Genome-wide Association Studies of Metabolites in CKD Patients Identifies Multiple Loci and Illuminates Tubular Transport Mechanisms

Yong Li[1]*, Peggy Sekula[1]*, Matthias Wuttke[1], Judith Wahrheit[2], Birgit Hausknecht[3], Ulla T. Schultheiss[1], Wolfram Gronwald[4], Pascal Schlosser[1], Sara Tucci[5], Arif B. Ekici[6], Ute Spiekerkoetter[5], the GCKD Investigators, Florian Kronenberg[7], Kai-Uwe Eckardt[3], Peter J. Oefner[4], Anna Köttgen[1]
* these authors contributed equally

Affiliations:
[1]Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center – University of Freiburg, Hugstetter Str. 49, 79106 Freiburg, Germany
[2]Biocrates Life Sciences AG, Eduard-Bodem-Gasse 8, 6020 Innsbruck, Austria
[3]Department of Nephrology and Hypertension, University of Erlangen-Nürnberg, 91054 Erlangen, Germany
[4]Institute of Functional Genomics, University of Regensburg, Am BioPark 9, 93053 Regensburg, Germany
[5]Department of General Pediatrics, Center for Pediatrics and Adolescent Medicine,
Faculty of Medicine and Medical Center – University of Freiburg, Mathildenstraße 1, 79106 Freiburg, Germany
[6]Insitute of Human Genetics, University of Erlangen-Nürnberg, 91054 Erlangen, Germany
[7]Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Schöpfstr. 41, 6020 Innsbruck, Austria

Supplementary Information

List of GCKD Study Investigators

University of Erlangen-Nürnberg: Kai-Uwe Eckardt, Heike Meiselbach, Markus Schneider, Thomas Dienemann, Hans-Ulrich Prokosch, Barbara Bärthlein, André Reis, Arif B. Ekici, Susanne Avendaño, Dinah Becker-Grosspitsch, Birgit Hausknecht, Rita Zitzmann, Anke Weigel, Andreas Beck, Thomas Ganslandt, Sabine Knispel

University of Freiburg: Gerd Walz, Anna Köttgen, Ulla Schultheiß, Simone Meder, Erna Mitsch, Ursula Reinhard

RWTH Aachen University: Jürgen Floege, Georg Schlieper, Turgay Saritas, Sabine Ernst, Astrid Blasius, Kerstin Schmitz

Charité, University Medicine Berlin: Elke Schaeffner, Seema Baid-Agrawal, Kerstin Theisen

Hannover Medical School: Hermann Haller, Jan Menne, Elisabeth Bahlmann

University of Heidelberg: Martin Zeier, Claudia Sommerer, Claudia Föllinger

University of Jena: Gunter Wolf, Martin Busch, Rainer Fuß

Ludwig-Maximilians University of München: Thomas Sitter, Claudia Blank

University of Würzburg: Christoph Wanner, Vera Krane, Karina Schönowsky, Antje Börner-Klein

Medical University of Innsbruck, Division of Genetic Epidemiology: Florian Kronenberg, Julia Raschenberger, Barbara Kollerits, Lukas Forer, Sebastian Schönherr, Hansi Weissensteiner

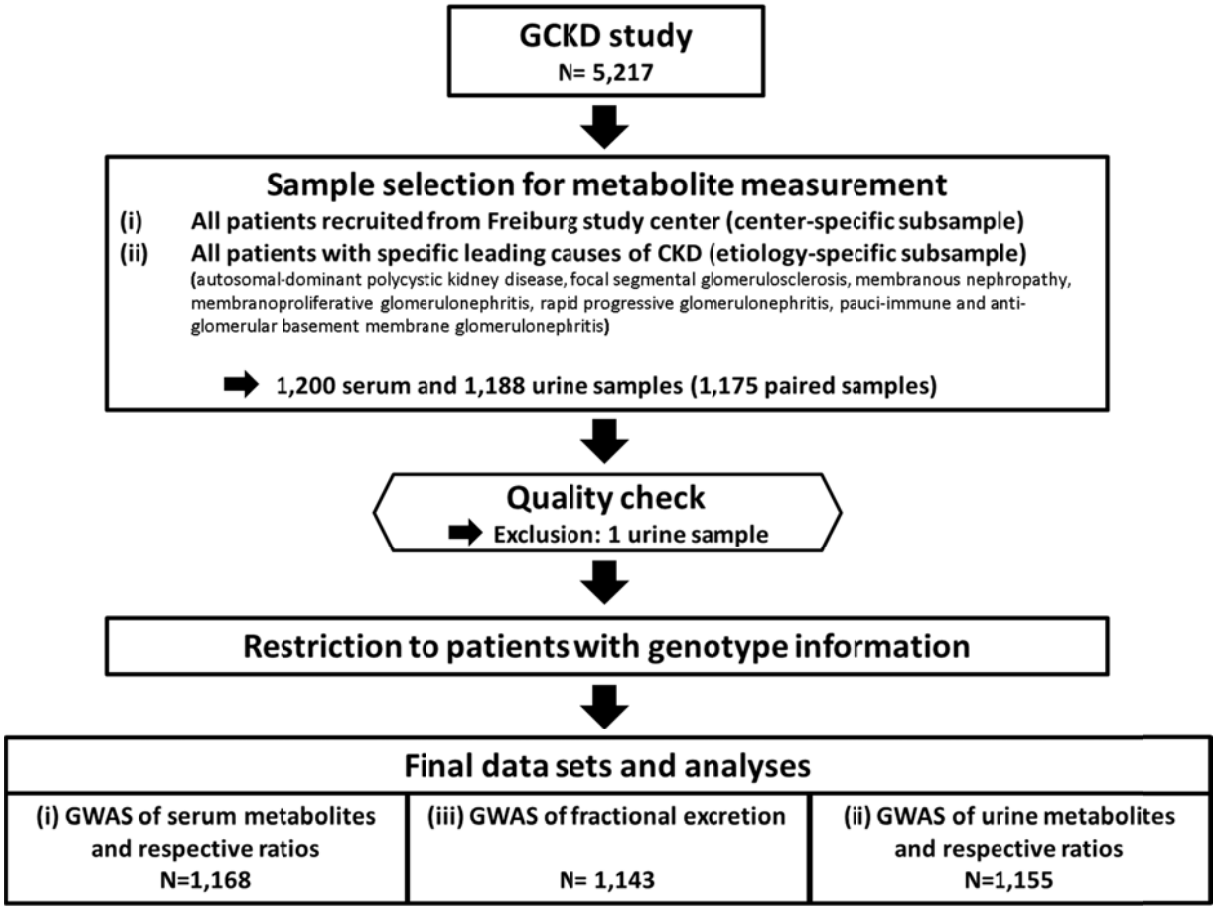
University of Regensburg, Institute of Functional Genomics: Peter Oefner, Wolfram Gronwald, Helena Zacharias

Department of Medical Biometry, Informatics and Epidemiology (IMBIE), University of Bonn: Matthias Schmid, Jennifer Nadal.

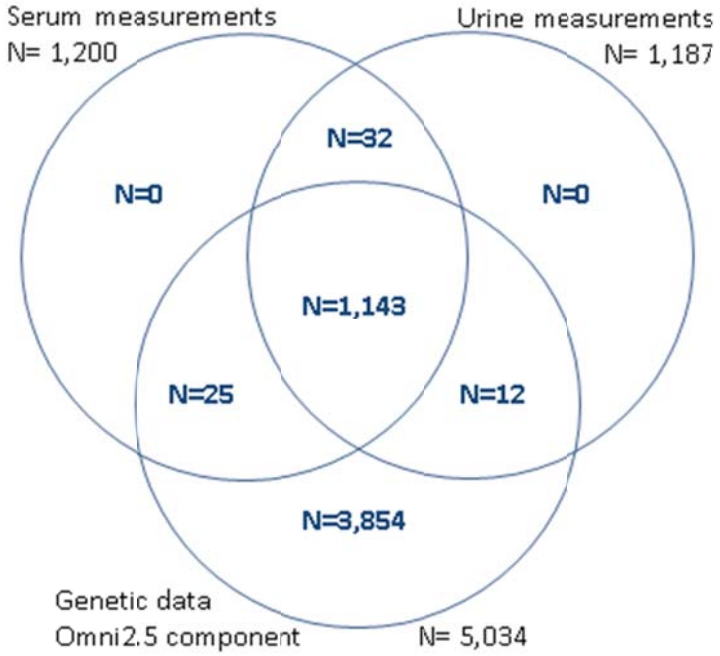
A list of nephrologists currently collaborating with the GCKD study is available at http://www.gckd.org.

Supplementary Information 1: Graphical illustration of patient selection and available data

(A) Flowchart showing the selection process of CKD patients



(B) Venn diagram showing the overlap of patients for which cleaned serum and urine metabolite measurements as well as genetic data were available



Supplementary Information 2: Criteria of quality control and results of genotype data cleaning

Raw data	Omni2.5 component	Exome Chip component
Number of individuals **Number of SNPs**	**5,123** **2,337,794**	**5,123** **230,051**
1) per-individual QC		
Exclusion of individuals with genotyping call rate <97%	48	39
Exclusion of individuals failing sex check	19	19
Exclusion of individuals outside 2 SD of mean heterozygosity	15	31
Cryptic relatedness[1]	11	18
Genetic ancestry outlier by PCA[2]	15	18
Number of individuals remaining[3]	**5,034**	**5,027**
2) per-SNP QC		
Exclusion of SNPs based on call rate[4]	40,681	3,453
Exclusion of SNPs with Hardy-Weinberg Equilibrium p-value <1E-5	6,339	825
Exclusion of SNPs with MAF <1%[5]	801,767	-
Exclusion of SNPs with duplicate positions[5]	10,728	-
Exclusion of SNPs with divergent AF to 1000 Genomes reference[5]	432	-
Number of SNPs remaining[3]	**1,479,401**	**226,233**
Final data		
Number of individuals **Number of SNPs**	**5,034** **1,479,401**	**5,027** **226,233**

[1]Cryptic relatedness was detected by calculating allele sharing IBD using Plink. One individual per pair with IBD proportion >0.1875 was excluded; [2]PCA was performed by Eigenstrat. Outlier detection across the first 10 PCs, thresholds were 8 SD for Omni2.5 and 6 SD for Exome Chip components; [3]The total of excluded individuals or SNPs is not the sum of the steps because of overlaps; [4]SNP call rate cut-off was 96% for Omni2.5 and 95% for Exome Chip; [5]These per-SNP QC steps were only applied to Omni2.5 data.

QC: quality control; SD: standard deviation; PCA: principal component analysis; (M)AF: (minor) allele frequency

Supplementary Information 3: Criteria of quality control and results of metabolite data cleaning

Delivered data	**Serum**		**Urine**	
Number of subjects **Number of analytes**	**1,200** **186**		**1,188** **185**[1]	
0) basic cleaning				
Exclusion of analytes with CV ≥25%	7		43	
Exclusion of analytes with <50% of measurements within range of quantification	40		100	
Number of analytes remaining	**139**		**42**	
1) subject-related cleaning				
Exclusion of subjects with ≥20% missing data	0		0	
Exclusion of subjects with >3 outliers[2]	0		1	
Number of subjects remaining	**1,200**		**1,187**	
2) analyte-related cleaning				
Exclusion of analytes with ≥20% missing data	0		0	
Exclusion of analytes with SD <0.01	0		1	
Number of analytes remaining	**139**		**41**	
Final data				
Number of subjects **Number of analytes**	**1,200** **139**		**1,187** **41**	
Completeness of data (untransformed) after removal of single outliers[1] median (1st quartile)	per subject 100 (99.3)	per analyte 99.8 (99.8)	per subject 97.6 (95.2)	per analyte 99.5 (99.2)

Derived analytes	**Serum Metabolite Ratios**		**Fractional Excretion**[3]		**Urine Metabolite Ratios**	
Number of subjects **Number of analytes**	**1,200** **9,591**		**1,175 (Overlap)** **34**		**1,187** **820**	
Completeness of data (untransformed) after removal of single outliers[1] median (1st quartile)	per subject 99.9 (98.9)	per analyte 99.8 (99.7)	per subject 97.1 (94.3)	per analyte 99.4 (98.8)	per subject 95.0 (90.6)	per analyte 99.2 (98.0)

[1]excluding arginine due to measurement errors; [2]defined as any value outside mean ± 5*SD;
[3]Fractional Excretion (Metabolite)=(100%*$Metabolite_{urine}$*$Creatinine_{serum}$) / ($Metabolite_{serum}$*$Creatinine_{urine}$); excluding FE for creatinine
CV: coefficient of variation; SD: standard deviation
All derived analytes were checked analogous to the single analyte measurements, but did not lead to any additional exclusions.
For details about which metabolites were excluded, please see **Supplementary Table 1**.