Identification of genetic determinants of kidney function and chronic kidney disease

Ghasemi Semeskandeh, D.

Chapter 1: General introduction

With a prevalence of more than 10% in the global adult population, chronic kidney disease (CKD) is a major public health concern and a major cause of morbidity and mortality globally [1]. While diabetes, hypertension, vascular disease, and glomerulonephritis are some of the known causes of CKD, the aetiology of CKD is often obscure [2]. The main surrogate marker for CKD is the reduced glomerular filtration rate (GFR) which can be estimated from the serum creatinine level [1]. Whereas the underlying serum creatinine is a metabolite from muscle metabolism [3] and thus may not only reflect kidney function, the creatinine-based estimated GFR (eGFRcrea) is a widely used biomarker of kidney function in the clinic, and it has a strong heritable component [4]. Kidney function declines with aging, while early stages of CKD remain asymptomatic, leading to the diagnosis of CKD usually in late stages [5]. Increased serum creatinine is not evident until roughly 50% of the renal filtration function is lost [6], making CKD a silent disease and creating a blind spot for prognosis of kidney disease [7].

Recent efforts have begun to systematically prioritize causal genes and regulatory variants. Kidney Disease Genetic Scorecard framework is shaped to identify potential drug targets and relevant cell types involved in CKD pathogenesis [8]. Despite recent therapeutic advances for CKD such as renin-angiotensin-aldosterone system (RAAS) inhibitors [9], sodium-glucose co-transporter 2 (SGLT2) inhibitors [10], and the mineralocorticoid receptor antagonist finerenone [11], the treatments remain limited. Many potential therapies such as those targeting APOL1-associated kidney disease [12], are still in experimental stages. Given the still limited therapeutic options and the incomplete understanding of CKD pathophysiology, there is a growing need to identify new biomarkers that could enable early diagnosis and potentially provide novel therapeutic targets [5]. In this context, genetic research is promising for advancing drug discovery. Recent evidence suggests that drug targets supported by human genetic findings—identified, for example, through genome-wide association studies (GWAS) or investigations of Mendelian disorders—are significantly more likely to lead to successful therapeutic development. Specifically, such targets are estimated to be 2.6 times more likely to progress through the drug development pipeline compared to targets without genetic evidence [13]. These findings highlight the potential of genetic studies in guiding the discovery of new therapeutic strategies for CKD.

Multi-omics approach to CKD

Despite growing efforts, finding early indicators of CKD has not yet been translated to clinical practice. Mass spectrometry-based metabolomic profiling allows non-invasive measurement of many small molecules in readily available blood and urine samples. In targeted metabolomics**,** hundreds of known metabolites are precisely measured. In contrast, untargeted metabolomics can explore thousands of metabolites and identify new ones but without precise measurement. Both approaches succeed in studying kidney disease, offering complementary strengths for biomarker discovery and quantification [14].

Proteomics has been increasingly employed in kidney disease. The goal of using assays of select peptides and proteins is to improve diagnosis and risk prognostication. They provide pathophysiologic insight of the cause and metabolic consequences of kidney disease, both in blood and urine, which are critical for the clinical practice of nephrologists [15]. Plasma proteins hold significant potential as novel disease biomarkers, and their genetic underpinnings offer valuable insights into underlying networks and causal pathways, pointing to novel drug targets [16]. For example, clinical assessment of kidney function frequently employs biomarkers such as GFR, estimated from serum cystatin C levels, to evaluate renal filtration capacity. Complement factors and parathyroid hormone, also measured in serum, provide insights into the etiological and metabolic consequences of kidney disease, respectively. Urinary albumin excretion is a well-established marker of severity of kidney damage and can aid in elucidating underlying renal disease processes.

Current proteomic tools permit the high-throughput analysis of the blood proteome in large cohorts. While urine and kidney tissue analyses are important for kidney disease and large-scale results in these areas are pending, the extensive body of epidemiological research are built upon blood proteomics as the most frequently employed biofluid in clinical settings. Numerous proteins associated with cross-sectional measures of kidney function, as well as with the longitudinal risk of CKD progression have been identified. Representative signals include an association between levels of testican-2 and favorable kidney prognosis and an association between levels of TNFRSF1A and TNFRSF1B and worse kidney prognosis [15]. Given that, kidney function significantly influences blood protein levels, complicating the identification of causal roles for these proteins in kidney disease. Nevertheless, large-scale proteomic studies aim to develop improved diagnostics and therapeutic targets. Expanding these studies to larger, more diverse patient populations will facilitate the discovery of generalizable biomarkers and treatment strategies.

In the context of a targeted multi-omics approach to CKD, bile acids (BAs) represent a biologically plausible and underexplored molecular class with potential relevance to renal pathology. BAs are synthesized from cholesterol, helping digestion and absorption of fat in humans. The kidney expresses BA receptors, including farnesoid X receptor (FXR) and G protein-coupled bile acid receptor 1 (TGR5), implicated in metabolic regulation [17]. Renal BA transporters manage BA reabsorption and excretion, with alterations occurring in cholestasis. Disruptions in BA levels and FXR expression accompany various renal diseases [18]. Bile cast nephropathy, a severe form of acute kidney injury, exemplifies the toxic impact of bilirubin and BAs on renal tubules [19]. A comprehensive understanding of BA-kidney interactions is essential for developing targeted therapies and enhancing patient prognosis.

General epidemiology of CKD

One of the recent GWAS meta-analysis of eGFRcrea was conducted by CKD Genetics (CKDGen) Consortium, resulting in 264 associated genetic loci at which the lead variants explained approximately 20% of the genetic heritability of eGFRcrea [20]. A significant challenge in GWAS for eGFRcrea lies in differentiating the genetic influences on biomarker metabolism from those directly affecting kidney function. While blood urea nitrogen (BUN) offers an alternative kidney function biomarker [21], it replicated only 147 of the 264 associations identified in the GWAS on eGFRcrea (**Figure 1a**) [20]. Although GFR estimated by serum cystatin C (eGFRcys) might provide a more accurate reflection of GFR (**Figure 1b**), apart from the limited samples for GWAS and cost considerations [22], it can be influenced by factors beyond GFR itself, such as inflammation, obesity, and diabetes [23]. Overall, a recent GWAS meta-analysis that combined results from the CKDGen with data from the UK Biobank, identified 202 of 424 eGFRcrea-associated loci as overlapping across all three kidney function markers.

With the world-wide GWAS efforts, the number of genes and variants underneath association signals is the main challenge. A trans-ancestry GWAS meta-analysis study [20] was conducted on eGFR values derived from 121 GWAS encompassing 765,348 individuals of different ancestries. Primarily, they discovered 308 loci containing at least one eGFR-associated SNP at genome-wide significance ($P<5\times10^{-8}$), of which 200 were not reported by previous GWAS of eGFR. The minor alleles across index SNPs showed both decreasing and increasing effects on eGFR, with larger effects observed for lower-frequency SNPs. The 308 index SNPs explained

7.1% of the eGFR variance, nearly doubling recent GWAS-based estimates, and 19.6% of eGFR genetic heritability (h^2=39%, 95% credible interval=32%, 47%), estimated in a participating general-population-based pedigree study. The effects of index SNPs were largely homogeneous across studies and ancestry groups.

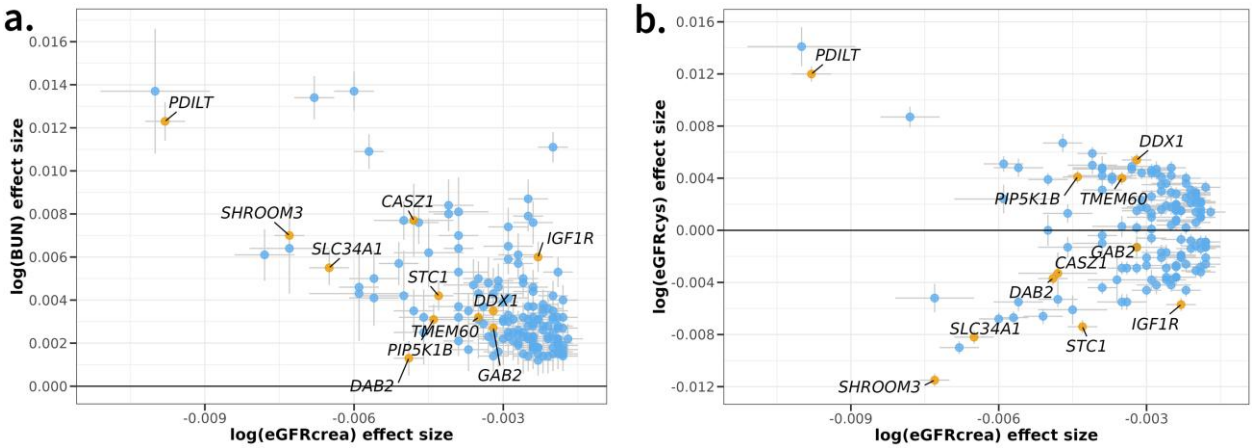

Figure 1. Supporting alternative biomarker association for 147 loci. *Highlighted loci in orange are those that showed a signal in the CHRIS study (see* ***Chapter 2****), otherwise indicated in blue. (a) Scatterplot comparing effect sizes for log(eGFRcrea) and log(BUN) with 95% confidence intervals. (b) Scatterplot comparing effect sizes for log(eGFRcrea) and log(eGFRcys). Summary statistics for eGFRcrea and BUN were obtained from trans-ethnic GWAS meta-analysis by CKDGen* [20]*. Summary statistics for eGFRcys were obtained from a meta-analysis of CKDGen trans-ethnic and UK Biobank* [21]*.*

Secondly, they investigated the effects of the eGFR index SNPs on CKD in CKDGen studies (*n*=625,219, including 64,164 CKD cases). GWAS meta-analysis of CKD identified 23 genome-wide-significant loci, including 17 likely relevant for kidney function (*SDCCAG8*, *LARP4B*, *DCDC1*, *WDR72*, *UMOD–PDILT*, *MYO19*, *AQP4*, *NFATC1*, *PSD4*, *HOXD8*, *NRIP1*, *SHROOM3*, *FGF5*, *SLC34A1*, *DAB2*, *UNCX* and *PRKAG2*). They found that 224 of the 264 replicated eGFR index SNPs were associated with CKD at nominal significance level, pertaining 130 likely relevant for kidney function.

Thirdly, they demonstrated that a lower genetic risk score based on the combined effect of the 147 eGFR index SNPs, reflecting genetically lower eGFR, was associated with higher odds ratios of clinically diagnosed chronic renal failure, glomerular diseases, acute renal failure and hypertensive diseases in the UK Biobank (*n*=452,264).

Finally, they assessed genome-wide genetic correlations of eGFR associations with 748 complex traits and diseases. Focusing on 37 significant correlations, there were negative correlations between eGFR and serum creatinine, serum citrate, serum urate, anthropometric

traits including lean mass and physical fitness such as left-hand grip strength. While the inverse correlation with muscle-mass-related traits probably reflects higher creatinine generation leading to lower creatinine-based eGFR, the correlations with citrate and urate levels probably reflect reduced filtration function. Similarly, the positive correlation between eGFRcrea with GFR estimated from cystatin C (**Figure 1b**) emerges decreased kidney function. Therefore, significant genetic correlations with eGFRcrea reflect the two biological components that govern serum creatinine concentrations: its excretion via the kidney and its generation in muscle.

Genetic and molecular data resources

The CHRIS study

The Cooperative Health Research in South Tyrol (CHRIS) study [24] is a prospective study set up in 2011 in the Middle and Upper Val Venosta/Vinschgau, South Tyrol, Italy. Recruited into the studies were 13,393 participants from thirteen municipalities, each one characterized by a central territory, small villages and scattered mountain farms. Settlements are located at an altitude of 600 to 2000 meters above sea level. Participants cover more than one third of the whole population of the target region. The study protocol [24] as well as details on the biological sampling and measurements [25], electrocardiographic and smoking measurements [26], and neurological questionnaires [27-29], have been previously described. At the study center, all participants underwent physical examinations, including anthropometry, electrocardiographic analysis, blood pressure measurement, and tremor assessment. They also underwent a structured clinical interview and completed self-administered questionnaires. Additionally, information on medication use during the seven days prior to the study center visit was collected by scanning the barcodes of drug boxes brought by the participants to the interview and classified using the Anatomical Therapeutic Chemical (ATC) coding system. Administration mode, frequency, and duration of therapy were registered for each scanned drug. Ninety standard blood and urine parameters were measured [25] along with targeted and untargeted metabolomic analysis for a subset of the study population [30, 31]. All individuals underwent genotyping and 3600 of them whole exome sequencing. To date, high quality imputed genotypes are available for all study participants at ~20 million genetic variants. Microbiome data in a subset have also been collected [32, 33]. Finally, an extensive biobank has been created that stores several aliquots of serum, plasma, urine, DNA, and buffy coats for all participants [24].

Genetic data

Genomic data covering one million variants for the study was generated using Illumina HumanOmniExpressExome or Omni2.5Exome arrays on GRCh37. Variants with low GenTrain scores (<0.6), cluster separation scores (<0.4), or call rates (<80%), violating Hardy-Weinberg equilibrium ($P<10^{-6}$), and monomorphic ones were discarded. Further quality control excluded samples with call rates below 98%. After phasing the genotypes, the genotype data were imputed based on the HRC and TOPMed reference panels. Whole exome sequencing (WES) was also performed on 3840 participants using the xGen® Exome Research Panel v1.0, retaining 3600 samples after quality control as detailed previously [31, 34]. WES data served as a reference panel for genotype imputation in the larger CHRIS study, with combined variants from both genotyping and WES. As detailed elsewhere [34], genotypes were phased and imputation was conducted, discarding variants with an imputation quality index Rsq <0.3.

Metabolomics data

For the subset of 7252 participants of the CHRIS study, targeted metabolomics analysis was conducted on the serum samples with the AbsoluteIDQ p180 kit (Biocrates Life Sciences AG, Innsbruck, Austria). Normalization and quality controls of the 188 measured metabolites are described elsewhere [30]. To increase sample homogeneity, pregnant women and individuals of non-European descent were excluded. QC left 175 high-quality metabolites available on 6642 samples.

Proteomics data

From the in total 13,393 participants of the CHRIS study 3,632, participating to the study between August 2011 and August 2014, were selected for high throughput mass spectrometry-based quantification of their plasma proteome based on Scanning SWATH [35]. The biological sampling and measurements, quality controls, batch effect removals taken place on the proteins have been described elsewhere [36]. The 148 highly abundant protein groups involved immunoglobulins, coagulation factors, metabolic proteins, and components of the innate immune system. Of the 148 plasma proteins consistently quantified, 134 could be assigned to 25 different functional classes. The most prominent were immunoglobulins (n=27), protease inhibitors (22), serine proteases (13), components of the complement system (10) and apolipoproteins (9). The final data set used for this study was log2-transformed protein concentrations.

Bile acids profiling

Bile acid (BA) analysis was performed from serum samples of 1336 individuals in the MICROS study by liquid chromatography-tandem mass spectrometry (LC-MS/MS) as previously described [37]. This resulted in 14 quantitative and 4 binary BAs, which were qc-ed and values below LOD were imputed. Some of the primary and secondary BAs consists of cholic acid (CA), chenodeoxycholic acid (CDCA), and their taurine- or glycine-bound derivatives (TCA and TCDCA, GCA and GCDCA), deoxycholic acid (DCA), a CA derivative, and lithocholic acid (LCA), a CDCA derivative, being the most prevalent [38]. All BAs considered as quantitative traits were log10-transformed.

Research objectives and rationale

The overall aim of this project was to improve the characterization of genetic loci associated with kidney function and CKD, with a specific focus on those loci previously identified through GWAS. While GWAS have identified hundreds of loci associated with complex traits such as kidney function, the biological mechanisms underlying many of these associations remain poorly understood. These genetic loci often contain multiple candidate genes, few of which have been thoroughly investigated, and mechanistic models remain incomplete.

Given this context, our study focused on refining the understanding of loci that have shown robust associations with kidney function, particularly those implicated through overlapping associations with eGFRcrea, BUN, eGFRcys. This approach aimed to prevent from pursuing loci that may be related merely to biomarker metabolism, thereby prioritizing those most likely to reflect kidney function itself.

Using the CHRIS study as the primary resource—owing to its rich genomic, metabolomic, and proteomic datasets—this project aimed to identify and further characterize the effects of these loci. As shown in **Figure 2**, the study sought to explore potential mediating pathways linking genetic variants to kidney function using clinical traits and omics data available in the CHRIS study and in publicly available datasets such as GTEx and deCODE studies. Moreover, among significant loci showing evidence of mediation, we conducted haplotype-based analyses using WES data to identify functional genetic profiles across multiple phenotypes. Later, we aimed to develop a systematic analytical pipeline for the in-depth characterization of multiple loci with

similar approaches. In the end, leveraging the unique data available in the CHRIS study, we investigated the genetic basis of primary and secondary BAs, with particular attention to potential sex-specific genetic effects on BA variability.

In summary, this project was designed to:

- Define kidney-related phenotypes within the CHRIS study.
- Identify kidney function-associated loci using GWAS in the CHRIS study.
- Investigate mediating biological pathways through integrated analysis of phenotypic, metabolic, and proteomic data.
- Apply haplotype-based methods integrating GWAS and WES data to identify subgroups of the CHRIS population with distinct phenotypic profiles.
- Develop a generalizable analytical pipeline for the systematic study of kidney function loci.
- Explore genetic architecture of bile acid metabolism in the CHRIS study.

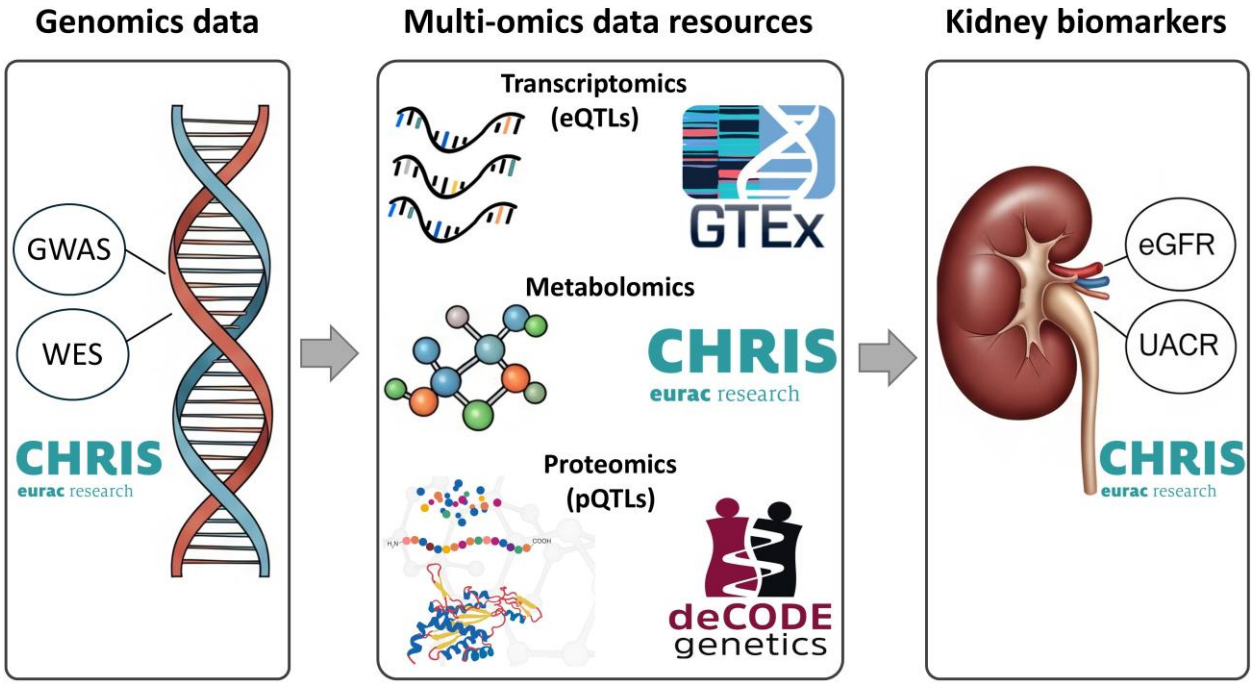

Figure 2. Study overview. *Flowchart depicting the integration of multi-omics data (genomics, transcriptomics, proteomics, and metabolomics) from various studies (CHRIS, GTEx, deCODE) to identify and understand kidney phenotypes (eGFR, UACR).*

Thesis structure

In chapter 2, we identify loci associated with eGFRcrea from the CKDGen GWAS meta-analysis [20] in the CHRIS study. We evaluate linkage disequilibrium (LD) for haplotype analysis and perform mediation analyses with various biochemical parameters to identify potential interacting pathways with local environment. In chapter 3, we reconstruct haplotypes including functional variants at the *FAM47E-SHROOM3* locus, where serum magnesium mediates the association with kidney function, and explore their associations with various clinical traits, metabolites, and protein concentrations. We identify specific phenotypic and metabolic signatures associated with distinct haplotype groups via hierarchical clustering. In chapter 4, we introduce an efficient bioinformatic pipeline for executing haplotype-based association analysis on GWAS-identified genetic loci in populations enriched with molecular omics data. This user-friendly tool generates reports summarizing variant information, constructed haplotypes, and association results. In chapter 5, we investigate the genetic architecture of 18 BAs acids in blood plasma of a subset of the CHRIS study using GWAS meta-analysis and sex-stratified GWAS. This chapter addresses the understudied area of genetic regulation of BAs in the general population, identifying associations with both common and rare variants. In chapter 6, we discuss the main findings of the thesis and provide insights into further research.

References

1. Rovin, B.H., et al., *KDIGO 2024 Clinical Practice Guideline for the management of LUPUS NEPHRITIS.* Kidney International, 2024. **105**(1): p. S1-S69.
2. Mazzieri, A., et al., *Molecular Targets of Novel Therapeutics for Diabetic Kidney Disease: A New Era of Nephroprotection.* International Journal of Molecular Sciences, 2024. **25**(7): p. 3969.
3. Patel, S.S., et al., *Serum creatinine as a marker of muscle mass in chronic kidney disease: results of a cross-sectional study and review of literature.* Journal of cachexia, sarcopenia and muscle, 2013. **4**: p. 19-29.
4. ArpegAard, J., et al., *Comparison of heritability of cystatin C-and creatinine-based estimates of kidney function and their relation to heritability of cardiovascular disease.* Journal of the American Heart Association, 2015. **4**(1): p. e001467.
5. Lin, J.-s., et al., *Proteomic profiling of longitudinal changes in kidney function among middle-aged and older men and women: the KORA S4/F4/FF4 study.* BMC Medicine, 2023. **21**(1): p. 245.
6. Mischak, H., et al., *Proteomic biomarkers in kidney disease: issues in development and implementation.* Nature Reviews Nephrology, 2015. **11**(4): p. 221-232.

7. Sanchez-Niño, M.D., et al., *Clinical proteomics in kidney disease as an exponential technology: heading towards the disruptive phase.* Clinical kidney journal, 2017. **10**(2): p. 188-191.
8. Liu, H., et al., *Kidney multiome-based genetic scorecard reveals convergent coding and regulatory variants.* Science, 2025. **387**(6734): p. eadp4753.
9. Bhandari, S., et al., *Renin–Angiotensin System Inhibition in Advanced Chronic Kidney Disease.* New England Journal of Medicine, 2022. **387**(22): p. 2021-2032.
10. Terzo, C., et al., *Investigational new drugs for the treatment of chronic renal failure: an overview of the literature.* Expert Opinion on Investigational Drugs, 2024. **33**(4): p. 319-334.
11. Bakris, G.L., et al., *Effect of Finerenone on Chronic Kidney Disease Outcomes in Type 2 Diabetes.* New England Journal of Medicine, 2020. **383**(23): p. 2219-2229.
12. Hung, A.M., et al., *Genetic Inhibition of APOL1 Pore-Forming Function Prevents APOL1-Mediated Kidney Disease.* Journal of the American Society of Nephrology, 2023. **34**(11).
13. Minikel, E.V., et al., *Refining the impact of genetic evidence on clinical success.* Nature, 2024. **629**(8012): p. 624-629.
14. Moritz, L., et al., *A systematic review of metabolomic findings in adult and pediatric renal disease.* Clinical Biochemistry, 2023: p. 110703.
15. Schlosser, P., M.E. Grams, and E.P. Rhee, *Proteomics: progress and promise of high-throughput proteomics in chronic kidney disease.* Molecular & Cellular Proteomics, 2023. **22**(6).
16. Kutalik, Z. and M. Robinson, *48th European Mathematical Genetics Meeting (EMGM) 2020. Lausanne, Switzerland, April 16–17, 2020.* Human Heredity, 2020. **84**(4/5): p. 203-229.
17. Li, S., C. Li, and W. Wang, *Bile acid signaling in renal water regulation.* American Journal of Physiology-Renal Physiology, 2019. **317**(1): p. F73-F76.
18. Yang, J., M. Pontoglio, and F. Terzi, *Bile Acids and Farnesoid X Receptor in Renal Pathophysiology.* Nephron, 2024: p. 1-13.
19. El Chediak, A., K. Janom, and S.H. Koubar, *Bile cast nephropathy: when the kidneys turn yellow.* Renal Replacement Therapy, 2020. **6**(1): p. 15.
20. Wuttke, M., et al., *A catalog of genetic loci associated with kidney function from analyses of a million individuals.* Nat Genet, 2019. **51**(6): p. 957-972.
21. Stanzick, K.J., et al., *Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals.* Nature Communications, 2021. **12**(1): p. 4350.
22. Mayanja, R., et al., *Genome-wide association analysis of cystatin-C kidney function in continental Africa.* EBioMedicine, 2023. **95**.
23. Stevens, L.A., et al., *Factors other than glomerular filtration rate affect serum cystatin C levels.* Kidney international, 2009. **75**(6): p. 652-660.
24. Pattaro, C., et al., *The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results.* Journal of Translational Medicine, 2015. **13**(1): p. 348.
25. Noce, D., et al., *Sequential recruitment of study participants may inflate genetic heritability estimates.* Hum Genet, 2017. **136**(6): p. 743-757.

26. Murgia, F., et al., *Effects of smoking status, history and intensity on heart rate variability in the general population: The CHRIS study.* PloS one, 2019. **14**(4): p. e0215053.
27. Zanigni, S., et al., *Association between restless legs syndrome and migraine: a population‐based study.* European Journal of Neurology, 2014. **21**(9): p. 1205-1210.
28. Giannini, G., et al., *Association between restless legs syndrome and hypertension: a preliminary population‐based study in South Tyrol, Italy.* European Journal of Neurology, 2014. **21**(1): p. 72-78.
29. Melotti, R., et al., *Structural consistency of the Pain Sensitivity Questionnaire in the Cooperative Health Research in South Tyrol (CHRIS) population-based study.* The Journal of Pain, 2018. **19**(12): p. 1424-1434.
30. Verri Hernandes, V., et al., *Age, Sex, Body Mass Index, Diet and Menopause Related Metabolites in a Large Homogeneous Alpine Cohort.* Metabolites, 2022. **12**(3): p. 205.
31. Emmert, D.B., et al., *Genetic and Metabolic Determinants of Atrial Fibrillation in a General Population Sample: The CHRIS Study.* Biomolecules, 2021. **11**(11): p. 1663.
32. Antonello, G., et al., *Smoking and salivary microbiota: a cross-sectional analysis of an Italian alpine population.* Scientific Reports, 2023. **13**(1): p. 18904.
33. Motta, B.M., et al., *Microbiota, type 2 diabetes and non-alcoholic fatty liver disease: protocol of an observational study.* Journal of Translational Medicine, 2019. **17**: p. 1-12.
34. König, E., et al., *Whole Exome Sequencing Enhanced Imputation Identifies 85 Metabolite Associations in the Alpine CHRIS Cohort.* Metabolites, 2022. **12**(7): p. 604.
35. Messner, C.B., et al., *Ultra-fast proteomics with Scanning SWATH.* Nat Biotechnol, 2021. **39**(7): p. 846-854.
36. Dordevic, N., et al., *Extensive modulation of the circulating blood proteome by hormonal contraceptive use across two population studies.* Commun Med, 2025. **5**(1): p. 131.
37. Scherer, M., et al., *Rapid quantification of bile acids and their conjugates in serum by liquid chromatography-tandem mass spectrometry.* J Chromatogr B Analyt Technol Biomed Life Sci, 2009. **877**(30): p. 3920-5.
38. Chiang, J.Y., *Bile acid metabolism and signaling.* Compr Physiol, 2013. **3**(3): p. 1191-212.