nature medicine

Supplementary information

https://doi.org/10.1038/s41591-025-03872-8

The proteogenomic landscape of the human kidney and implications for cardio-kidney-metabolic health

In the format provided by the authors and unedited

Supplementary Note

Sample preparation for liquid chromatography-mass spectrometry

Kidney lysate samples were prepared on the Hamilton Vantage mostly as previously described[1,2]. Briefly, samples were suspended in a lysis buffer (3% SDS, 50mM HEPES pH 8.5, 75mM NaCl + protease inhibitors), and then reduced with 5 mM dithiothreitol, alkylated with 15 mM iodoacetamide and quenched with 5 mM dithiothreitol at 25°C for 30 min each.

Proteins were purified by using single-pot solid phase enhanced sample preparation (SP3) beads. Proteins were bound to beads at 75% acetonitrile (ACN) at 10 ug beads to 1 ug protein ratio for 18 min. Protein-bound beads were then washed twice with 70% ethanol and followed by a 100% ACN wash. Beads were then resuspended in digestion buffer (50 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, 10 mM calcium chloride) and digested overnight with 33.3 ng/µL Trypsin/LysC mix (Promega) at 37ºC, 1000 rpm.

Supernatants, which contained digested peptides, were transferred to fresh tubes and were labeled with TMTpro™ 18plex reagent set (ThermoFisher Scientific) at 5:1 TMT (Tandem Mass Tag):peptide ratio. Samples were mixed and incubated for 1 hour at 25°C. Samples were then quenched at 0.5 % hydroxylamine, mixed and incubated at room temperature for 15 minutes. A small amount of each sample was combined and analyzed via liquid chromatography-mass spectrometry (LC-MS) to confirm the labeling efficiency and ratio balancing across the channels before combining into one set. Following labeling efficiency and quality checks, the samples were subjected to the peptide clean-up protocol. Here, peptides were bound to beads at 75% ACN, washed twice with 95% isopropanol (IPA), followed by a 100% IPA wash. Peptide-bound beads were blow dried by ultra-pure nitrogen, and eluted twice with 5% ACN.

To achieve deeper coverage, the cleaned-up TMT labeled peptides were partially dried down to and resuspended in 5% acetonitrile, 10 mM ammonium bicarbonate, pH 8.0, sonicated for 10 min and then

fractionated by medium pH reverse phase HPLC (Zorbax 300Extend C18, 4.6 × 250 mm column, 5 μm particle size, Agilent) at 25°C. Peptides were eluted with a gradient with initial starting condition of 100% buffer A (5% ACN, 10 mM ammonium bicarbonate) and 0% buffer B (95% ACN, 10 mM ammonium bicarbonate). Buffer B was increased to 35% over 60 minutes, then ramped up to 100% B in 6 seconds, where it held for 5 minutes. Buffer B was then decreased to 0% over 6 seconds and held for 10 minutes to re-equilibrate the column to original conditions. The samples were fractionated into 96 fractions, then pooled into 24 fractions, as previously described. The fractions were dried under vacuum and resuspended in 5% ACN/5% formic acid and approximately 1 μg per sample was injected for analysis on an Orbitrap Eclipse equipped as described below.

LC-MS analysis

Peptides were analyzed on an Orbitrap Eclipse mass spectrometer equipped with FAIMS Pro (Thermo Fisher Scientific) coupled to an Ultimate 3000 (Thermo Fisher Scientific) liquid chromatography system as described[3] with some modifications. In brief, peptides were separated on an IonOpticks Aurora microcapillary column (75 μm inner diameter, 25 cm long, C18 resin, 1.6 μm, 120 Å). The total LC-MS run length for each sample was 185 min including a 165-min gradient from 6% to 35% acetonitrile in 0.1% formic acid. The flow rate was 300 nl / min, and the column was heated at 60 °C. Four different FAIMS Pro compensation voltages (−40, −50, −60 and −70 V) was used and the data was collected in a data-dependent acquisition (DDA) mode. Each of the four compensation voltages had a 1.25 s cycle time. High-resolution precursor (MS1) scans were acquired in the Orbitrap mass analyzer with a resolution of 120,000 and m/z scan range 400 to 1600. The automatic gain control (AGC) target was set to standard with 'Auto' max injection time and RF lens value of 30%. Dynamic exclusion was enabled (after 1 time) with a duration of 60 s and a mass tolerance of 10 ppm. For MS2 spectra, ions were isolated with the quadrupole mass filter using a 0.7 m/z isolation window. The MS2 product ion population was analyzed in the quadrupole ion trap (CID, AGC 1×10^4, normalized

collision energy 35%, and max injection time of 35 ms) and the MS3 scan was analyzed in the Orbitrap (HCD, 50,000 resolution, AGC 1×10^5, max injection time 200 ms, normalized collision energy 45%) and SPS precursor ranges were set to 455–1,600 m/z. Up to ten fragment ions from each MS2 spectrum were selected for MS3 analysis using synchronous precursor selection (SPS). The human uniprot proteome (version UP000005640) was used for real time search and a maximum of two missed cleavages and one variable modification was allowed. The maximum search time was set to 35 ms and a cross-correlation score of 1, delta cross-correlation score of 0.1 and precursor of 10 ppm for charge state 2, 3 and 4 was used.

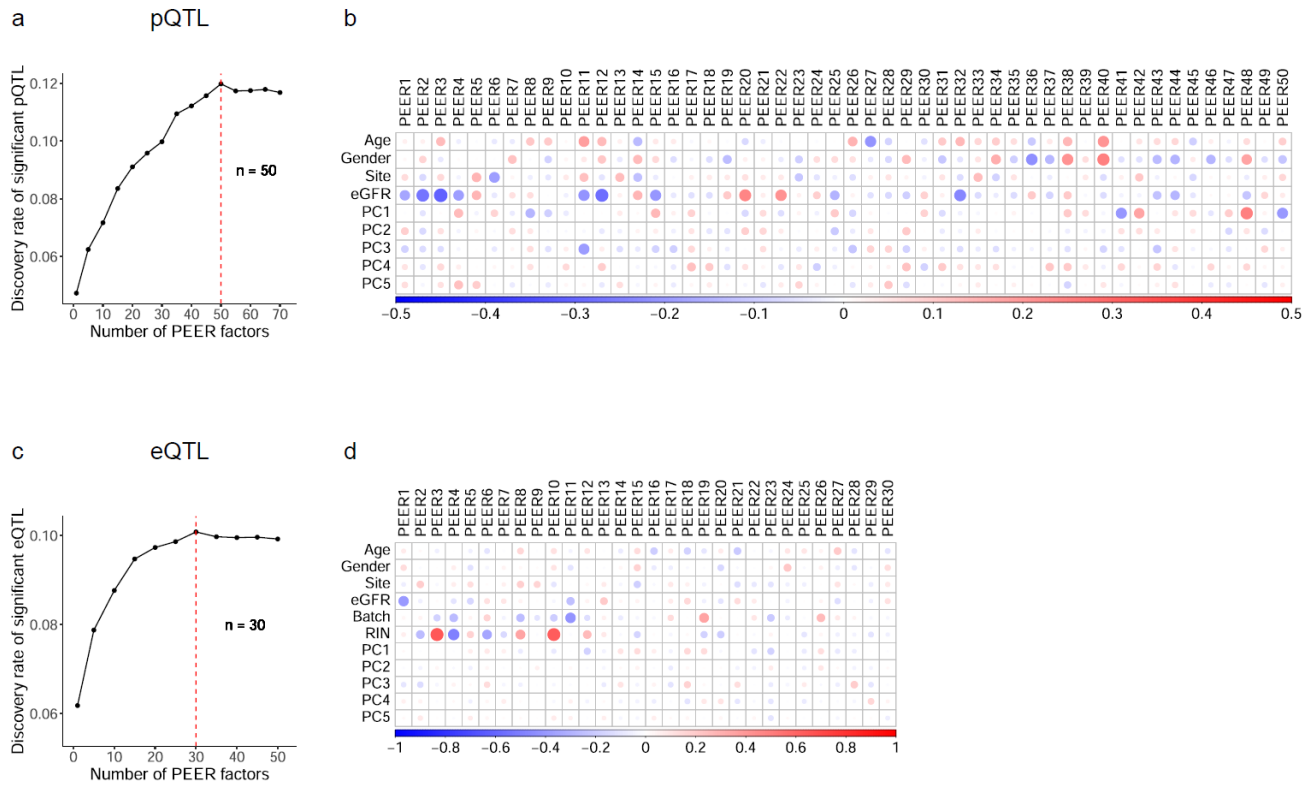
Proteomics data analysis for LC-MS

The data was analyzed using Masspike software licensed from Harvard (GFY core version 3.8) with a built-in statistical package[3]. Thermo Fisher Scientific. raw files were converted to mzXML using ReAdW.exe. MS2 spectra assignment was performed using the SEQUEST algorithm[4] v.28 by searching the data against the reference proteomes for Homo Sapien along with common contaminants such as human keratins and trypsin (same as the Real-Time-Search database). MS/MS spectra were matched with fully tryptic peptides from this composite dataset using a precursor ion tolerance of 20 ppm and a product ion tolerance of 1Da. TMT modification of peptide N-termini and lysine residues (+304.2071 Da) and carbamidomethylation of cysteine residues (+57.02146 Da) were set as static modifications. Oxidation of methionine residues (+15.99492 Da) was set as a variable modification. Peptide spectral matches were filtered to a 1% false discovery rate (FDR) using linear discriminant analysis (LDA) as previously described[5]. Non-unique peptides that matched to multiple proteins were assigned to proteins that contained the largest number of matched redundant peptides sequences using the principle of Occam's razor[5]. Quantification of TMT reporter ion intensities was performed by extracting the most intense ion within a 0.003 m/z window at the predicted m/z value for each reporter ion. Peptide level intensities and the corresponding reporter ion scans were analyzed using the

previously described framework for multiplexed mass spectrometry proteomics analysis[3].

Reference for Supplementary Note

1. Gaun, A. *et al.* Automated 16-Plex Plasma Proteomics with Real-Time Search and Ion Mobility Mass Spectrometry Enables Large-Scale Profiling in Naked Mole-Rats and Mice. *J Proteome Res* **20**, 1280-1295 (2021).
2. Nguyen, T. *et al.* Differential nuclear import sets the timing of protein access to the embryonic genome. *Nat Commun* **13**, 5887 (2022).
3. O'Brien, J.J. *et al.* A data analysis framework for combining multiple batches increases the power of isobaric proteomics experiments. *Nat Methods* **21**, 290-300 (2024).
4. Eng, J.K., McCormack, A.L. & Yates, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**, 976-89 (1994).
5. Ting, L., Rad, R., Gygi, S.P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods* **8**, 937-40 (2011).



Supplementary Fig. 1 Impact of PEER factors on pQTL and eQTL discovery.

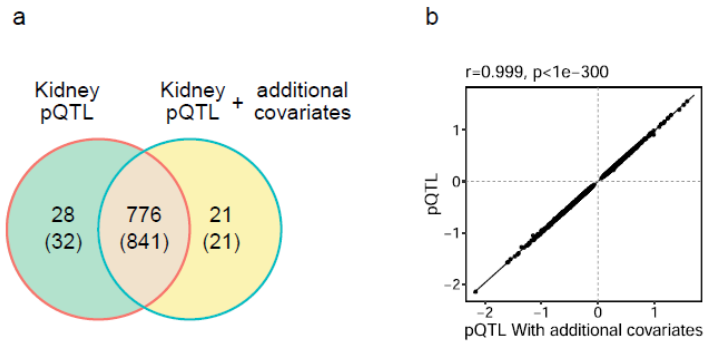
(a) Impact of PEER factor count on pQTL discovery: This plot demonstrates how the inclusion of different numbers of PEER factors affects the discovery rate of pQTLs, with the y-axis representing the number of identified pQTLs and the x-axis showing the number of PEER factors used.

(b) Correlation of PEER factors and clinical variables, and genotyping PCA across 321 pQTL samples. Pearson's correlation coefficient was calculated and shown in red (positive correlation) and blue (negative correlation).

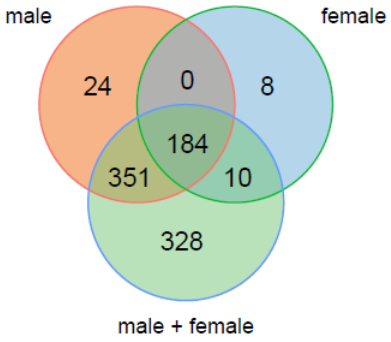
(c) Impact of PEER factor count on eQTL discovery: This plot demonstrates how the inclusion of different numbers of PEER factors affects the discovery rate of eQTLs, with the y-axis representing the number of identified eQTLs and the x-axis showing the number of PEER factors used.

(d) Correlation of PEER factors and clinical, technical variables, and genotyping PCA across 311 eQTL samples. Pearson's correlation coefficient was calculated and shown in red (positive correlation) and blue (negative correlation).

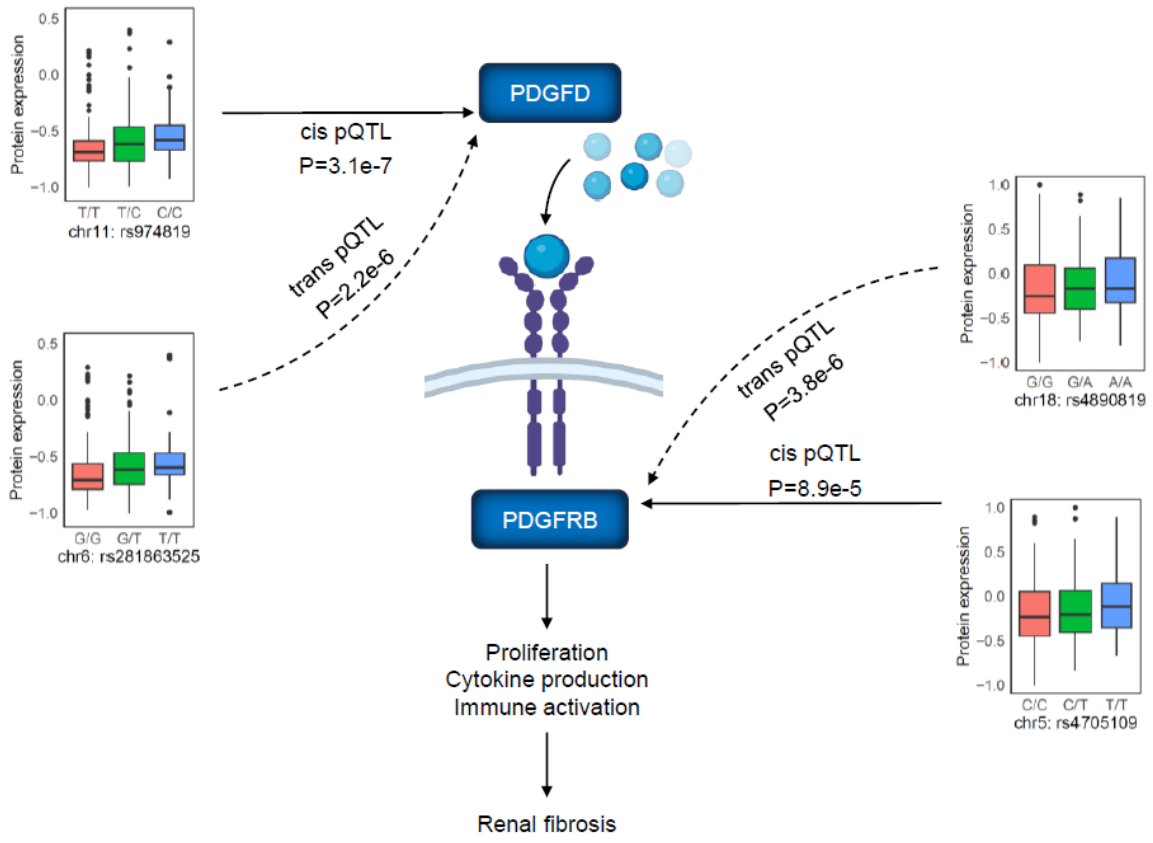
RIN, RNA integrity number.



Supplementary Fig. 2 Comparison of kidney tissue pQTLs with additional covariates.
(a) Venn diagram illustrating the unique and overlapping pQTL proteins identified in the kidney tissue analysis using the main covariates from the primary pQTL analysis, with the addition of DM, hypertension, and sample storage duration as covariates (parentheses indicate unique SOMAmer counts).
(b) Scatter plot showing the correlation of effect sizes between the pQTL analysis with and without the additional covariates. Correlation coefficient was calculated using Spearman's rho (r) statistic and two-sided p value was calculated using asymptotic t approximation.

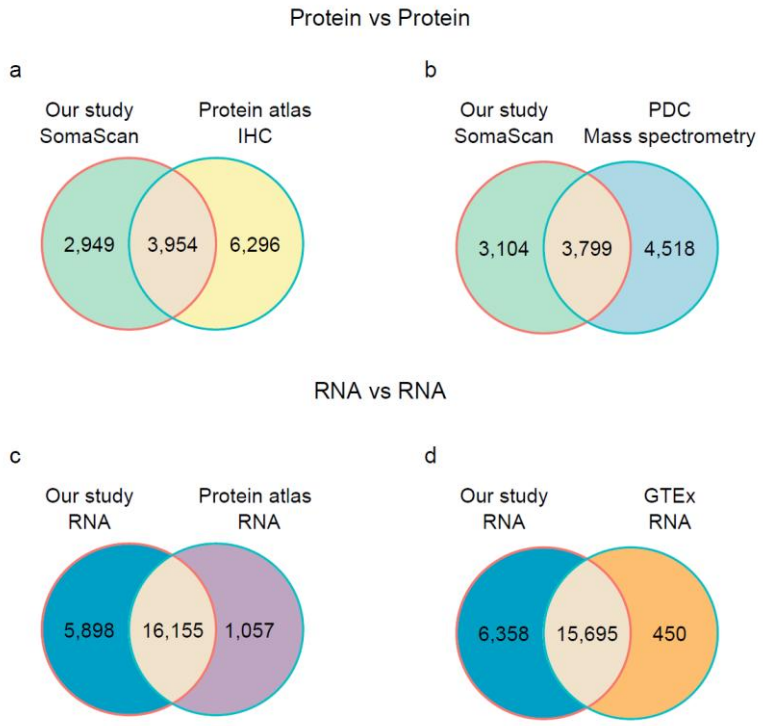


Supplementary Fig. 3 Sex stratified human kidney pQTL analysis.
Venn diagram depicting the overlap of unique SOMAmer counts among male- or female-only kidney pQTLs, and combined (male + female) analyses. The combined analysis represents the primary kidney pQTL results.



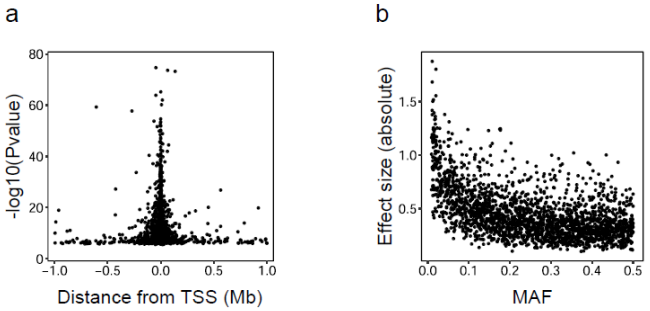
Supplementary Fig. 4 Example of ligand-receptor interaction identified through cis and trans pQTLs in human kidney samples.

This figure illustrates an example of a ligand-receptor interaction between platelet-derived growth factor D (PDGFD) and platelet-derived growth factor receptor beta (PDGFRB), identified through both cis and trans pQTLs. Solid and dashed arrows indicate cis and trans associations, respectively, with corresponding p-values provided. Potential downstream signaling pathways of PDGFRB are also shown. Box plots display protein expression levels for significant variants, showing the median, 25th and 75th percentiles, and values within 1.5× the interquartile range. Each dot represents an individual patient. Group differences were assessed using one-way ANOVA.



Supplementary Fig. 5. Comparison of kidney protein and RNA expression in our dataset with reported expression in existing databases.

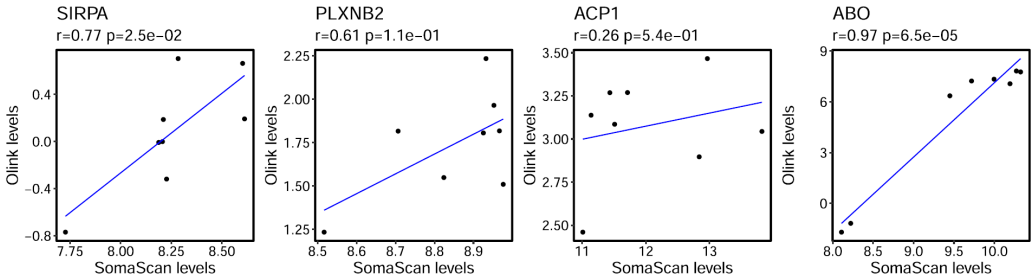
This figure compares the number of proteins and genes detected in the kidney across different datasets, focusing on differences between protein-level and RNA-level detection. Panels **(a)** and **(b)** compare protein-level expression detected by SomaScan with that reported in the Protein Atlas (immunohistochemistry) and the Protein Data Commons (mass spectrometry), respectively, while panels **(c)** and **(d)** compare kidney-expressed genes detected in our RNA-seq data with those reported by the Protein Atlas and GTEx, respectively.



Supplementary Fig. 6 Overview of kidney eQTL analysis.

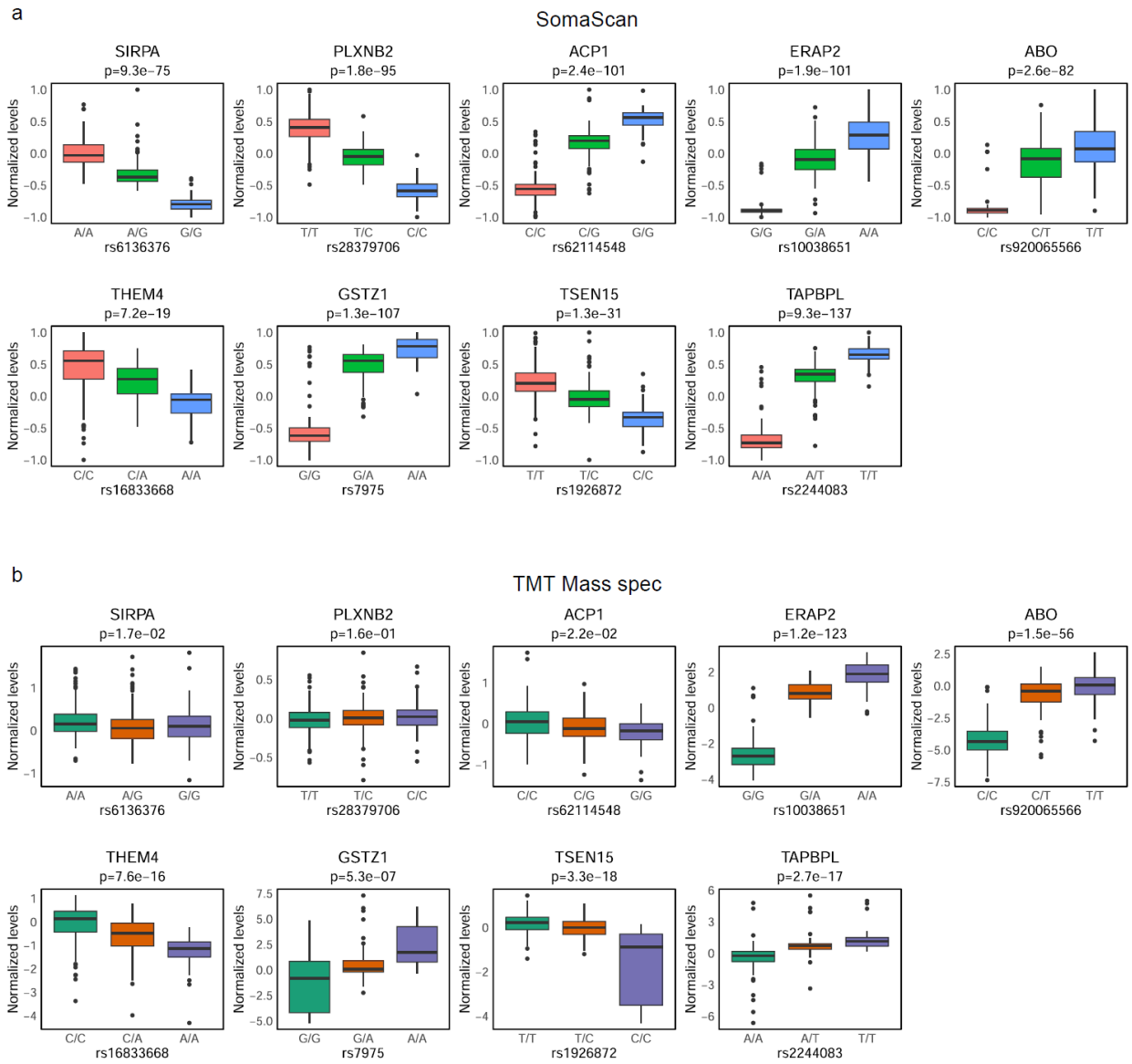
(a) Significance level of cis associations according to distance from the transcription start site (TSS) within a 1 Mb window in kidney eQTL analysis.

(b) Relationship between minor allele frequency (MAF) and absolute effect size for kidney eQTLs, showing the distribution of allele frequencies and their impacts on gene expression.



Supplementary Fig. 7 Correlation analysis of protein levels between SomaScan and Olink assays.

Scatter plots (SomaScan protein x-axis, Olink protein y axis) show the correlation of protein levels for five selected significant pQTL proteins: SIRPA, PLXNB2, ACP1, and ABO, as measured by SomaScan (log2 expression) and Olink (normalized expression in NPX units) platforms, with a sample size of n=8. These proteins, identified among the top 10 significant kidney tissue pQTLs by SomaScan, are also available on the Olink platform. Each point represents paired protein measurements from the same kidney sample using both assays, allowing direct comparison. The correlation of each protein is quantified by Spearman's rho (r) statistic, and the significance is indicated by the p-value.

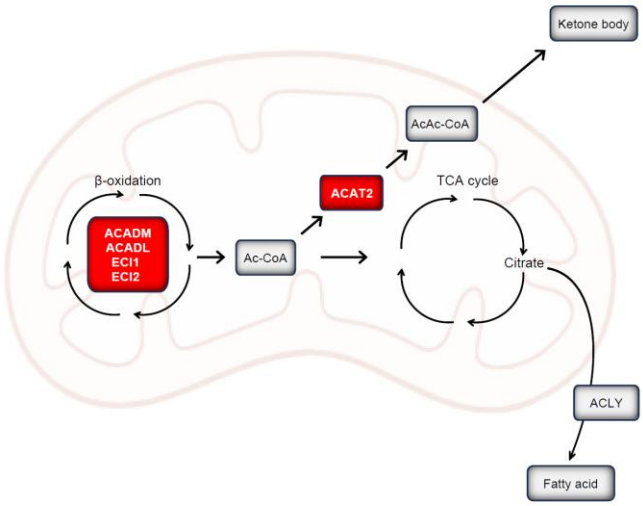


Supplementary Fig. 8 Comparative modality of top pQTL protein expression.

(a) SomaScan analysis: Box plots with normalized protein levels illustrating the allelic expression modality of SIRPA, PLXNB2, ACP1, ERAP2, ABO, THEM4, GSTZ1, TSEN15 and TAPBPL across different genotypes, as measured by the SomaScan platform.

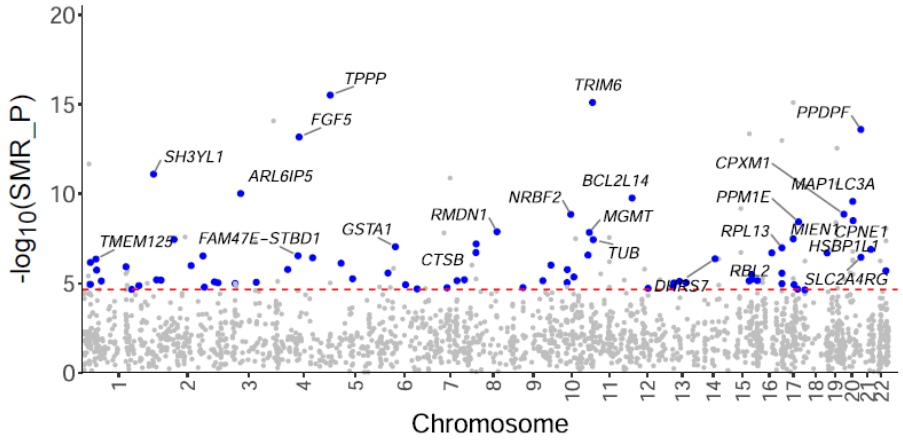
(b) Quantitative multiplexed mass spectrometry analysis: Parallel box plots with normalized protein levels representing the allelic expression modality of the same top pQTL proteins, but as measured by quantitative multiplexed mass spectrometry.

Box plots display protein expression levels for significant variants, showing the median, 25th and 75th percentiles, and values within 1.5× the interquartile range. Each dot represents an individual patient. Group differences were assessed using one-way ANOVA.

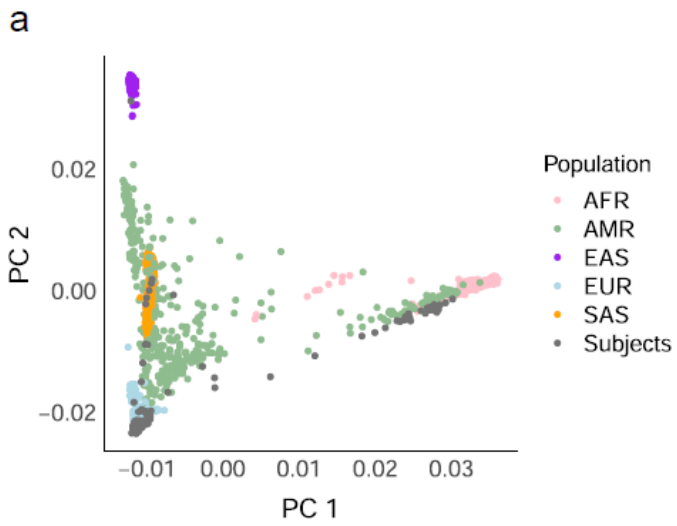


Supplementary Fig. 9 Overview of fatty acid metabolism.

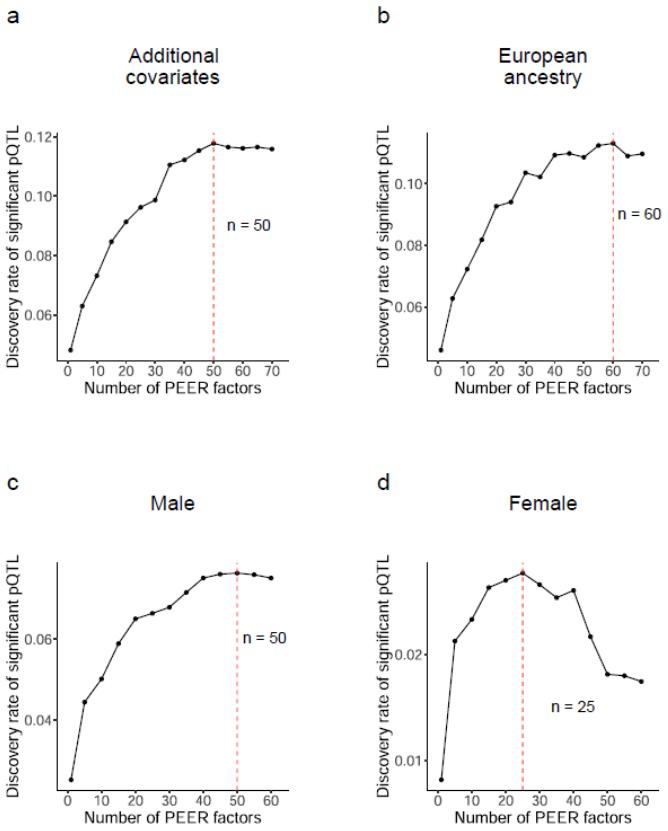
This figure provides a streamlined representation of fatty acid metabolism, including β-oxidation, ketone body synthesis, and fatty acid synthesis. Acetyl-CoA (Ac-CoA) generated through β-oxidation can enter the tricarboxylic acid (TCA) cycle, contribute to ketone body formation via acetoacetyl-CoA (AcAc-CoA) and acetyl-CoA acetyltransferase 2 (ACAT2), or be converted to citrate for transport to the cytoplasm, where ATP citrate lyase (ACLY) regenerates Ac-CoA for fatty acid synthesis. Proteins associated with cis-protein quantitative trait loci are highlighted in red. ACADM, medium-chain acyl-CoA dehydrogenase; ACADL, long-chain acyl-CoA dehydrogenase; ECI1, enoyl-CoA isomerase 1; ECI2, enoyl-CoA isomerase 2.



Supplementary Fig. 10 Manhattan plot of SMR eQTL associations in human kidney tissues. This Manhattan plot illustrates the eQTL landscape in human kidney tissues. The x-axis represents chromosomal locations, while the y-axis indicates the strength of the SMR significance, expressed as $-\log_{10}$(SMR Pvalue). The top 25 most significant gene associations are labeled, displaying only coding genes for eQTLs. The red dashed line represents the significance threshold (P = 2.38e-5, Methods). Blue dots denote significant associations in both SMR and HEIDI analyses, whereas gray dots represent non-significant associations.



Supplementary Fig. 11 Global population structure and genetic diversity revealed by PCA.
PCA scatter plot: The scatter plot displays a PCA of 328 subjects from our dataset alongside 2,504 individuals from the 1,000 Genomes Project Phase1v3, showcasing the genetic diversity and population structure.



Supplementary Fig. 12 Effect of PEER factor count on pQTL discovery rate across different analysis conditions.

This figure demonstrates how varying the number of PEER factors influences the discovery rate of significant pQTLs under different analysis conditions. The y-axis represents the discovery rate of significant pQTLs, while the x-axis shows the number of PEER factors included in the analysis.

(a) pQTL analysis with additional covariates: DM, hypertension, and sample storage duration.

(b) pQTL analysis restricted to individuals of European ancestry.

(c) pQTL analysis conducted in males only.

(d) pQTL analysis conducted in females only.

Dashed red lines indicate the selected PEER factor count (n) that maximized the discovery rate for each condition.