St Martin, T., Nacht, M., and Teicher, B. A. (2011) Tumor endothelial marker 7 (TEM-7): a novel target for antiangiogenic therapy. *Microvasc Res*, 82: 253-262

64. Yamaji, Y., Yoshida, S., Ishikawa, K., Sengoku, A., Sato, K., Yoshida, A., Kuwahara, R., Ohuchida, K., Oki, E., Enaida, H., Fujisawa, K., Kono, T., and Ishibashi, T. (2008) TEM7 (PLXDC1) in neovascular endothelial cells of fibrovascular membranes from patients with proliferative diabetic retinopathy. *Invest Ophthalmol Vis Sci*, 49: 3151-3157

65. Ihara, K., Nishimura, T., Fukuda, T., Ookura, T., and Nishimori, K. (2012) Generation of Venus reporter knock-in mice revealed MAGI-2 expression patterns in adult mice. *Gene Expr Patterns*, 12: 95-101

66. Mount, P. F., and Power, D. A. (2015) Balancing the energy equation for healthy kidneys. *J Pathol*, 237: 407-410

67. Han, S. H., Malaga-Dieguez, L., Chinga, F., Kang, H. M., Tao, J., Reidy, K., and Susztak, K. (2016) Deletion of Lkb1 in Renal Tubular Epithelial Cells Leads to CKD by Altering Metabolism. *J Am Soc Nephrol*, 27: 439-453

68. Kang, H. M., Ahn, S. H., Choi, P., Ko, Y. A., Han, S. H., Chinga, F., Park, A. S., Tao, J., Sharma, K., Pullman, J., Bottinger, E. P., Goldberg, I. J., and Susztak, K. (2015) Defective fatty acid oxidation in renal tubular epithelial cells has a key role in kidney fibrosis development. *Nat Med*, 21: 37-46

69. Yamaji, T., Horie, A., Tachida, Y., Sakuma, C., Suzuki, Y., Kushi, Y., and Hanada, K. (2016) Role of Intracellular Lipid Logistics in the Preferential Usage of Very Long Chain-Ceramides in Glucosylceramide. *Int J Mol Sci*, 17

70. Montgomery, M. K., Brown, S. H., Lim, X. Y., Fiveash, C. E., Osborne, B., Bentley, N. L., Braude, J. P., Mitchell, T. W., Coster, A. C., Don, A. S., Cooney, G. J., Schmitz-Peiffer, C., and Turner, N. (2016) Regulation of glucose homeostasis and insulin action by ceramide acyl-chain length: A beneficial role for very long-chain sphingolipid species. *Biochim Biophys Acta*, 1861: 1828-1839

71. Ippolito, D. L., AbdulHameed, M. D., Tawa, G. J., Baer, C. E., Permenter, M. G., McDyre, B. C., Dennis, W. E., Boyle, M. H., Hobbs, C. A., Streicker, M. A., Snowden, B. S., Lewis, J. A., Wallqvist, A., and Stallings, J. D. (2016) Gene Expression Patterns Associated With Histopathology in Toxic Liver Fibrosis. *Toxicol Sci*, 149: 67-88

72. Lin, V. T., and Lin, F. T. (2011) TRIP6: an adaptor protein that regulates cell motility, antiapoptotic signaling and transcriptional activity. *Cell Signal*, 23: 1691-1697

73. Nakatani, S., Kakehashi, A., Ishimura, E., Yamano, S., Mori, K., Wei, M., Inaba, M., and Wanibuchi, H. (2011) Targeted proteomics of isolated glomeruli from the kidneys of diabetic rats: sorbin and SH3 domain containing 2 is a novel protein associated with diabetic nephropathy. *Exp Diabetes Res*, 2011: 979354

74. An, Y., Xu, F., Le, W., Ge, Y., Zhou, M., Chen, H., Zeng, C., Zhang, H., and Liu, Z. (2015) Renal histologic changes and the outcome in patients with diabetic nephropathy. *Nephrol Dial Transplant*, 30: 257-266

75. Takaori, K., Nakamura, J., Yamamoto, S., Nakata, H., Sato, Y., Takase, M., Nameta, M., Yamamoto, T., Economides, A. N., Kohno, K., Haga, H., Sharma, K., and Yanagita, M. (2016) Severity and Frequency of Proximal Tubule Injury Determines Renal Prognosis. *J Am Soc Nephrol*, 27: 2393-2406

76. Mise, K., Hoshino, J., Ueno, T., Hazue, R., Sumida, K., Hiramatsu, R., Hasegawa, E., Yamanouchi, M., Hayami, N., Suwabe, T., Sawa, N., Fujii, T., Hara, S., Ohashi, K., Takaichi, K., and Ubara, Y. (2015) Clinical and pathological predictors of estimated GFR decline in patients with type 2 diabetes and overt proteinuric diabetic nephropathy. *Diabetes Metab Res Rev*, 31: 572-581

77. Mathew, A., Cunard, R., and Sharma, K. (2011) Antifibrotic treatment and other new strategies for improving renal outcomes. *Contrib Nephrol,* 170: 217-227
78. Navarro-Gonzalez, J. F., Mora-Fernandez, C., Muros de Fuentes, M., and Garcia-Perez, J. (2011) Inflammatory molecules and pathways in the pathogenesis of diabetic nephropathy. *Nat Rev Nephrol,* 7: 327-340
79. Imig, J. D., and Ryan, M. J. (2013) Immune and inflammatory role in renal disease. *Compr Physiol,* 3: 957-976

12. Bibliography of the candidate's publications

12.1. The list of publications related to the Ph.D. thesis

I. **Ledo N**, Ko YA, Park ASD, Kang HM, Han SY, Choi P and Susztak K (2015) Functional genomic annotation of genetic risk loci highlights inflammation and epithelial biology networks in chronic kidney disease. *J Am Soc Nephrol,* 26(3):692-714.

II. Germain M, Pezzolesi MG, Sandholm N, McKnight AJ, Susztak K, Lajer M, Forsblom C, Marre M, Parving HH, Rossing P, Toppila I, Skupien J, Roussel R, Ko YA, **Ledo N**, Folkersen L, Civelek M, Maxwell AP, Tregouet DA, Groop PH, Tarnow L and Hadjadj S (2015) SORBS1 gene, a new candidate for diabetic nephropathy: results from a multi-stage genome-wide association study in patients with type 1 diabetes. *Diabetologia*, 58(3):543-8.

12.2. Other publications of the candidate

1. Ko YA, Yi H, Qiu C, Huang S, Park J, **Ledo N**, Köttgen A, Li H, Rader DJ, Pack MA, Brown CD, Susztak K (2017) Genetic-variation-driven gene-expression changes highlight genes with important functions for kidney disease. *Am J Hum Genet*, 100(6):940-953

2. **Ledó N**, Horváth VJ, Tislér A (2017) A volumenstátusz meghatározásának jelentősége és lehetőségei a szív- és vesebeteg populációban *Hypertonia és Nephrologia*, 21(1)

3. **Ledo N**, Susztak K, Palmer MB (2016) Cell Phenotype Transitions in Renal Fibrosis. *Curr Pathobiol Rep,* 4:19

4. Murber A, Fancsovits P, **Ledo N**, Szakacs M, Rigo J Jr, Urbancsek J (2011) Impact of highly purified versus recombinant follicle stimulating hormone on oocyte quality and embryo development in intracytoplasmic sperm injection cycles. *Acta Biol Hung*, 62(3): 255-64.

5. Pongor E, **Ledo N**, Altdorfer K, Lengyel G, Feher E (2010) Distribution and possible origin of neuropeptide-containing nerve elements in the mammalian liver. *Acta Vet Hung,* 58(2):177-87

6. Murber Á, **Ledó N**, Fancsovits P, Tóthné Gilán Zs, Rigó J jr, Urbancsek J (2010) In vitro fertilizáció (IVF) kezelések során alkalmazott stimulációs protokollok hatása a petesejt- és embrióminőségre, valamint a korai embriófejlődésre *Magyar Nőorvosok Lapja,* 73(1):23-29.

7. Murber A, Fancsovits P, **Ledo N**, Gilan ZT, Rigo J Jr, Urbancsek J (2009) Impact of GnRH analogues on oocyte/embryo quality and embryo development in in vitro fertilization/intracytoplasmic sperm injection cycles: a case control study. *Reprod Biol Endocrinol,* 7:103

13. Acknowledgements

First, I would like to thank my tutor, Dr. Katalin Suszták for the opportunity that I could work with her. I appreciate the great amount of help and knowledge she gave me, and I am grateful that she helped me to choose nephrology as my future specialty.

I would like to thank the help of my other tutor, Dr. András Tislér and I am grateful that he is the one who is guiding me on the way to become a clinical nephrologist.

I am also thankful for the help of the members of the Suszták Lab, especially I give thanks to Dr. Yi-An Ko, Dr. Pazit Beckerman, Dr. Hyun Mi Kang and Dr. Mariya Sweetwyne.

Finally, I am grateful for my family who had always supported me and my work. I owe special thanks to my daughter, Júlia, who let me work on the revision of the paper - even during labor and later when she was only a few weeks old.