

RESEARCH ARTICLE

Genetics of osteopontin in patients with chronic kidney disease: The German Chronic Kidney Disease study

Yurong Cheng^{1,2}, Yong Li¹, Nora Scherer^{1,3}, Franziska Grundner-Culemann¹, Terho Lehtimäki^{4,5,6}, Binisha H. Mishra^{4,5,6}, Olli T. Raitakari^{7,8,9}, Matthias Nauck¹⁰, Kai-Uwe Eckardt^{11,12}, Peggy Sekula¹³, Ulla T. Schultheiss^{1,13}*, on behalf of the GCKD investigators¹

1 Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center—University of Freiburg, Freiburg, Germany, **2** Faculty of Biology, University of Freiburg, Freiburg, Germany, **3** Spemann Graduate School of Biology and Medicine (SGBM), University of Freiburg, Freiburg, Germany, **4** Department of Clinical Chemistry, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland, **5** Finnish Cardiovascular Research Centre, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland, **6** Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland, **7** Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland, **8** Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland, **9** Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland, **10** Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany, **11** Department of Nephrology and Hypertension, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, **12** Department of Nephrology and Medical Intensive Care, Charité, University-Medicine, Berlin, Germany, **13** Department of Medicine IV, Nephrology and Primary Care, Faculty of Medicine and Medical Center—University of Freiburg, Freiburg, Germany

* These authors contributed equally to this work.

† a list of investigators participating in the GCKD Study can be found in [S1 Information](#)

* ulla.schultheiss@uniklinik-freiburg.de



OPEN ACCESS

Citation: Cheng Y, Li Y, Scherer N, Grundner-Culemann F, Lehtimäki T, Mishra BH, et al. (2022) Genetics of osteopontin in patients with chronic kidney disease: The German Chronic Kidney Disease study. *PLoS Genet* 18(4): e1010139. <https://doi.org/10.1371/journal.pgen.1010139>

Editor: Heather J Cordell, Newcastle University, UNITED KINGDOM

Received: October 13, 2021

Accepted: March 9, 2022

Published: April 6, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1010139>

Copyright: © 2022 Cheng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Public posting of individual level participant data is not covered by the informed patient consent form. As stated in the patient consent form and approved by the Ethics

Abstract

Osteopontin (OPN), encoded by *SPP1*, is a phosphorylated glycoprotein predominantly synthesized in kidney tissue. Increased OPN mRNA and protein expression correlates with proteinuria, reduced creatinine clearance, and kidney fibrosis in animal models of kidney disease. But its genetic underpinnings are incompletely understood. We therefore conducted a genome-wide association study (GWAS) of OPN in a European chronic kidney disease (CKD) population. Using data from participants of the German Chronic Kidney Disease (GCKD) study (N = 4,897), a GWAS (minor allele frequency [MAF] ≥ 1%) and aggregated variant testing (AVT, MAF < 1%) of ELISA-quantified serum OPN, adjusted for age, sex, estimated glomerular filtration rate (eGFR), and urinary albumin-to-creatinine ratio (UACR) was conducted. In the project, GCKD participants had a mean age of 60 years (SD 12), median eGFR of 46 mL/min/1.73m² (p25: 37, p75: 57) and median UACR of 50 mg/g (p25: 9, p75: 383). GWAS revealed 3 loci (p < 5.0E-08), two of which replicated in the population-based Young Finns Study (YFS) cohort (p < 1.67E-03): rs10011284, upstream of *SPP1* encoding the OPN protein and related to OPN production, and rs4253311, mapping into *KLKB1* encoding prekallikrein (PK), which is processed to kallikrein (KAL) implicated through the kinin-kallikrein system (KKS) in blood pressure control, inflammation, blood coagulation, cancer, and cardiovascular disease. The *SPP1* gene was also identified by

Committees, a dataset containing pseudonyms can be obtained by collaborating scientists upon approval of a scientific project proposal by the steering committee of the GCKD study: (<http://gckd.org>). Complete summary statistics used in this project from the GCKD study can be obtained here: <https://nxc-1453.imbi.uni-freiburg.de/s/gGrGW5xZRNPEHnc>.

Funding: The work of UTS was supported within the e:Med (<https://www.sys-med.de/en/>) junior consortium CKDNapp (<https://ckdn.app/>), which is funded by grants from the German Ministry of Education and Research (BMBF, grant number 01ZX1912B; <https://www.gesundheitsforschung-bmbf.de/de/ckdnapp-entwicklung-der-chronic-kidney-disease-nephrologists-app-10066.php>). The work of PS was partially funded by German Research Foundation (DFG) Project-ID 431984000 - SFB 1453. The GCKD study was funded by grants from the BMBF (grant number 01ER0804) and the KfH Foundation for Preventive Medicine (<https://www.kfh-stiftung-praeventivmedizin.de/content/stiftung>) and corporate sponsors. Genotyping and measurements of osteopontin were supported by Bayer Pharma AG (<https://www.bayer.com/en/>). The Young Finns Study has been financially supported by the Academy of Finland (<https://www.aka.fi/en/>): grants 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland (<https://www.kela.fi/web/en/>); Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001; <https://www.vsshp.fi/en/tutkijoi/rahoitus/Pages/default.aspx>); Juho Vainio Foundation (<https://juhovainionsaatio.fi/en/juhovainio-foundation/>); Paavo Nurmi Foundation (https://www.paavonurmensaatio.fi/saatio_e3.htm); Finnish Foundation for Cardiovascular Research (<https://www.sydanutkimussaatio.fi/en/foundation/>); Finnish Cultural Foundation (<https://skr.fi/en/>); The Sigrid Juselius Foundation (<https://www.sigridjuselius.fi/en/>); Tampere Tuberculosis Foundation (<http://www.tuberkuloosisaatio.fi/>); Emil Aaltonen Foundation (<https://emilaaaltonen.fi/apurahat/in-english/>); Yrjö Jahnsson Foundation (<https://www.yjs.fi/en/>); Signe and Ane Gyllenberg Foundation (<https://gyllenbergs.fi/en/>); Diabetes Research Foundation of Finnish Diabetes Association (<https://www.diabetes.fi/en/finnish-diabetes-association/association/the-diabetes-research-foundation>). This project has received funding from the European Union's Horizon 2020 research and innovation programme (<https://ec.europa.eu/programmes/horizon2020/en/home>) under grant agreements No 848146 (To Aition) and

AVT ($p = 2.5E-8$), comprising 7 splice-site and missense variants. Among others, downstream analyses revealed colocalization of the OPN association signal at *SPP1* with expression in pancreas tissue, and at *KLKB1* with various plasma proteins in *trans*, and with phenotypes (bone disorder, deep venous thrombosis) in human tissue. In summary, this GWAS of OPN levels revealed two replicated associations. The *KLKB1* locus connects the function of OPN with PK, suggestive of possible further post-translation processing of OPN. Further studies are needed to elucidate the complex role of OPN within human (patho) physiology.

Author summary

Osteopontin (OPN) is involved in many (patho)physiological processes of the human body. Among others, it is known to be associated with adverse kidney outcomes. Since its genetic underpinnings are incompletely understood, we conducted a genome-wide association study of OPN in a European chronic kidney disease (CKD) population (N = 4,897). Of the three detected signals, two could be replicated within a population-based study of Finns. One locus is located upstream of *SPP1* which encodes the OPN protein and is related to OPN production. This gene was also disclosed by an analysis of rare variants, all presumably effecting the gene product. Another locus maps into *KLKB1* encoding prekallikrein (PK) that after processing to kallikrein (KAL) is implicated in blood pressure control and inflammation among others. Overall, our results highlight the multi-functional role of OPN and its possible pathological role in CKD. Further studies are needed to elucidate the complex role of OPN in humans.

Introduction

Osteopontin (OPN) encoded by the *SPP1* gene was first described as a glycoprotein belonging to the SIBLING (Small Integrin-Binding LIgand N-linked Glycoprotein) family in 1985 [1]. OPN is expressed in a multitude of tissues like osteoblasts, osteocytes, odontoblasts (playing a role in mineralization and bone resorption [2,3]) macrophages, smooth muscle cells, and endothelial cells, but can also be found in the inner ear, the central nervous system, and the placenta [1,2]. Although, OPN can be detected in many cell types it is predominantly synthesized and expressed in kidney tissue. OPN production is stimulated by many factors including parathyroid hormone, calcitriol, calcium, phosphate, and cytokines. The protein is able to bind integrins through a specific peptide sequence, the arginine-glycine-aspartic acid (RGD) motif, making interaction with various cell types possible (via the nuclear factor kappa B pathway, [4,5]). In the kidney, integrins can be found in the Bowman's capsule, glomerular epithelium, and vascular epithelium [6,7]. OPN is synthesized in the thick ascending limb of Henle's loop and in the distal tubule [1,8].

In a review by Kaleta (2019), known (patho)physiological roles of OPN have been discussed [1]. Based on this review, the physiological role of OPN in the kidney is not fully understood yet, but it has been suggested as being essential for tubulogenesis [1]. *SPP1* mRNA as well as OPN protein expression were elevated in mostly rat models of kidney diseases and high OPN expression correlated with proteinuria, reduced kidney function, and fibrosis [1]. One study identified various polymorphisms in the *SPP1* promoter region affecting its transcriptional activity [9]. In the past several specific *SPP1* gene variants have been associated with the

No 755320 (TAXINOMISIS). This project has received funding from the European Research Council (ERC; <https://erc.europa.eu/>) advanced grants under grant agreement No 742927 (MULTIEPIGEN project); Tampere University Hospital Supporting Foundation (https://www.tays.fi/en-US/Research_and_development) and Finnish Society of Clinical Chemistry (<https://www.ifcc.org/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We acknowledge support by the Open Access Publication Fund of the University of Freiburg.

Competing interests: The authors have declared that no competing interests exist.

pathogenesis and progression of different kidney diseases. Other case-control studies reported on specific variants in the *SPP1* gene being associated with different kidney disease patients in comparison to a (healthy) control group: For example, rs1126616 was repeatedly reported as a marker for lupus nephritis and immunoglobulin A nephropathy [10–14]. In connection with diabetic nephropathy, the two SNPs in *SPP1*, rs11730582 and rs17524488, have been reported [15,16]. We therefore reasoned that the presence of reduced kidney function may represent a good study setting to further establish our understanding of the genetic underpinnings of OPN levels in kidney disease, as some biologic mechanisms might be upregulated and thus be easier to detect, which has been shown before [17–19]. In Jing et al. [19], for example, the magnitude of effects for known loci identified in a GWAS of serum urate in CKD patients were of similar or higher magnitude than those reported from population-based studies.

The German Chronic Kidney Disease (GCKD) study comprises a large cohort of CKD patients [20]. Besides demographic and clinical data, genetic data are available as well as baseline measurements of serum OPN, providing an ideal setting to explore the genetics of OPN. For this purpose, we performed a GWAS of serum OPN levels in the GCKD study.

Results

Description of the GCKD analysis set

Table 1 gives an overview of baseline characteristics of a selected number of variables for the complete GCKD study cohort and the GWAS analysis set in which participants with complete data on genetics, OPN measurements as well as estimated glomerular filtration rate (eGFR) and urinary albumin-to-creatinine ratio (UACR) are included (**S1 Fig**). There were no major discrepancies between the complete cohort and the analysis set.

Overall, the GWAS analysis set was characterized by a proportion of 60% men with a mean age of 60.2 years (SD: 12.0), with median values of 46.0 mL/min/1.73m² (p25: 37.0; p75: 57.0) for eGFR and of 50.2 mg/g (p25: 9.4; p75: 382.8) for UACR (**Table 1**). Among the included

Table 1. Study sample characteristics of the complete GCKD study cohort (N = 5,217) and the GWAS analysis set (N = 4,897).

	N = 5,217	N = 4,897
Osteopontin, ng/mL, median (p25; p75)	29.2 (20.7; 41.9)	29.2 (20.7; 41.8)
Age, years, mean (SD)	60.1 (12.0)	60.2 (12.0)
Male, N (%)	3,132 (60.0)	2,950 (60.2)
eGFR, mL/min/1.73m ² , median (p25; p75)	46.4 (37.1; 57.4)	46.0 (37.0; 57.0)
UACR, mg/g, median (p25; p75)	50.9 (9.7; 391.7)	50.2 (9.4; 382.8)
HDL, mg/dL, median (p25; p75)	48.4 (39.3; 61.4)	48.5 (39.4; 61.4)
Systolic blood pressure, mmHg, mean (SD)	139.5 (20.4)	139.4 (20.3)
BMI, kg/m ² , mean (SD)	29.8 (6.0)	29.8 (6.0)
Diabetes mellitus, N (%)	1,868 (35.8)	1,715 (35.0)
Smoking, current, N (%)	828 (15.9)	781 (16.0)
CVD, N (%)	1,591 (30.5)	1,489 (30.4)

Continuous variables are mean (SD: standard deviation) for normally distributed variables or median (p25; p75: 25th; 75th percentile) for variables with skewed distributions.

eGFR: estimated glomerular filtration rate; UACR: urinary albumin to creatinine ratio; HDL: high density lipoprotein; BMI: body mass index; CVD: history of cardiovascular disease.

Missingness per variable: N complete cohort (N GWAS cohort): Osteopontin 63 (0), eGFR 55 (0), UACR 90 (0), HDL 66 (8), systolic blood pressure 34 (29), BMI 54 (49), smoking 16 (14), CVD 2 (2).

<https://doi.org/10.1371/journal.pgen.1010139.t001>

participants, 35% had a prevalent diabetes mellitus, 16% were current smokers and 30% reported a history of cardiovascular disease (CVD).

Median OPN levels in the complete cohort were 29.2 ng/mL (p25: 20.7; p75: 41.9; [Table 1](#)). Levels of OPN increased on average across eGFR categories and UACR categories from a median of 25.4 ng/mL for CKD stage G1/2 to 38.5 ng/mL for CKD stage G4/5, as well as from median OPN values of 25.6 ng/mL for UACR stage A1 to mean OPN values of 34.2 ng/mL for UACR stage A3 ([S2 Fig](#)).

Genome-wide association study and fine-mapping

We conducted a GWAS for serum OPN levels (\log_2 -transformed) using ~7.7 million high-quality autosomal bi-allelic variants of the GCKD study with a minor allele frequency (MAF) of ≥ 0.01 ([S1 Table](#)). The quantile-quantile plot comparing observed and expected p-values from the OPN GWAS did not indicate inflation (inflation factor $\lambda = 1.01$), consistent with the absence of systematic errors ([S3 Fig](#)).

Overall, the Manhattan plot revealed three genome-wide significant regions associated with OPN levels (p-value $< 5.0E-08$; [S4 Fig](#)). Besides the three identified regions, conditional analysis did not reveal additional independent signals ([S5 Fig](#)). The respective association results for the three index SNPs (= SNP with the lowest p-value in the respective region) are presented in [Table 2](#). For all three SNPs the coded allele was present frequently (allele frequency range 0.5–0.75). The respective coded allele in our cohort decreased OPN levels on average ([S6 Fig](#)) with effect estimates per copy of the coded allele ranging from -0.10 to -0.18 (SE: 0.01–0.02; [Table 2](#)). One of the index SNPs on chromosome 4 (rs10011284, 4:88833389) is located upstream of *SPP1*, which encodes the protein OPN itself ([Fig 1A](#)). The other index SNP on chromosome 4 (rs4253311, 4:187174683) maps into the *KLKB1* gene (intronic variant), encoding the protein prekallikrein (PK) that is converted to kallikrein (KAL); a protease implicated in the surface-dependent activation of coagulation, bradykinin (BK) release, and potentially the renin angiotensin aldosterone system ([Fig 2A](#)). The index SNP on chromosome 5 (rs2731673, 5:176839898) maps closest to the *F12* gene, which encodes coagulation factor XII, a serine protease that cleaves *KLKB1*-encoded PK to KAL, among other functions and is also related to blood coagulation, fibrinolysis, and the generation of BK ([S7 Fig](#)). A summary of

Table 2. Association results for the 3 index SNPs genome-wide-significantly associated with serum osteopontin levels in the GWAS discovery of the GCKD study (N = 4,897) and in the replication cohort of the YFS (N = 1,979).

SNP	Position (GRCh37)	Gene(s)	Coded allele / non-coded allele	Study	Quality (quality score)	Frequency, coded allele	Beta (SE)	p-value, 2-sided
rs10011284	4:88833389	<i>MEPE</i> (dist = 65421), <i>SPP1</i> (dist = 63413)	A/G	GCKD	imputed (0.999)	0.57	-0.10 (0.01)	8.59E-11
				YFS	imputed (0.997)	0.52	-0.07 (0.02)	2.11E-05
rs4253311	4:187174683	<i>KLKB1</i>	G/A	GCKD	genotyped	0.50	-0.14 (0.01)	5.29E-20
				YFS	imputed (0.997)	0.58	-0.10 (0.02)	1.93E-08
rs2731673	5:176839898	<i>F12</i> (dist = 3321), <i>GRK6</i> (dist = 13789)	C/T	GCKD	imputed (0.987)	0.75	-0.18 (0.02)	4.47E-25
				YFS	imputed (0.985)	0.74	-0.03 (0.02)	1.63E-01

Associations with OPN were adjusted for age, sex, $\log(eGFR)$, $\log(UACR)$ in GCKD (GWAS discovery) and for age, sex, and $\log_2(eGFR)$ in YFS (replication cohort). Statistical significant association p-values are marked in bold: discovery: p-value $< 5E-08$, replication: 1-sided p-value $< 0.05/3$.

<https://doi.org/10.1371/journal.pgen.1010139.t002>

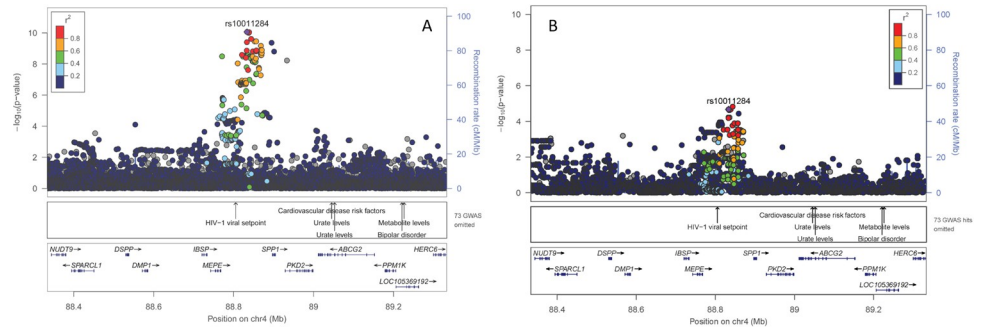


Fig 1. Regional association plot for the region around rs10011284 on chromosome 4: (A) GCKD study (discovery) and (B) YFS cohort (replication). Plots are produced in LocusZoom and show the most strongly associated SNP (purple diamond), SNP colors reflect LD correlation (r^2) using 1000G EUR population as reference. The $-\log_{10}$ p-values (left y-axis) of SNPs are shown according to their chromosomal positions (x-axis, GRCh37); the genetic recombination rates are shown on the right y-axis. The $-\log_{10}$ p-values are shown for both genotyped and imputed SNPs distributed in a 0.8-megabase genomic region.

<https://doi.org/10.1371/journal.pgen.1010139.g001>

annotations combined from different publicly accessible data bases and related to all three SNPs is provided in [S2 Table](#).

We next tested whether these three index SNPs were associated with serum OPN levels in the Young Finns Study (YFS) cohort, a population-based study with a mean age of 38 years (SD: 5.0) and a mean eGFR of 92.6 mL/min/1.73m² (SD: 20.7; [S1 Methods](#)). Both SNPs on chromosome 4, rs10011284 and rs4253311, were significantly associated with OPN levels showing also direction consistency ([Table 2 and Figs 1B and 2B](#)). In contrast, rs2731673 closest to the *F12* gene did not replicate in the YFS cohort ([S7 Fig](#)).

The two replicated SNPs on chromosome 4 explained 1% of OPN levels each within the GCKD study and did not show a non-additive effect on OPN levels ([S8 Fig](#)). Moreover, statistical fine-mapping was performed for the two replicated loci to resolve associated loci into potentially causal variants by constructing credible sets that collectively accounted for 99% posterior probability of containing the variant or variants that cause the association signal (PPA; [Material and methods, \[21\]](#)). However, fine-mapping results are inconclusive as both constructed sets are large and single variants included only exhibit low PPA estimates ([S3 Table](#)).

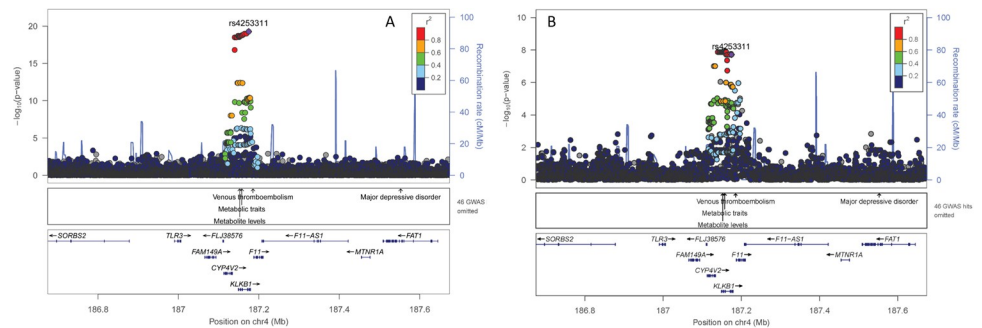


Fig 2. Regional association plot for the region around rs4253311 on chromosome 4: (A) GCKD study (discovery) and (B) YFS cohort (replication). Plots are produced in LocusZoom and show the most strongly associated SNP (purple diamond), SNP colors reflect LD correlation (r^2) using 1000G EUR population as reference. The $-\log_{10}$ p-values (left y-axis) of SNPs are shown according to their chromosomal positions (x-axis, GRCh37); the genetic recombination rates are shown on the right y-axis. The $-\log_{10}$ p-values are shown for both genotyped and imputed SNPs distributed in a 0.8-megabase genomic region.

<https://doi.org/10.1371/journal.pgen.1010139.g002>

Colocalization analyses

In order to learn more about the molecular mechanisms and associated phenotypes underlying the identified association signals for OPN, we compared patterns of OPN GWAS results in predefined regions to respective GWAS summary statistics from three other sources using human data (see **Material and methods** for details). Comparable patterns may indicate a common biological basis.

Gene expression. First, we performed colocalization analyses of OPN GWAS summary statistics related to the two replicated loci with the corresponding GWAS summary statistics of gene expression in *cis* using data from the GTEx project and the NEPTUNE study (**Material and methods**). Colocalization (posterior probability of H4 [p12] >0.8) of the OPN association signals were detected with the expression of *MEPE* in lung (**Fig 3A**) and of *SPP1* in pancreas (**Fig 3B** and **S4 Table**). Furthermore, colocalization was found between the OPN association signal and expression of *F11* in six other tissues (in descending order of H4): tibial artery (**Fig 3C**), brain cortex, terminal ileum part of the small intestine, muscularis of the esophagus, transverse colon, and aortic artery (**S4 Table**). Except for the colocalization of the OPN signal with expression of *MEPE* in lung, the effect direction of both traits (OPN and gene expression) was concordant ($\alpha_{12} > 0$; **S4 Table**).

Plasma proteome. In addition, colocalization analyses were conducted using GWAS summary statistics of SNP associations in *cis* and in *trans* with levels of ~3,000 different plasma proteins (pGWAS) reported by Sun *et al.* (**Material and methods**, [22]). While no colocalization was present for the summary statistics of the GWAS of OPN and proteins in *cis*, pGWAS results for 87 proteins from various protein classes were found to *trans* colocalize with OPN GWAS results for rs4253311 at *KLKB1* (**S5 Table**). For the majority of colocalization results

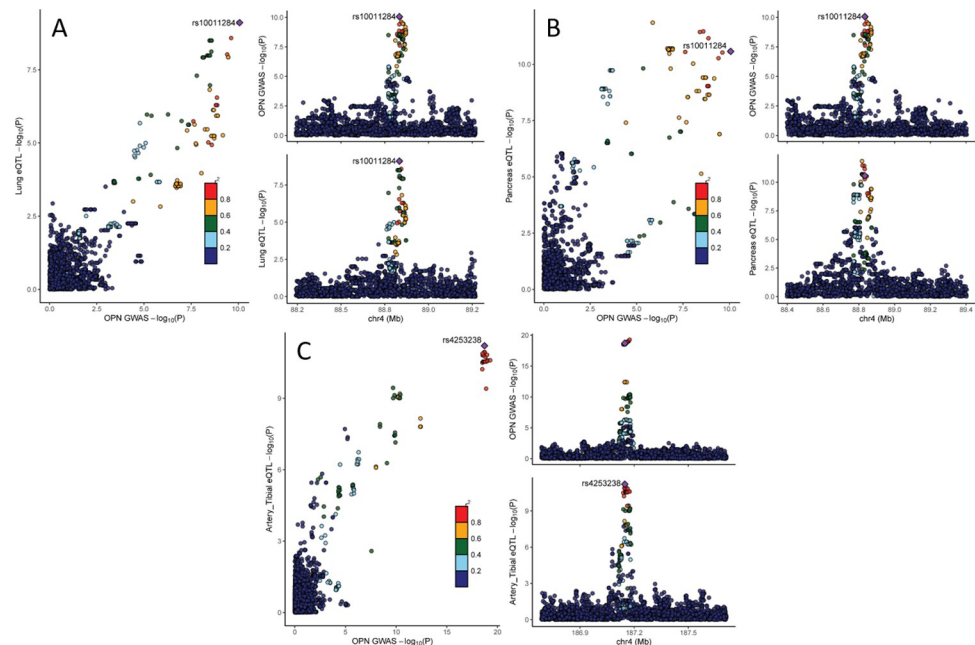


Fig 3. Comparing summary statistics from OPN GWAS and GTEx tissue colocalizing: (A) *MEPE*: OPN and lung tissue (H4: p12 = 0.99), (B) *SPP1*: OPN and pancreas (H4: p12 = 0.85); (C) *F11*: OPN and tibial artery (H4: P12 = 0.98). Left: scatter plot comparing association p-values from both sources against each other (-log10 scale). Upper right: OPN GWAS results for the region of interest. Lower right: GTEx GWAS results for the region of interest of respective organ. Colors reflect LD correlation (r^2) using 1000G EUR population as reference.

<https://doi.org/10.1371/journal.pgen.1010139.g003>

(60/87, 69%), the effect direction of OPN and proteins was concordant, which is in accordance with the fact that the activated KKS is involved in a broad spectrum of processes, like inflammation, cancer, cardiovascular disease, as well as (patho)physiological roles in kidney and the central nervous system.

For 86 mapped proteins, a Gene Ontology (GO) enrichment analysis was conducted to assess whether their encoding genes were enriched in terms representing specific cellular components and molecular functions (**Material and methods, S6 Table**). Because of a hierarchical structure of reference lists, implicated terms are partially dependent on each other showing e.g. overlapping upregulated molecular functions as well as upregulated hormone and receptor activities (**S9 Fig**).

UK Biobank (UKB) diseases. In order to address the interplay between genetic modulation of OPN levels and phenotypes we further conducted a colocalization analysis with binary UKB disease traits that showed a genome-wide significant association in the GeneAtlas resource [23]. Based on marginal association statistics, no positive colocalization was detected between OPN and various disease traits (**S7 Table**). Still, for four of seven traits with an association signal different from the OPN signal in the region (H3: $p_{1.2} > 0.8$), a second independent association signal for the respective disease trait in UKB was detected. We then performed additional colocalization analyses based on the obtained conditional association statistics, and identified a positive colocalization between the OPN association signal at rs4253311 at the *KLKB1* locus and rs1593 for deep venous thrombosis (DVT; H4: $p_{1.2} = 0.96$; **S7 Table and S10 Fig**). In the GeneAtlas GWAS of DVT, the major allele of rs1593 (A, allele frequency = 0.87) was associated with a higher risk for DVT (OR = 1.2, p-value 2.4×10^{-16}), as was the major allele of rs4253311 (major allele: G, allele frequency = 0.51, OR = 1.13, p-value = 2.6×10^{-16}).

Rare variant analysis

Based on 4,879 GCKD participants with available exome chip data, we additionally conducted aggregated rare variant testing (**S1 Fig**). Variants with a MAF <1% and having a major effect on the gene product (nonsynonymous, stop gain/loss, splicing; “qualifying variants”) as annotated by dbNSFP v.2.0 were aggregated (**Material and methods, [24,25]**).

While there was no significant association result when the burden test was used, we found a significant association using the sequence kernel association test (SKAT) for the *SPP1* gene (**Table 3**, p-value = 2.5×10^{-8}), which remained significant after adjustment for the two replicated SNPs from the initial GWAS (p-value = 9.4×10^{-8}). Seven variants were aggregated for the analysis of the *SPP1* gene. In the single variant analysis, effect estimates of the seven variants ranged from -1.20 to 0.24 with rs139555315 (4,889,011,97), a splice-site variant (CADD score: 23.7) showing the most significant association (effect estimate = -1.20, SE = 0.19, p-value = 2.5×10^{-10}) and thus likely driving the association. This is supported by the non-significant result (p-value_{SKAT} = 3.5×10^{-1}) when rs139555315 was excluded from the variant set.

Other genes implicated by GWAS of common variants reached only nominal statistical significance (p < 0.05; *MEPE*: p-value_{SKAT} = 4.6×10^{-3} ; *KLKB1*: p-value_{Burden} = 5.6×10^{-4} , p-value_{SKAT} = 2.8×10^{-3} ; *F12*: p-value_{Burden} = 3.4×10^{-2} , p-value_{SKAT} = 3.0×10^{-2}).

Discussion

In this study, we focused on characterizing the genetics of OPN within a CKD cohort, because OPN levels are known to be associated with adverse kidney outcomes, but genetic underpinnings of this kidney-enhanced protein are not fully understood. The use of a CKD patient cohort might present an advantageous setting in which the transcription of kidney-specific genes may be altered in comparison to the general population, making identification of

Table 3. Association results for *SPP1* (chromosome 4) for the rare variant analysis.

(A) Association results of aggregated variant testing							
Model	Gene	Burden test		SKAT test	No of SNPs aggregated	Total MAC	Cumulative MAF of SNPs
		p-value	Beta (SE)	p-value			
adjusted for age, sex, log(eGFR), log (UACR)	<i>SPP1</i>	1.36E-03	-0.31 (0.10)	2.51E-08	7	59.03	6.05E-03
Same as above plus 2 replicated common SNPs*	<i>SPP1</i>	1.50E-03	-0.30 (0.10)	9.43E-08	7	59.03	6.05E-03
(B) Association results of single variant analysis for aggregated variants							
Model	Variant	Position (GRCh37)	MAF	Beta (SE)	p-value	Exonic effect (CADD)	
adjusted for age, sex, log(eGFR), log (UACR)	rs139555315	88901197	1.54E-03	-1.20 (0.19)	2.54E-10	splicing (23.7)	
	rs140258871	88901249	1.95E-03	0.24 (0.17)	1.62E-01	nonsynonymous (NA)	
	rs138638879	88902774	7.18E-04	-0.18 (0.28)	5.18E-01	nonsynonymous (26.8)	
	rs7435825	88903774	3.08E-04	0.04 (0.42)	9.17E-01	nonsynonymous (12.8)	
	rs149833253	88903825	3.08E-04	-0.45 (0.42)	2.85E-01	nonsynonymous (8.8)	
	rs146563765	88903899	1.02E-03	-0.10 (0.23)	6.83E-01	nonsynonymous (4.2)	
	rs4660	88904005	2.05E-04	-0.52 (0.52)	3.22E-01	nonsynonymous (9.8)	

MAC: minor allele count; MAF: minor allele frequency; NA: not available

* rs10011284, rs4253311. Statistical significant association p-values are marked in bold (aggregated variant testing: p-value < 1.4E-06, single variant analysis: p-value < 5.0E-08. Exonic effects: source dbNSFP v.2.0.

<https://doi.org/10.1371/journal.pgen.1010139.t003>

specific SNPs possibly easier. We identified three loci, two on chromosome 4 (rs10011284, rs4253311), that could be replicated in an external population-based cohort, and rs2731673 on chromosome 5, which could not be replicated. When aggregating rare variants, the *SPP1* gene encoding OPN was detected.

To our knowledge this is the first GWAS of serum OPN levels quantified via ELISA. Other studies like Sun *et al.* conducted GWAS of plasma proteins (pGWAS) including OPN. Here proteins were quantified differently via an aptamer-based technology (SOMAscan, [22]). While this study was likely too small (N = 3,301) to detect any genetic signals for OPN, a study by Pietzner *et al.* (N up to 10,708) provided signals on chromosome 3, 4, and 10 associated with SOMAscan-measured OPN [26]. The detected signal on chromosome 4 (rs5860110, 4:88897106, MAF = 0.30) is a common, intronic indel located within *SPP1* not in linkage disequilibrium (LD) with common variants identified in our project. In a next step, Pietzner *et al.* reported association statistics for the SOMAscan detected loci using a different technique to measure proteins (Olink, antibody-based protein panels). Olink measurements were, however, only available for a small fraction of the study population and none of the three SOMAscan-measured OPN signals could be confirmed. The systematic comparison of protein levels quantified by these two techniques revealed varying correlations (median 0.38, IQR: 0.08–0.64). For OPN, a correlation coefficient of 0.51 was reported. A similar comparison of even more proteomics platforms also reported on a wide range of correlations among measurements [27]. Differences in the detection of genetic signals could thus not only be explained by differences in power but by technical, protein and variant characteristics. Across platforms, any comparison results is thus difficult and meta-analyses could lead to wrong inferences.

rs10011284; 4:88833389 (*SPP1/MEPE* intergenic)

The index SNP rs10011284 (MAF = 0.43) of the first locus on chromosome 4 maps between the *SPP1* and *MEPE* genes. Whether this variant or another is the causal variant underlying the observed association remains unclear as results from statistical fine-mapping are inconclusive. Nevertheless, when aggregating rare variants, *SPP1*, the gene encoding the protein OPN, showed a significant association driven by a splice-site variant. Any errors occurring during the splicing process can lead to false intron removal causing alterations of the open reading frame. In turn this may either lead to formation of a premature stop codon and a shortened protein or more likely to a faster mRNA degradation called nonsense mediated decay [28]. The variant driving this association at *SPP1*, a splice-site variant (rs139555315, 4:88901197, MAF = 1.54E-03), was reported to be associated with pediatric systemic lupus erythematosus [29].

SPP1 is made up of 7 exons containing 942 transcribed nucleotides from the start codon in exon 2 to the stop codon (within exon 7, [30]). OPN belongs to the SIBLING glycoprotein family of secreted phosphoproteins; other members of this family comprise dentin matrix protein 1, dentin-sialophosphoprotein, statherin, bone sialoprotein, and matrix extracellular phosphoglycoprotein (MEPE). Results from our colocalization analysis using GeneAtlas are pointing towards a connection of *SPP1* with bone disorders. Fitting with these results, one of OPN's main physiological functions in the body is the regulation of biomineralization processes [31]. Conflicting results have been reported on OPN as well as *SPP1* polymorphisms and susceptibility to nephrolithiasis in the past [32–36]. From in vitro studies it may be inferred, that OPN inhibits nucleation, growth, and aggregation of calcium oxalate crystals [37], but clinical studies draw a more unclear picture. Some researchers report on a protective role of OPN, where others do not [38,39]. Nonetheless, a recent study from South Asia found a significant association of the *SPP1* rs2853744:G>T polymorphism with urolithiasis [40].

OPN is mostly secreted, but an intracellular form has also been reported [41]. Using reverse-transcription-PCR OPN was found to be expressed in normal human adult kidney, further immunohistochemical analyses and in situ hybridization revealed OPN expression to be restricted to the distal convoluted and straight tubules in kidney cortex and medulla in monkey kidney [42]. Looking at GTEx tissue expression data, a positive colocalization of the OPN association signal at *SPP1* with pancreas tissue was detected. This is in line with findings in the literature where OPN has been suggested to have a role in type 2 diabetes. One study performed by Cai *et al.* investigated a diabetic mouse model SUR1-E1506K^{+/+} and islets from human donors and was able to demonstrate that in islets from human cadaver donors, *OPN* gene expression was elevated in diabetic islets, and externally added OPN significantly increased glucose-stimulated insulin secretion from diabetic but not normal glycemic donors [43]. Many other studies have also investigated OPN's role in pancreatic cancer, here, OPN was found to be a prognostic marker associating higher levels with poor overall prognosis in patients [44].

MEPE (Matric Extracellular PhosphoglycoprotEin) is the gene encoding the secreted calcium-binding phosphoprotein MEPE. A common feature of SIBLING proteins is the Acidic Serine Aspartate Rich MEPE associated motif (ASARM), involved in the regulation of mineralization, bone turnover, mechanotransduction, phosphate and energy metabolism [45]. The ASARM motif is also the connecting link between SIBLINGs and FGF23 thereby being part of the physiological bone-kidney link [45]. MEPE is involved in the regulation of the phosphate homeostasis controlled by the kidney and intestine [46,47]. Colocalization of the OPN association signal at *MEPE* leads to detection of an association with lung tissue. So far a connection between several members of the SIBLING family with lung cancer have been reported, but a definite connection between *MEPE* and lung has not been made before [48]. Since there are

multiple transcript variants known due to alternative splicing a connection between *MEPE* and lung cannot be ruled out and could offer possible research areas in the future. Results from our colocalization analysis using GeneAtlas are pointing towards a connection of *MEPE* with bone disorder. Diseases associated with *MEPE* are osteomalacia and autosomal-dominant hypophosphatemic rickets supporting this connection between *MEPE* and bone disorders [49]. Another phenotype seemingly related to rs10011284 is gout. This association is most likely driven by *ABCG2*, which is located in close proximity to *SPPI* and *MEPE*, and is one of the best described uric acid transporter genes to date [19,50].

rs4253311; 4:187174683 (*KLKB1* intronic)

The 2nd replicated index SNP rs4253311 (MAF = 0.50) on chromosome 4 is an intronic variant of the *KLKB1* gene. Again, whether this variant is the causal variant responsible for the observed association cannot be answered by this study.

KLKB1 encodes prekallikrein (PK) a single-chain zymogen that, after activation to kallikrein (KAL), a serine protease, becomes involved in the surface-dependent activation of blood coagulation, fibrinolysis, kinin generation and inflammation. Diseases associated with *KLKB1* include PK deficiency and malignant essential hypertension [51,52]. PK and subsequently KAL are part of the kallikrein-kinin system (KKS). Main function of KAL includes the release of bradykinin (BK) [53]. Genome-wide association studies in the past identified associated SNPs in *KLKB1* with vasoactive peptides or precursors of vasoactive peptides (BK [54,55], active renin [56], B-type natriuretic peptide [57], aldosterone/renin ratio [57], midregional proadrenomedullin and C-terminal-pro-endothelin-1 [58], L-arginine [59]), and apolipoprotein A IV [60], but not OPN.

Pathways related to this gene include complement and coagulation cascades, as well as degradation of the extracellular matrix [61,62]. An important paralog of *KLKB1* is the gene *F11*. Since our analysis revealed colocalizations between OPN association signal for rs4253311 and expression in multiple tissues for *F11*, it could be presumed that rs4253311 is linked to *F11* rather than *KLKB1*. Interestingly, OPN contains several protease cleavage sites that regulate its activity [31]. Some OPN interaction sites require cleavage by thrombin, another serine protease, to become fully functional. In return, OPN has been shown to be a substrate for other proteases, that regulate its activity [31]. One might speculate that inflammatory processes within the kidney of CKD patients bring together, on the one hand, an activated KKS and, on the other hand, higher OPN levels, thus it might be plausible that new bioactive OPN fragments could possibly be generated by KAL.

Conditional colocalization analyses with SNPs located around *KLKB1* resulted in positive results for rs1593, mapping intronically into the *F11* gene, and DVT. DVT is a serious disease influenced by both genetic and environmental risk factors, but 60% of the variation in risk for DVT has been attributed to genetic risk factors in the past [63]. Genetic studies of DVT have reported several common SNPs in the 4q35.2 locus to be associated with DVT [63]. These common SNPs were localized within *KLKB1* and *F11* amongst others [63].

Other colocalization results showed association signals between the OPN locus at *KLKB1* and 87 plasma proteins. These proteins showed enrichment for proteins of the neuronal cell body in plasma and cerebrospinal fluid of patients. BK, the principal effector of the plasma KKS, is generated systemically and locally (vessel wall) and acts in a paracrine or autocrine way influencing vascular tone and ultrastructure via two G protein-coupled receptors [64,65]. Components of the KKS and in particular BK have been shown to have important functions in the central nervous system by regulating cerebrovascular resistance, vessel capacity and permeability of the blood-brain-barrier. Maintenance of a vascular permeability equilibrium in the

central nervous system is critical for maintaining brain integrity. Associations of the KKS in CNS pathology include several disease states among which are neuropsychiatric lupus, Alzheimer's disease, schizophrenia, and epileptic syndrome [66]. These facts in turn explain the enrichment analysis results of the molecular functions: hormone activity via signaling receptor binding as well as receptor regulatory activity.

rs2731673; 5:176839898 (*F12/GRK6* intergenic)

Finally, the index SNP rs2731673 (MAF = 0.25) that could not be replicated in the population-based YFS cohort is located between the genes *F12* and *GRK6*. While this non-confirmation may indicate a false positive result of our GWAS, replication may have failed for some unknown reason such as being a specific result relevant for CKD patients. In the absence of another cohort (whether population-based or CKD cohort) with necessary data on OPN and genetics, we were unable to validate this any further.

The gene nearest to the locus is the *F12* gene that encodes coagulation factor XII, which, together with plasma PK, belongs to the contact activation system [67,68]. While KAL can activate factor XII (factor XIIa: active enzyme of factor XII) that, in turn promotes inflammation via the KKS, including PK [69]. Since CKD patients markedly have more inflammation and fibrosis as the joined common final path of kidney disease progression these insights and connection may encourage further validation of this locus. Inflammatory processes also play a major role in CVD and CKD patients, who are well known to suffer from excessive CVD promoting higher morbidity and mortality. In the human cardiovascular system, OPN is primarily expressed in endothelial cells, macrophages, and smooth muscle derived foam cells and can also be detected in human atherosclerotic plaques of the arterial system [70,71]. Higher serum OPN levels were found in patients with acute coronary syndrome vs chronic coronary syndrome. In coronary artery disease (CAD) patients high OPN levels were associated with rapid coronary plaque progression and in-stent restenosis [72]. OPN has been known to be associated with adverse outcomes in patients with CVD [73–75], but its function in CVD is diverse. Acute increases of OPN in CVD are associated with wound healing and neovascularization [76,77]. Chronically increased OPN, however, is associated with a poor prognosis of major adverse cardiovascular events [78].

GRK6 on the other hand is involved in blood pressure regulation and was found to have a decreased kidney expression in spontaneously hypertensive rats, providing a similar rationale [79].

Strengths and limitations

The presented analyses have several strengths and limitations: Firstly, our analyses are based on a CKD patient population of European ancestry and mostly CKD stage 3 under regular nephrologist care. While biological mechanisms may be upregulated in impaired kidney function and thus detected more easily in CKD patients, results potentially compromise generalizability to the general population as well as to other ethnicities. Secondly, we could replicate two of three identified loci in a population-based cohort of the YFS, who also applied an ELISA technique to measure OPN, thus confirming the potential to transfer findings from a CKD cohort to the general population. Regarding the third, non-replicated finding, one should await further validation of this result as regarding it as false-positive would be a premature conclusion. Thirdly, serum OPN was measured in the GCKD study from baseline samples using a state-of-the-art ELISA assay. Although serum has been validated for use in the used OPN assay, it is not the recommended sample type, because of proteolytic cleavage by thrombin during the clotting process. In contrast, OPN measurements were obtained from plasma

samples in the YFS cohort. While a comparable assay was used, differences between levels in CKD patients and YFS participants might thus be explainable by more than disease status.

Conclusions

In this first GWAS of serum OPN levels in a large CKD cohort two replicated associations on chromosome 4 were detected. One locus closest to the *SPP1* gene, as well as a locus mapping into the *KLKB1* gene, connecting OPN to its production and the KKS. Further studies are needed to fully explain OPN's role in kidney (patho)physiology and elucidate functions of OPN in connection with the KKS and possibly inflammatory processes during kidney fibrosis.

Material and methods

Ethics statement

The German Chronic Kidney Disease (GCKD) study was approved by all Ethics Committees of participating institutions in Germany that also covers the present project. It was registered in the national registry for clinical studies (DRKS 00003971; [S1 Information](#)). Written informed consent was obtained for all participants.

Study population

The GCKD study cohort consists of 5,217 adult CKD patients of European ancestry with (i) an eGFR between 30–60 mL/min per 1.73m² or (ii) an eGFR >60 mL/min/1.73 m² and 'overt' albuminuria/proteinuria at baseline [20]. At the baseline visit (2010–2012), trained personnel obtained data using a standardized questionnaire and physical examinations. Biosamples were obtained, directly processed and then stored at -80°C in a central biobank [80]. Study procedures and main baseline findings have been reported before [20,81].

Baseline variables and measurements

A standardized set of biomarkers was measured in a central certified laboratory using standardized protocols [81]. Among others, creatinine and albumin from serum and urine were quantified using an IDMS traceable methodology (Creatinine plus, Roche, Germany) and a turbidimetric method (Tina-quant, Roche, Germany) Roche/Hitachi MODULAR P, respectively.

Glomerular filtration rate (GFR) was estimated using the creatinine-based CKD-EPI formula (unit: mL/min/1.73m², [82]). UACR was calculated as measured urinary albumin/urinary creatinine (mg/g, [83]). Age and sex were self-reported at the baseline visit.

In 2015, OPN was measured from baseline serum samples of the complete GCKD study cohort using a quantitative sandwich enzyme immunoassay technique (solid-phase ELISA; Quantikine Human OPN Immunoassay DOST00 from R&D Systems (R&D Systems Europe, Abingdon, UK)). Quantification was carried out at the Institute of Clinical Chemistry and Laboratory Medicine, Greifswald, Germany. Coefficients of variation (intra-assay) were 4.5%, 5.3% and 3.5% for low, median and high levels, respectively. The inter-assay coefficient of variation was 6.4%. Reagents and secondary standards were used as recommended by the manufacturer.

Genotyping, quality control and imputation

Detailed information on genotyping and data cleaning in the GCKD study has been described previously [18]. Briefly, DNA was isolated from whole blood and genotyped at 2,612,357 variants for 5,123 GCKD participants using the Illumina HumanOmni2.5 Exome BeadChip array

(Illumina, GenomeStudio, Genotyping Module Version 1.9.4) at the Helmholtz Center Munich. Data cleaning was carried out separately for the Omni2.5 content and the exome chip content of the array.

Based on standardized protocols [84], custom written scripts (R, Perl) and Plink1.9 [85] software was used for quality control (QC) of the Omni2.5 content. Sample-based QC steps included checks of call rate, sex, heterozygosity, genetic ancestry and relatedness, leading to the exclusion of 89 samples. On the variant level, single nucleotide polymorphisms (SNPs) were excluded if the call rate was <0.96 , and whenever the assumption of the Hardy-Weinberg equilibrium was violated (p -value $<1.0E-05$). After removing SNPs on duplicate positions, the cleaned dataset contained 5,034 individuals and 2,337,794 SNPs (S1 Fig). Genotypes were then imputed using minimac3 v2.0.1 at the Michigan Imputation Server [86]. The Haplotype Reference Consortium (HRC) haplotypes version r1.1 were used as the reference panel, and Eagle 2.3 was used for phasing. The final dataset contains data of 5,034 participants and 7,750,367 high-quality autosomal bi-allelic variants (imputation quality of $R^2 \geq 0.3$, $MAF \geq 1\%$).

For the exome chip content, QC was similarly conducted [18]. In addition, checks specific for exome variants were added [87]. In brief, 96 individuals and 3,818 SNPs were removed, the latter of which had a call rate <0.95 and a Hardy-Weinberg equilibrium p -value $<1.0E-05$. The final exome chip dataset contains 5,027 participants with 226,233 variants (S1 Fig). For the exome chip association analysis, the genotypes were post-processed using zCall with a z -score threshold of six [88]. Genomic positions base on human genome build GRCh37.

Genome-wide association study of common variants

As previously reported [17,18], GWAS was conducted for GCKD participants with complete genotyping (Omni2.5), eGFR, UACR and $\log_2(OPN)$ measurement ($N = 4,897$) data using linear regression of $\log_2(OPN)$ on SNPs (additive genetic model) with a $MAF \geq 1\%$, adjusted for age, sex, $\log(eGFR)$, and $\log(UACR)$ (S1 Fig). Association analysis was performed using SNPTEST v2.5 [89]. Summary statistics were checked for quality using GWAToolbox [90] and for inflation using genomic control [91]. A genomic control correction, however, was not requested ($\lambda = 1.01$). Associations with a p -value $<5.0E-08$ were considered significant. Per chromosome, an index SNP was defined as the SNP with the lowest genome-wide p -value with a 1-Mb interval centered around this SNP. This approach was repeated until no further SNP outside the interval(s) was available passing the genome-wide significance threshold. In order to discover further independent signals, we repeated GWAS analysis for chromosomes with significant results by conditioning on the genotype of the SNP with the lowest association p -value of the respective chromosome. This procedure was repeated until no further genome-wide signal was observed.

Functional annotation of variants was conducted using ANNOVAR[92], SNIIPA [93], Open Targets Genetics [94], FAVOR [95], and RegulomeDB [96]. Regional association plots were plotted using LocusZoom v1.3 [97].

Fine-Mapping

Statistical fine-mapping [21] was carried out as previously described [17] for the two replicated SNPs within a region ± 500 kb. Approximate Bayes factors (ABFs) were then derived from the original GWAS statistics estimates. The SD prior was chosen as 0.61 because 95% of the effect size estimates fell within the -1.2 to 1.2 interval [21]. The ABF of the SNPs were used to calculate the posterior probability for each variant driving the association signal (PPA, 'causal variant'). Credible sets were determined by summing up PPA-ranked variants until the cumulative PPA was $>99\%$.

Colocalization analyses

In order to further understand the molecular mechanisms and associated phenotypes underlying the associations, we performed colocalization analyses of the OPN GWAS summary statistics related to the two replicated OPN loci with GWAS summary statistics from three other sources as outlined below. For all colocalization analyses, we used the ‘coloc.fast’ function from the R package *gtx* with default parameters and prior definitions (<https://github.com/tobyjohnson/gtx>), an implementation of an adapted version of the colocalization method introduced by Giambartolomei *et al.* [98]. We consider a positive colocalization when the posterior probability of a shared causal variant at the association locus for both traits (H4, p12) was > 0.8 .

Gene expression. First, we used GWAS summary statistics of gene expression data from the GTEx project [99] and the NEPTUNE study [100]. The eQTL data from the GTEx V8 (49 tissues) and the NEPTUNE study (NephQTL from glomerulus and tubulointerstitial kidney portions) were downloaded from the GTEx Portal (<https://www.gtexportal.org/home/>) and NephQTL web site (<http://nephqtl.org/>), respectively.

The analysis steps of colocalization have been described in detail elsewhere [17]. Firstly, GWAS summaries of GTEx and NephQTL in genomic regions $\pm 100\text{kb}$ of the two OPN SNPs were extracted. The genes in the extracted GWAS are identified and for each gene, a *cis* window of 500kb flanking the start and end of the gene are defined. Then, for every such *cis* gene window, with at least one SNP having an association p-value < 0.001 , the GWAS summaries of GTEx and NephQTL tissue as well as the OPN GWAS were extracted and used as input for colocalization analysis.

Plasma proteome. In addition, we used GWAS summary statistics of plasma proteins (pGWAS) by Sun *et al.* [22] to run colocalization analysis to identify consistent association signals between OPN and proteins with effects in *cis* as well as in *trans*. In contrast to data from GTEx and NephQTL, the genome-wide available pGWAS summary does allow the assessment of both effects.

In order to detect colocalization with *cis*-pQTLs, pGWAS summary statistics of any protein-gene region (gene region $\pm 500\text{kb}$, *cis* region) were extracted. Per OPN locus and a 100kb region around it, we checked if any of the *cis* pGWAS extracts overlapped and had a pGWAS association p-value of $< 0.05/2$ (Bonferroni correction for two OPN loci). For all hereby selected proteins, we then extracted the protein-gene-region from the OPN GWAS summary statistics and ran colocalization within the protein-gene-region.

For potential colocalization with *trans*-pQTLs, we selected all proteins with pGWAS association p-values $< 0.05/2/3,000$ (Bonferroni correction for the two OPN loci and number of proteins evaluated in pGWAS) within a 100kb region around an OPN-associated index SNP. For all hereby selected proteins, colocalization analyses were conducted within the $\pm 500\text{kb}$ region of the OPN-associated index SNP.

For colocalizing proteins, a Gene ontology (GO) enrichment analysis (<http://geneontology.org/>, [101,102]) in form of a PANTHER [103] overrepresentation test with the two annotation data sets of GO cellular component and GO molecular function (homo sapiens) as references was conducted to assess enriched categories to which identified proteins were assigned to. Overall, 20,595 human genes are mapped to various terms related to cellular component and molecular function. A category is considered enriched if both, the Bonferroni-corrected p-value of the Fisher’s exact test and the false discovery rate based on the Benjamini-Hochberg procedure, are < 0.05 .

UKB diseases. Finally, we used the GWAS from GeneAtlas database (<http://geneatlas.roslin.ed.ac.uk/>) to perform colocalization analysis for the two replicated OPN loci ($\pm 500\text{kb}$)

and all UKB binary disease traits that showed genome-wide significant associations (p-value $<5.0E-08$) in at least one of the two replicated OPN loci. Overall, GeneAtlas comprises GWAS results of 660 binary disease traits of ~450,000 UKB participants [23].

In addition, we adopted the conditional colocalization analysis approach which was firstly applied in a GWAS of plasma proteome [104]. Performing colocalization on conditionally independent association statistics could reveal true colocalization signals that were missing when using marginal association statistics in the presence of multiple independent association signals. We applied GCTA COJO Slct algorithm to identify independent association signals in the OPN region for the seven traits [105], which showed a trait association signal different from the OPN signal (H3: $p_{1.2} > 0.8$). The cleaned and imputed GCKD genotype dataset mentioned before was used as LD reference by GCTA. We set the collinearity cutoff at 0.1 to be conservative. For loci with more than 1 independent signal, an approximate conditional analysis was conducted by GCTA COJO-Cond algorithm to generate conditional association statistics conditioned on the other independent SNPs in the region [105]. Finally the colocalization analyses were performed as before for each of the independent SNPs using the conditional association statistics as input.

Aggregated rare variant testing

Overall, 4,879 GCKD participants with complete data on genotyping (Exome chip), eGFR, UACR and $\log_2(\text{OPN})$ measurements were included in the analysis of aggregated rare variant testing (S1 Fig). As previously described [106], two types of rare variant aggregation tests (burden test, sequence kernel association test [SKAT]) implemented in the R package *seqMeta* (v1.6.7, [107]) were conducted using exome chip data and $\log_2(\text{OPN})$ measurements (outcome). Per gene, variants with MAF $< 1\%$ and having a major effect on the gene product (non-synonymous, stop gain/loss, splicing; “qualifying variants”) as annotated by dbNSFP v.2.0 were aggregated [24,25]. Results were filtered to retain genes with cumulative minor allele count (MAC) ≥ 10 and with ≥ 2 contributing variants per gene. Analyses were adjusted for age, sex, $\log(\text{eGFR})$, and $\log(\text{UACR})$. To adjust for multiple testing, the statistical significance level was corrected for the number of assessed genes ($N = 17,575$) and the two conducted tests: $0.05/(2 \times 17,575) = 1.4E-06$. Moreover, analyses were repeated for significantly associated genes additionally adjusted for the two replicated OPN loci.

Replication of identified loci in Young Finns Study

The three OPN loci identified in the GWAS of GCKD participants were tested for replication in the Cardiovascular Risk in Young Finns Study (YFS) cohort. Here, plasma OPN was measured by enzyme-linked immunosorbent assay (Human Osteopontin Quantikine kit, R&D Systems, USA) from samples thawed for the first time for the assay in 2007. Samples of 2,442 participants and 546,677 genotyped SNPs were available for further analysis after QC and imputation. Further details can be found in S1 Methods.

Per selected locus, association analysis of $\log_2(\text{OPN})$ on SNP dosage (additive) was performed by fitting linear regression models adjusted for age, sex, and eGFR by using SNPTEST v2.5.4 [89]. GFR was estimated with the MDRD study equation and \log_2 -transformed prior to analysis [108]. Replication was defined by a one-sided association p-value $< 0.05/3$ (Bonferroni correction for three OPN loci).

Supporting information

S1 Fig. Flow chart showing exclusion of patients and analysis sets.
(PDF)

S2 Fig. Osteopontin (OPN, ng/mL) measurements in GCKD.

(PDF)

S3 Fig. Quantile-Quantile plot of results from GWAS of \log_2 (OPN).

(PDF)

S4 Fig. Manhattan plot of results from GWAS of \log_2 (OPN).

(PDF)

S5 Fig. Regional association plots obtained in the course of conditional analysis to identify independent signals.

(PDF)

S6 Fig. Levels of osteopontin (\log_2 -transformed) in the overall cohort and across genotypes of discovered SNPs in GCKD.

(PDF)

S7 Fig. Regional association plot for the region around rs2731673 on chromosome 5.

(PDF)

S8 Fig. Effects of rs10011284 and rs4253311 (chromosome 4) on OPN levels.

(PDF)

S9 Fig. Relationship of selected GO terms.

(PDF)

S10 Fig. Comparing summary statistics for the *KLKB1* locus from OPN GWAS (unconditional statistics, A) with respective results for the UK Biobank phenotype deep venous thrombosis (DVT, conditional statistics, B).

(PDF)

S1 Methods. The Cardiovascular Risk in Young Finns Study (YFS) cohort.

(DOCX)

S1 Information. List of institutions and investigators participating in the GCKD study.

(DOCX)

S1 Table. Genome-wide association results for common variants ($MAF \geq 0.01$) with p-value $< 1E-06$.

(XLSX)

S2 Table. Extended annotation of the three top SNPs identified in OPN GWAS in the GCKD study.

(XLSX)

S3 Table. SNPs in 99% credible sets for the replicated OPN loci.

(XLSX)

S4 Table. Colocalization analysis: results for GTEx tissues.

(XLSX)

S5 Table. Colocalization analysis: results for pQTLs, in *trans*.

(XLSX)

S6 Table. GO overrepresentation analysis results for colocalized pQTLs, in *trans*.

(XLSX)

S7 Table. Colocalization analysis: results for GeneAtlas.
(XLSX)

Acknowledgments

We are grateful for the willingness of the CKD patients to participate in the GCKD study. The enormous effort of the study personnel of the various regional centers is highly appreciated. We thank the large number of nephrologists who provide routine care for the patients and collaborate with the GCKD study. The GCKD Investigators are listed in the [S1 Information](#). A complete list of nephrologists currently collaborating with the GCKD study is available at (<http://gckd.org>).

Author Contributions

Conceptualization: Peggy Sekula, Ulla T. Schultheiss.

Data curation: Yurong Cheng, Yong Li, Nora Scherer, Franziska Grundner-Culemann, Terho Lehtimäki, Binisha H. Mishra, Olli T. Raitakari, Matthias Nauck, Kai-Uwe Eckardt, Peggy Sekula, Ulla T. Schultheiss.

Formal analysis: Yurong Cheng, Yong Li, Nora Scherer, Franziska Grundner-Culemann, Binisha H. Mishra, Peggy Sekula, Ulla T. Schultheiss.

Funding acquisition: Terho Lehtimäki, Olli T. Raitakari, Kai-Uwe Eckardt, Peggy Sekula, Ulla T. Schultheiss.

Investigation: Binisha H. Mishra, Olli T. Raitakari, Peggy Sekula, Ulla T. Schultheiss.

Methodology: Yurong Cheng, Yong Li, Nora Scherer, Franziska Grundner-Culemann, Binisha H. Mishra, Matthias Nauck, Peggy Sekula, Ulla T. Schultheiss.

Project administration: Peggy Sekula, Ulla T. Schultheiss.

Resources: Peggy Sekula, Ulla T. Schultheiss.

Software: Yurong Cheng, Yong Li, Binisha H. Mishra, Peggy Sekula, Ulla T. Schultheiss.

Supervision: Nora Scherer, Franziska Grundner-Culemann, Peggy Sekula, Ulla T. Schultheiss.

Validation: Yurong Cheng, Yong Li, Nora Scherer, Franziska Grundner-Culemann, Terho Lehtimäki, Binisha H. Mishra, Olli T. Raitakari, Matthias Nauck, Kai-Uwe Eckardt, Peggy Sekula, Ulla T. Schultheiss.

Visualization: Yurong Cheng, Yong Li, Peggy Sekula, Ulla T. Schultheiss.

Writing – original draft: Peggy Sekula, Ulla T. Schultheiss.

Writing – review & editing: Yurong Cheng, Yong Li, Nora Scherer, Franziska Grundner-Culemann, Terho Lehtimäki, Binisha H. Mishra, Olli T. Raitakari, Matthias Nauck, Kai-Uwe Eckardt, Peggy Sekula, Ulla T. Schultheiss.

References

1. Kaleta B. The role of osteopontin in kidney diseases. *Inflamm Res*. 2019; 68(2):93–102. <https://doi.org/10.1007/s00011-018-1200-5> PMID: 30456594.
2. Denhardt DT, Guo X. Osteopontin: a protein with diverse functions. *Faseb j*. 1993; 7(15):1475–82. Epub 1993/12/01. PMID: 8262332.
3. Ross FP, Chappel J, Alvarez JI, Sander D, Butler WT, Farach-Carson MC, et al. Interactions between the bone matrix proteins osteopontin and bone sialoprotein and the osteoclast integrin alpha v beta 3

- potentiate bone resorption. *The Journal of biological chemistry*. 1993; 268(13):9901–7. Epub 1993/05/05. PMID: [8486670](#).
4. Green PM, Ludbrook SB, Miller DD, Horgan CM, Barry ST. Structural elements of the osteopontin SVVYGLR motif important for the interaction with alpha(4) integrins. *FEBS Lett*. 2001; 503(1):75–9. [https://doi.org/10.1016/s0014-5793\(01\)02690-4](https://doi.org/10.1016/s0014-5793(01)02690-4) PMID: [11513858](#).
 5. Hu DD, Lin EC, Kovach NL, Hoyer JR, Smith JW. A biochemical characterization of the binding of osteopontin to integrins alpha v beta 1 and alpha v beta 5. *J Biol Chem*. 1995; 270(44):26232–8. <https://doi.org/10.1074/jbc.270.44.26232> PMID: [7592829](#).
 6. Xie Y, Sakatsume M, Nishi S, Narita I, Arakawa M, Gejyo F. Expression, roles, receptors, and regulation of osteopontin in the kidney. *Kidney Int*. 2001; 60(5):1645–57. <https://doi.org/10.1046/j.1523-1755.2001.00032.x> PMID: [11703581](#).
 7. Rabb H, Barroso-Vicens E, Adams R, Pow-Sang J, Ramirez G. Alpha-V/beta-3 and alpha-V/beta-5 integrin distribution in neoplastic kidney. *Am J Nephrol*. 1996; 16(5):402–8. <https://doi.org/10.1159/000169032> PMID: [8886177](#).
 8. Lopez CA, Hoyer JR, Wilson PD, Waterhouse P, Denhardt DT. Heterogeneity of osteopontin expression among nephrons in mouse kidneys and enhanced expression in sclerotic glomeruli. *Laboratory investigation; a journal of technical methods and pathology*. 1993; 69(3):355–63. Epub 1993/09/01. PMID: [8377476](#).
 9. Giacomelli F, Marciano R, Pistorio A, Catarsi P, Canini S, Karsenty G, et al. Polymorphisms in the osteopontin promoter affect its transcriptional activity. *Physiological genomics*. 2004; 20(1):87–96. Epub 2004/10/14. <https://doi.org/10.1152/physiolgenomics.00138.2004> PMID: [15479859](#).
 10. Salimi S, Noora M, Nabizadeh S, Rezaei M, Shahraki H, Milad MK, et al. Association of the osteopontin rs1126616 polymorphism and a higher serum osteopontin level with lupus nephritis. *Biomed Rep*. 2016; 4(3):355–60. Epub 2016/03/22. <https://doi.org/10.3892/br.2016.589> PMID: [26998275](#); PubMed Central PMCID: [PMC4774351](#).
 11. Forton AC, Petri MA, Goldman D, Sullivan KE. An osteopontin (SPP1) polymorphism is associated with systemic lupus erythematosus. *Hum Mutat*. 2002; 19(4):459. Epub 2002/04/05. <https://doi.org/10.1002/humu.9025> PMID: [11933203](#).
 12. Xu AP, Bai J, Lü J, Liang YY, Li JG, Lai DY, et al. Osteopontin gene polymorphism in association with systemic lupus erythematosus in Chinese patients. *Chin Med J (Engl)*. 2007; 120(23):2124–8. Epub 2008/01/03. PMID: [18167187](#).
 13. Fathy R, Elsayed GR, Hasan AS. Osteopontin 9250 C/T Gene Polymorphism in Egyptian Lupus Nephritis Patients. *SL Clinical Medicine: Research*. 2021; 4(1):120.
 14. Kaleta B, Krata N, Zagodzón R, Mucha K. Osteopontin Gene Polymorphism and Urinary OPN Excretion in Patients with Immunoglobulin A Nephropathy. *Cells*. 2019; 8(6). <https://doi.org/10.3390/cells8060524> PMID: [31159229](#); PubMed Central PMCID: [PMC6628186](#).
 15. Cheema BS, Iyengar S, Ahluwalia TS, Kohli HS, Sharma R, Shah VN, et al. Association of an Osteopontin gene promoter polymorphism with susceptibility to diabetic nephropathy in Asian Indians. *Clinica chimica acta; international journal of clinical chemistry*. 2012; 413(19–20):1600–4. Epub 2012/05/16. <https://doi.org/10.1016/j.cca.2012.04.028> PMID: [22584029](#).
 16. Cheema BS, Iyengar S, Sharma R, Kohli HS, Bhansali A, Khullar M. Association between Osteopontin Promoter Gene Polymorphisms and Haplotypes with Risk of Diabetic Nephropathy. *J Clin Med*. 2015; 4(6):1281–92. Epub 2015/08/05. <https://doi.org/10.3390/jcm4061281> PMID: [26239559](#); PubMed Central PMCID: [PMC4485000](#).
 17. Schlosser P, Li Y, Sekula P, Raffler J, Grundner-Culemann F, Pietzner M, et al. Genetic studies of urinary metabolites illuminate mechanisms of detoxification and excretion in humans. *Nat Genet*. 2020; 52(2):167–76. Epub 2020/01/22. <https://doi.org/10.1038/s41588-019-0567-8> PMID: [31959995](#); PubMed Central PMCID: [PMC7484970](#).
 18. Li Y, Sekula P, Wuttke M, Wahrheit J, Hausknecht B, Schultheiss UT, et al. Genome-Wide Association Studies of Metabolites in Patients with CKD Identify Multiple Loci and Illuminate Tubular Transport Mechanisms. *Journal of the American Society of Nephrology: JASN*. 2018; 29(5):1513–24. Epub 2018/03/17. <https://doi.org/10.1681/ASN.2017101099> PMID: [29545352](#); PubMed Central PMCID: [PMC5967769](#).
 19. Jing J, Ekici AB, Sitter T, Eckardt KU, Schaeffner E, Li Y, et al. Genetics of serum urate concentrations and gout in a high-risk population, patients with chronic kidney disease. *Sci Rep*. 2018; 8(1):13184. <https://doi.org/10.1038/s41598-018-31282-z> PMID: [30181573](#).
 20. Eckardt KU, Barthlein B, Baid-Agrawal S, Beck A, Busch M, Eitner F, et al. The German Chronic Kidney Disease (GCKD) study: design and methods. *Nephrol Dial Transplant*. 2012; 27(4):1454–60. Epub 2011/08/25. <https://doi.org/10.1093/ndt/gfr456> PMID: [21862458](#).

21. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol.* 2009; 33(1):79–86. Epub 2008/07/22. <https://doi.org/10.1002/gepi.20359> PMID: 18642345.
22. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. *Nature.* 2018; 558(7708):73–9. Epub 2018/06/08. <https://doi.org/10.1038/s41586-018-0175-2> PMID: 29875488; PubMed Central PMCID: PMC6697541.
23. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet.* 2018; 50(11):1593–9. Epub 2018/10/24. <https://doi.org/10.1038/s41588-018-0248-z> PMID: 30349118; PubMed Central PMCID: PMC6707814.
24. Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD, et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One.* 2013; 8(7):e68095. Epub 2013/07/23. <https://doi.org/10.1371/journal.pone.0068095> PMID: 23874508; PubMed Central PMCID: PMC3709915 alter the authors' adherence to all the PLOS ONE policies on sharing data and materials. All authors have declared that no competing interests exist.
25. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat.* 2016; 37(3):235–41. Epub 2015/11/12. <https://doi.org/10.1002/humu.22932> PMID: 26555599; PubMed Central PMCID: PMC4752381.
26. Pietzner M. Cross-platform proteomics to advance genetic prioritisation strategies. *bioRxiv.* 2021.
27. Raffield LM, Dang H, Pratte KA, Jacobson S, Gillenwater LA, Ampleford E, et al. Comparison of Proteomic Assessment Methods in Multiple Cohort Studies. *Proteomics.* 2020; 20(12):e1900278. <https://doi.org/10.1002/pmic.201900278> PMID: 32386347; PubMed Central PMCID: PMC7425176.
28. Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of applied genetics.* 2018; 59(3):253–68. Epub 2018/04/24. <https://doi.org/10.1007/s13353-018-0444-7> PMID: 29680930; PubMed Central PMCID: PMC6060985.
29. Kaleta B. Role of osteopontin in systemic lupus erythematosus. *Archivum immunologiae et therapeuticae experimentalis.* 2014; 62(6):475–82. Epub 2014/06/12. <https://doi.org/10.1007/s00005-014-0294-x> PMID: 24917428; PubMed Central PMCID: PMC4244532.
30. Rodrigues LR. SPP1 (secreted phosphoprotein 1). *Atlas Genet Cytogenet Oncol Haematol.* 2009; 13(10):733–6.
31. Scatena M, Liaw L, Giachelli CM. Osteopontin: a multifunctional molecule regulating chronic inflammation and vascular disease. *Arterioscler Thromb Vasc Biol.* 2007; 27(11):2302–9. Epub 2007/08/25. <https://doi.org/10.1161/ATVBAHA.107.144824> PMID: 17717292.
32. Gao B, Yasui T, Itoh Y, Li Z, Okada A, Tozawa K, et al. Association of osteopontin gene haplotypes with nephrolithiasis. *Kidney Int.* 2007; 72(5):592–8. Epub 2007/05/24. <https://doi.org/10.1038/sj.ki.5002345> PMID: 17519954.
33. Safarinejad MR, Shafiei N, Safarinejad S. Association between polymorphisms in osteopontin gene (SPP1) and first episode calcium oxalate urolithiasis. *Urolithiasis.* 2013; 41(4):303–13. Epub 2013/06/20. <https://doi.org/10.1007/s00240-013-0582-7> PMID: 23784265.
34. Xiao X, Dong Z, Ye X, Yan Y, Chen X, Pan Q, et al. Association between OPN genetic variations and nephrolithiasis risk. *Biomed Rep.* 2016; 5(3):321–6. Epub 2016/07/27. <https://doi.org/10.3892/br.2016.724> PMID: 27602211; PubMed Central PMCID: PMC4998211.
35. Gogebakan B, Igci YZ, Arslan A, Igci M, Erturhan S, Oztuzcu S, et al. Association between the T-593A and C6982T polymorphisms of the osteopontin gene and risk of developing nephrolithiasis. *Arch Med Res.* 2010; 41(6):442–8. <https://doi.org/10.1016/j.arcmed.2010.08.014> PMID: 21044748.
36. Li X, Liu K, Pan Y, Zhang J, Lv Q, Hua L, et al. Roles of osteopontin gene polymorphism (rs1126616), osteopontin levels in urine and serum, and the risk of urolithiasis: a meta-analysis. *Biomed Res Int.* 2015; 2015:315043. Epub 2015/02/15. <https://doi.org/10.1155/2015/315043> PMID: 25785266; PubMed Central PMCID: PMC4345067.
37. Worcester EM, Beshensky AM. Osteopontin inhibits nucleation of calcium oxalate crystals. *Ann N Y Acad Sci.* 1995; 760:375–7. <https://doi.org/10.1111/j.1749-6632.1995.tb44661.x> PMID: 7785921.
38. Nishio S, Hatanaka M, Takeda H, Aoki K, Iseda T, Iwata H, et al. Calcium phosphate crystal-associated proteins: alpha2-HS-glycoprotein, prothrombin F1, and osteopontin. *Mol Urol.* 2000; 4(4):383–90. PMID: 11156706.
39. Hedgepeth RC, Yang L, Resnick MI, Marengo SR. Expression of proteins that inhibit calcium oxalate crystallization in vitro in the urine of normal and stone-forming individuals. *Am J Kidney Dis.* 2001; 37(1):104–12. <https://doi.org/10.1053/ajkd.2001.20594> PMID: 11136174.
40. Amar A, Afzal A, Hameed A, Ahmad M, Khan AR, Najma H, et al. Osteopontin promoter polymorphisms and risk of urolithiasis: a candidate gene association and meta-analysis study. *BMC Med*

- Genet. 2020; 21(1):172. Epub 20200825. <https://doi.org/10.1186/s12881-020-01101-2> PMID: 32842990; PubMed Central PMCID: PMC7446165.
41. Inoue M, Shinohara ML. Intracellular osteopontin (iOPN) and immunity. *Immunologic research*. 2011; 49(1–3):160–72. Epub 2010/12/08. <https://doi.org/10.1007/s12026-010-8179-5> PMID: 21136203; PubMed Central PMCID: PMC3509172.
 42. Ogbureke KU, Fisher LW. Renal expression of SIBLING proteins and their partner matrix metalloproteinases (MMPs). *Kidney Int*. 2005; 68(1):155–66. Epub 2005/06/16. <https://doi.org/10.1111/j.1523-1755.2005.00389.x> PMID: 15954904.
 43. Cai M, Bompada P, Salehi A, Acosta JR, Prasad RB, Atac D, et al. Role of osteopontin and its regulation in pancreatic islet. *Biochem Biophys Res Commun*. 2018; 495(1):1426–31. Epub 2017/11/29. <https://doi.org/10.1016/j.bbrc.2017.11.147> PMID: 29180017.
 44. Moorman HR, Poschel D, Klement JD, Lu C, Redd PS, Liu K. Osteopontin: A Key Regulator of Tumor Progression and Immunomodulation. *Cancers*. 2020; 12(11). Epub 2020/11/19. <https://doi.org/10.3390/cancers12113379> PMID: 33203146; PubMed Central PMCID: PMC7698217.
 45. Rowe PS. The chicken or the egg: PHEX, FGF23 and SIBLINGs unscrambled. *Cell biochemistry and function*. 2012; 30(5):355–75. Epub 2012/05/11. <https://doi.org/10.1002/cbf.2841> PMID: 22573484; PubMed Central PMCID: PMC3389266.
 46. Rowe PS, Kumagai Y, Gutierrez G, Garrett IR, Blacher R, Rosen D, et al. MEPE has the properties of an osteoblastic phosphatonin and minihibin. *Bone*. 2004; 34(2):303–19. Epub 2004/02/14. <https://doi.org/10.1016/j.bone.2003.10.005> PMID: 14962809; PubMed Central PMCID: PMC3357088.
 47. Marks J, Churchill LJ, Debnam ES, Unwin RJ. Matrix extracellular phosphoglycoprotein inhibits phosphate transport. *J Am Soc Nephrol*. 2008; 19(12):2313–20. Epub 2008/11/14. <https://doi.org/10.1681/ASN.2008030315> PMID: 19005008; PubMed Central PMCID: PMC2588094.
 48. Fisher LW, Jain A, Tayback M, Fedarko NS. Small integrin binding ligand N-linked glycoprotein gene family expression in different cancers. *Clin Cancer Res*. 2004; 10(24):8501–11. Epub 2004/12/30. <https://doi.org/10.1158/1078-0432.CCR-04-1072> PMID: 15623631.
 49. Rappaport N, Twik M, Plaschkes I, Nudel R, Iny Stein T, Levitt J, et al. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res*. 2017; 45(D1):D877–d87. Epub 2016/12/03. <https://doi.org/10.1093/nar/gkw1012> PMID: 27899610; PubMed Central PMCID: PMC5210521.
 50. Tin A, Marten J, Halperin Kuhns VL, Li Y, Wuttke M, Kirsten H, et al. Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat Genet*. 2019; 51(10):1459–74. Epub 2019/10/04. <https://doi.org/10.1038/s41588-019-0504-x> PMID: 31578528; PubMed Central PMCID: PMC6858555.
 51. Hathaway WE, Wuepper KD, Weston WL, Humbert JR, Rivers RP, Genton E, et al. Clinical and physiologic studies of two siblings with prekallikrein (Fletcher factor) deficiency. *Am J Med*. 1976; 60(5):654–64. Epub 1976/05/10. [https://doi.org/10.1016/0002-9343\(76\)90500-3](https://doi.org/10.1016/0002-9343(76)90500-3) PMID: 1020754
 52. Hiilme E, Herlitz H, Gyzander E, Hansson L. Urinary kallikrein excretion is low in malignant essential hypertension. *J Hypertens*. 1992; 10(8):869–74. Epub 1992/08/01. PMID: 1325521.
 53. Colman RW, Schmaier AH. Contact system: a vascular biology modulator with anticoagulant, profibrinolytic, antiadhesive, and proinflammatory attributes. *Blood*. 1997; 90(10):3819–43. Epub 1997/11/14. PMID: 9354649
 54. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*. 2014; 46(6):543–50. Epub 2014/05/13. <https://doi.org/10.1038/ng.2982> PMID: 24816252; PubMed Central PMCID: PMC4064254.
 55. Suhre K, Shin SY, Petersen AK, Mohny RP, Meredith D, Wägele B, et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature*. 2011; 477(7362):54–60. Epub 2011/09/03. <https://doi.org/10.1038/nature10354> PMID: 21886157; PubMed Central PMCID: PMC3832838.
 56. Biswas N, Maihofer AX, Mir SA, Rao F, Zhang K, Khandrika S, et al. Polymorphisms at the F12 and KLKB1 loci have significant trait association with activation of the renin-angiotensin system. *BMC medical genetics*. 2016; 17:21. Epub 2016/03/13. <https://doi.org/10.1186/s12881-016-0283-5> PMID: 26969407; PubMed Central PMCID: PMC4788869.
 57. Musani SK, Fox ER, Kraja A, Bidulescu A, Lieb W, Lin H, et al. Genome-wide association analysis of plasma B-type natriuretic peptide in blacks: the Jackson Heart Study. *Circ Cardiovasc Genet*. 2015; 8(1):122–30. Epub 2015/01/07. <https://doi.org/10.1161/CIRCGENETICS.114.000900> PMID: 25561047; PubMed Central PMCID: PMC4426827.
 58. Verweij N, Mahmud H, Mateo Leach I, de Boer RA, Brouwers FP, Yu H, et al. Genome-wide association study on plasma levels of midregional-proadrenomedullin and C-terminal-pro-endothelin-1. *Hypertension*. 2013; 61(3):602–8. Epub 2013/02/06. <https://doi.org/10.1161/HYPERTENSIONAHA.111.203117> PMID: 23381795.

59. Zhang W, Jerneerén F, Lehne BC, Chen MH, Luben RN, Johnston C, et al. Genome-wide association reveals that common genetic variation in the kallikrein-kinin system is associated with serum L-arginine levels. *Thromb Haemost*. 2016; 116(6):1041–9. Epub 2016/09/23. <https://doi.org/10.1160/TH16-02-0151> PMID: 27656708; PubMed Central PMCID: PMC6215702.
60. Lamina C, Friedel S, Coassin S, Rueedi R, Younsri NA, Seppälä I, et al. A genome-wide association meta-analysis on apolipoprotein A-IV concentrations. *Hum Mol Genet*. 2016; 25(16):3635–46. Epub 2016/07/15. <https://doi.org/10.1093/hmg/ddw211> PMID: 27412012; PubMed Central PMCID: PMC5179953.
61. Schmaier AH. The contact activation and kallikrein/kinin systems: pathophysiologic and physiologic activities. *Journal of thrombosis and haemostasis: JTH*. 2016; 14(1):28–39. Epub 2015/11/14. <https://doi.org/10.1111/jth.13194> PMID: 26565070.
62. Schmaier AH, Stavrou EX. Factor XII—What's important but not commonly thought about. *Research and practice in thrombosis and haemostasis*. 2019; 3(4):599–606. Epub 2019/10/19. <https://doi.org/10.1002/rth2.12235> PMID: 31624779; PubMed Central PMCID: PMC6781921.
63. Li Y, Bezemer ID, Rowland CM, Tong CH, Arellano AR, Catanese JJ, et al. Genetic variants associated with deep vein thrombosis: the F11 locus. *Journal of thrombosis and haemostasis: JTH*. 2009; 7(11):1802–8. Epub 2009/07/09. <https://doi.org/10.1111/j.1538-7836.2009.03544.x> PMID: 19583818.
64. Mombouli JV, Vanhoutte PM. Kinins and endothelial control of vascular smooth muscle. *Annu Rev Pharmacol Toxicol*. 1995; 35:679–705. Epub 1995/01/01. <https://doi.org/10.1146/annurev.pa.35.040195.003335> PMID: 7598512.
65. Zhao Y, Qiu Q, Mahdi F, Shariat-Madar Z, Røjkjaer R, Schmaier AH. Assembly and activation of HK-PK complex on endothelial cells results in bradykinin liberation and NO formation. *Am J Physiol Heart Circ Physiol*. 2001; 280(4):H1821–9. Epub 2001/03/15. <https://doi.org/10.1152/ajpheart.2001.280.4.H1821> PMID: 11247797.
66. Costa-Neto CM, Dillenburg-Pilla P, Heinrich TA, Parreiras-e-Silva LT, Pereira MG, Reis RI, et al. Participation of kallikrein-kinin system in different pathologies. *International immunopharmacology*. 2008; 8(2):135–42. Epub 2008/01/10. <https://doi.org/10.1016/j.intimp.2007.08.003> PMID: 18182216.
67. Calafell F, Almasy L, Sabater-Lleal M, Buil A, Mordillo C, Ramírez-Soriano A, et al. Sequence variation and genetic evolution at the human F12 locus: mapping quantitative trait nucleotides that influence FXII plasma levels. *Hum Mol Genet*. 2010; 19(3):517–25. Epub 2009/11/26. <https://doi.org/10.1093/hmg/ddp517> PMID: 19933701; PubMed Central PMCID: PMC2798724.
68. Maas C, Renné T. Coagulation factor XII in thrombosis and inflammation. *Blood*. 2018; 131(17):1903–9. Epub 2018/02/28. <https://doi.org/10.1182/blood-2017-04-569111> PMID: 29483100.
69. Renné T, Stavrou EX. Roles of Factor XII in Innate Immunity. *Frontiers in immunology*. 2019; 10:2011. Epub 2019/09/12. <https://doi.org/10.3389/fimmu.2019.02011> PMID: 31507606; PubMed Central PMCID: PMC6713930.
70. O'Brien ER, Garvin MR, Stewart DK, Hinohara T, Simpson JB, Schwartz SM, et al. Osteopontin is synthesized by macrophage, smooth muscle, and endothelial cells in primary and restenotic human coronary atherosclerotic plaques. *Arterioscler Thromb*. 1994; 14(10):1648–56. <https://doi.org/10.1161/01.atv.14.10.1648> PMID: 7918316.
71. Ikeda T, Shirasawa T, Esaki Y, Yoshiki S, Hirokawa K. Osteopontin mRNA is expressed by smooth muscle-derived foam cells in human atherosclerotic lesions of the aorta. *J Clin Invest*. 1993; 92(6):2814–20. <https://doi.org/10.1172/JCI116901> PubMed Central PMCID: PMC288482. PMID: 8254036
72. Mazzone A, Parri MS, Giannessi D, Ravani M, Vaggetti M, Altieri P, et al. Osteopontin plasma levels and accelerated atherosclerosis in patients with CAD undergoing PCI: a prospective clinical study. *Coron Artery Dis*. 2011; 22(3):179–87. <https://doi.org/10.1097/MCA.0b013e3283441d0b> PMID: 21407077.
73. Abdalrhim AD, Marroush TS, Austin EE, Gersh BJ, Solak N, Rizvi SA, et al. Plasma Osteopontin Levels and Adverse Cardiovascular Outcomes in the PEACE Trial. *PLoS One*. 2016; 11(6):e0156965. Epub 2016/06/10. <https://doi.org/10.1371/journal.pone.0156965> PMID: 27284698; PubMed Central PMCID: PMC4902195.
74. Klingel K, Kandolf R. Osteopontin: a biomarker to predict the outcome of inflammatory heart disease. *Semin Thromb Hemost*. 2010; 36(2):195–202. Epub 2010/04/22. <https://doi.org/10.1055/s-0030-1251504> PMID: 20414835.
75. Yousefi K, Irion CI, Takeuchi LM, Ding W, Lambert G, Eisenberg T, et al. Osteopontin Promotes Left Ventricular Diastolic Dysfunction Through a Mitochondrial Pathway. *J Am Coll Cardiol*. 2019; 73(21):2705–18. <https://doi.org/10.1016/j.jacc.2019.02.074> PMID: 31146816; PubMed Central PMCID: PMC6546303.

76. Lok ZSY, Lyle AN. Osteopontin in Vascular Disease. *Arterioscler Thromb Vasc Biol.* 2019; 39(4):613–22. <https://doi.org/10.1161/ATVBAHA.118.311577> PMID: 30727754; PubMed Central PMCID: PMC6436981.
77. Shirakawa K, Endo J, Kataoka M, Katsumata Y, Yoshida N, Yamamoto T, et al. IL (Interleukin)-10-STAT3-Galectin-3 Axis Is Essential for Osteopontin-Producing Reparative Macrophage Polarization After Myocardial Infarction. *Circulation.* 2018; 138(18):2021–35. <https://doi.org/10.1161/CIRCULATIONAHA.118.035047> PMID: 29967195.
78. Shirakawa K, Sano M. Osteopontin in Cardiovascular Diseases. *Biomolecules.* 2021; 11(7). Epub 20210716. <https://doi.org/10.3390/biom11071047> PMID: 34356671; PubMed Central PMCID: PMC8301767.
79. Yang J, Villar VA, Armando I, Jose PA, Zeng C. G Protein-Coupled Receptor Kinases: Crucial Regulators of Blood Pressure. *J Am Heart Assoc.* 2016; 5(7). Epub 2016/07/09. <https://doi.org/10.1161/JAHA.116.003519> PMID: 27390269; PubMed Central PMCID: PMC5015388.
80. Prokosch HU, Mate S, Christoph J, Beck A, Köpcke F, Stephan S, et al. Designing and implementing a biobanking IT framework for multiple research scenarios. *Studies in health technology and informatics.* 2012; 180:559–63. Epub 2012/08/10. PMID: 22874253.
81. Titze S, Schmid M, Kottgen A, Busch M, Floege J, Wanner C, et al. Disease burden and risk profile in referred patients with moderate chronic kidney disease: composition of the German Chronic Kidney Disease (GCKD) cohort. *Nephrol Dial Transplant.* 2015; 30(3):441–51. Epub 2014/10/02. <https://doi.org/10.1093/ndt/gfu294> PMID: 25271006.
82. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med.* 2009; 150(9):604–12. <https://doi.org/10.7326/0003-4819-150-9-200905050-00006> PMID: 19414839; PubMed Central PMCID: PMC2763564.
83. Levey AS, Eckardt KU, Tsukamoto Y, Levin A, Coresh J, Rossert J, et al. Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int.* 2005; 67(6):2089–100. Epub 2005/05/11. <https://doi.org/10.1111/j.1523-1755.2005.00365.x> PMID: 15882252.
84. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc.* 2010; 5(9):1564–73. Epub 2010/11/19. <https://doi.org/10.1038/nprot.2010.116> PMID: 21085122; PubMed Central PMCID: PMC3025522.
85. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81(3):559–75. Epub 2007/08/19. <https://doi.org/10.1086/519795> PMID: 17701901; PubMed Central PMCID: PMC1950838.
86. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016; 48(10):1284–7. <https://doi.org/10.1038/ng.3656> PMID: 27571263; PubMed Central PMCID: PMC5157836.
87. Guo Y, He J, Zhao S, Wu H, Zhong X, Sheng Q, et al. Illumina human exome genotyping array clustering and quality control. *Nat Protoc.* 2014; 9(11):2643–62. Epub 2014/10/17. <https://doi.org/10.1038/nprot.2014.174> PMID: 25321409; PubMed Central PMCID: PMC4441213.
88. Goldstein JI, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M, et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics.* 2012; 28(19):2543–5. Epub 2012/07/31. <https://doi.org/10.1093/bioinformatics/bts479> PMID: 22843986; PubMed Central PMCID: PMC3463112.
89. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet.* 2010; 11(7):499–511. Epub 2010/06/03. <https://doi.org/10.1038/nrg2796> PMID: 20517342.
90. Fuchsberger C, Taliun D, Pramstaller PP, Pattaro C. GWAToolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics.* 2012; 28(3):444–5. Epub 2011/12/14. <https://doi.org/10.1093/bioinformatics/btr679> PMID: 22155946.
91. Devlin B, Roeder K. Genomic control for association studies. *Biometrics.* 1999; 55(4):997–1004. Epub 2001/04/21. <https://doi.org/10.1111/j.0006-341x.1999.00997.x> PMID: 11315092.
92. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010; 38(16):e164. Epub 2010/07/06. <https://doi.org/10.1093/nar/gkq603> PMID: 20601685; PubMed Central PMCID: PMC2938201.
93. Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmüller G. SNIIPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics.* 2015; 31(8):1334–6. Epub 2014/11/29. <https://doi.org/10.1093/bioinformatics/btu779> PMID: 25431330; PubMed Central PMCID: PMC4393511.
94. Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A, et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-

- associated loci. *Nat Genet.* 2021; 53(11):1527–33. Epub 20211028. <https://doi.org/10.1038/s41588-021-00945-5> PMID: 34711957; PubMed Central PMCID: PMC7611956.
95. Li X, Li Z, Zhou H, Gaynor SM, Liu Y, Chen H, et al. Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nat Genet.* 2020; 52(9):969–83. Epub 20200824. <https://doi.org/10.1038/s41588-020-0676-4> PMID: 32839606; PubMed Central PMCID: PMC7483769.
 96. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012; 22(9):1790–7. <https://doi.org/10.1101/gr.137323.112> PMID: 22955989; PubMed Central PMCID: PMC3431494.
 97. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010; 26(18):2336–7. Epub 2010/07/17. <https://doi.org/10.1093/bioinformatics/btq419> PMID: 20634204; PubMed Central PMCID: PMC2935401.
 98. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014; 10(5):e1004383. Epub 2014/05/17. <https://doi.org/10.1371/journal.pgen.1004383> PMID: 24830394; PubMed Central PMCID: PMC4022491.
 99. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020; 369(6509):1318–30. Epub 2020/09/12. <https://doi.org/10.1126/science.aaz1776> PMID: 32913098; PubMed Central PMCID: PMC7737656.
 100. Gillies CE, Putler R, Menon R, Otto E, Yasutake K, Nair V, et al. An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *Am J Hum Genet.* 2018; 103(2):232–44. Epub 2018/07/31. <https://doi.org/10.1016/j.ajhg.2018.07.004> PMID: 30057032; PubMed Central PMCID: PMC6081280.
 101. The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res.* 2021; 49(D1):D325–d34. Epub 2020/12/09. <https://doi.org/10.1093/nar/gkaa1113> PMID: 33290552; PubMed Central PMCID: PMC7779012.
 102. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000; 25(1):25–9. Epub 2000/05/10. <https://doi.org/10.1038/75556> PMID: 10802651; PubMed Central PMCID: PMC3037419.
 103. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 2019; 47(D1):D419–d26. Epub 2018/11/09. <https://doi.org/10.1093/nar/gky1038> PMID: 30407594; PubMed Central PMCID: PMC6323939.
 104. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurler MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet.* 2020; 52(10):1122–31. Epub 2020/09/09. <https://doi.org/10.1038/s41588-020-0682-6> PMID: 32895551; PubMed Central PMCID: PMC7610464.
 105. Yang J, Ferreira T, Morris AP, Medland SE, Madden PA, Heath AC, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet.* 2012; 44(4):369–75, s1-3. Epub 2012/03/20. <https://doi.org/10.1038/ng.2213> PMID: 22426310; PubMed Central PMCID: PMC3593158.
 106. Cheng Y, Schlosser P, Hertel J, Sekula P, Oefner PJ, Spiekerkoetter U, et al. Rare genetic variants affecting urine metabolite levels link population variation to inborn errors of metabolism. *Nature communications.* 2021; 12(1):964. Epub 2021/02/13. <https://doi.org/10.1038/s41467-020-20877-8> PMID: 33574263; PubMed Central PMCID: PMC7878905 authors declare no competing interests.
 107. Vorman A, Brody J, Chen H, Lumley T, Davis B. seqMeta: an R Package for meta-analyzing region-based tests of rare DNA variants. 2014.
 108. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med.* 1999; 130(6):461–70. Epub 1999/03/13. <https://doi.org/10.7326/0003-4819-130-6-199903160-00002> PMID: 10075613.