The landscape of allele-specific expression in human kidneys

Ana C. Onuchic-Whitford[1,2,3,4] †, Junmo Sung[1,2,3,4] †, Eric D. Sakkas[2,3,4] †, Michelle T. McNulty[2,3,4], Christopher L. O'Connor[5], Anya Greenberg[2,3,4], Jihoon G. Yoon[6,7], Sowmya Badina[8], NEPTUNE (Nephrotic Syndrome Study Network), Laura H. Mariani[5], Markus Bitzer[5], Matthew G. Sampson[2,3,4]*, Dongwon Lee[2,3,4]*

[1] Division of Nephrology, Brigham and Women's Hospital; Boston, 02115, USA.

[2] Division of Nephrology, Boston Children's Hospital; Boston, 02115, USA.

[3] Harvard Medical School; Boston, 02115, USA.

[4] Broad Institute of MIT and Harvard; Cambridge, 02142, USA.

[5] Department of Medicine, University of Michigan; Ann Arbor, 48109, USA.

[6] Department of Laboratory Medicine, Yonsei University College of Medicine; Seoul, 03722, Republic of Korea.

[7] Department of Laboratory Medicine, Gangnam Severance Hospital; Seoul, 06273, Republic of Korea.

[8] Centre for Digital Health and Precision Medicine, The Apollo University; Chittoor, 517127, India.

*Corresponding authors: dongwon.lee@childrens.harvard.edu, matthew.sampson@childrens.harvard.edu

† These authors contributed equally to this work.

Abstract

Allele-specific expression (ASE), the preferential expression of one gene copy, is a key mechanism of genomic regulation. However, its role in human kidney disease remains poorly understood. In this study, we generated a high-quality, genome-wide ASE map using paired whole-genome sequencing and RNA-seq from microdissected glomerular (GLOM) and tubulointerstitial (TUBE) compartments of patients with proteinuric kidney disease. We showed that the majority of common ASE events were deterministic and sequence-mediated. We also found that diseased kidneys exhibited significantly more ASE in GLOM than TUBE. Unexpectedly, higher ASE in GLOM than TUBE was associated with improved kidney disease outcomes. Differential gene expression analysis suggests this reflects an active, protective transcriptional response, including ribosome and ATP synthesis upregulation, rather than pathogenic dysregulation. Our work reveals glomerular ASE as a marker of adaptive transcriptional activity in proteinuric kidney disease.

Introduction

Humans are diploid organisms and our cells carry two copies of each autosomal gene, which can be differentiated by heterozygous loci (*1*). The contribution of each gene copy, or allele, to a gene's total expression is not always balanced, and predominant expression of one allele over the other constitutes the phenomenon of allele-specific expression (ASE), or allelic imbalance (*2*). ASE results from multiple regulatory mechanisms, with a large heritable component, but is also impacted by non-genetic factors (*3*, *4*). Thus, ASE analysis is a versatile method to assess genomic and epigenetic phenomena, including nonsense-mediated decay, non-coding cis-regulatory variants, genomic imprinting, and alternative splicing (*5*).

As an intra-individual analysis, ASE minimizes batch effects and environmental confounders to uncover true cis-acting regulatory mechanisms and rare variant effects (*6*). However, ASE analysis is subject to technical biases requiring rigorous quality control (QC), particularly regarding haplotype phasing, genotyping, and RNA read mapping (*7*). Errors or biases in any of these steps can result in substantial inaccuracies that are difficult to correct at a cohort level, especially when analyzing rare variants (*1*, *7–10*). Several tools exist for ASE estimation to address some of these issues, but accurately quantifying ASE remains technically challenging (*1*).

Prior studies for ASE analysis have often relied on RNA-seq alone or on single nucleotide polymorphism (SNP) arrays for genotyping (*11*, *12*). Although RNA-seq can be used alone for both genotype calling and expression quantification, extreme ASE can result in heterozygous sites being miscalled as homozygous (*1*). This is a well-described caveat of RNA-only ASE analysis and impedes identification of strong allelic imbalance, including monoallelic expression. SNP arrays, conversely, require imputation that can introduce genotype errors (*7*). Thus, whole-genome sequencing (WGS) is considered the gold-standard genotyping strategy in ASE analysis, as it enables more accurate phasing, higher confidence heterozygous variant calling, minimization of imputation, and assessment of rare variants (*6*).

ASE has been extensively studied in diseases such as cancer, neurodevelopmental disorders and heart disease, with demonstrated instances of causality (*13–15*). However, ASE evaluation in kidney has previously focused on non-diseased tissue either from tumor nephrectomies (*11*, *12*, *16*) or post-mortem samples (*17*). Furthermore, datasets such as GTEx (*18*) do not account for the complexity of kidney structure, which comprises two distinct functional compartments: the glomeruli and tubulointerstitium (*19*). As glomeruli comprise only a tiny fraction (<10%) of the total cell count in the kidney cortex (*20*), the relative abundance of glomerular transcripts in bulk tissue is much lower than that of the tubulointerstitium. This makes it challenging to study ASE of glomerular-specific genes. Such compartmental resolution is particularly important when studying proteinuric diseases, characterized by loss of large amounts of protein in the urine due to defects in glomerular filtration (*21*). These disorders account for up to 90% of end-stage kidney disease (ESKD) worldwide (*22*). Specifically, a common mechanism is injury to specialized glomerular cells, podocytes, and most of the >70 monogenic causes of proteinuria involve genes expressed in this cell type (*23*). However, many patients with glomerular disorders remain undiagnosed, and several known genetic etiologies lack a clear mechanistic pathway to disease (*24*, *25*).

The goal of this study was to discover the genome-wide landscape of ASE in the human kidney and its role in the pathogenesis and progression of proteinuric kidney diseases (**Fig. 1A**). We use a unique dataset of WGS with paired RNA-seq from microdissected compartments of

kidney biopsy from the Nephrotic Syndrome Study Network (NEPTUNE) (*26*). We create a rigorous, generalizable framework for ASE analysis, integrating both available and novel computational tools. Our pipeline includes new steps to address major sources of bias in ASE, including overlapping genes and concealed genotyping errors. We generate a high-confidence ASE dataset and characterize distinct ASE patterns in glomeruli (GLOM) and tubulointerstitium (TUBE) of diseased kidneys, compared to controls. We then perform extensive investigation into the molecular mechanisms, cell-type specificity, clinical correlations, and differential gene expression of ASE in the human kidney.

Results

A new workflow for generation of high-confidence ASE data

We assembled a rigorous pipeline for generation of high-confidence ASE calling to assess allele-specific expression in kidney tissues from the NEPTUNE proteinuric kidney disease cohort (**Methods**). Biopsies from these individuals were microdissected into GLOM and TUBE compartments, followed by bulk RNA sequencing. In total, we initially assessed 240 NEPTUNE individuals with available paired WGS and GLOM RNA-seq, of which 218 also had TUBE RNA-seq. Our ASE analysis workflow is summarized in **Fig. S1** and described in detail in **Methods** and **Supplementary Text**. Here, we highlight three key features introduced in this work: 1) novel QC steps to blacklist variants prone to genotyping errors; 2) bias reduction from misassigned RNA reads in loci containing overlapping genes; and 3) ASE definition using multiple criteria to minimize false positive calls. Briefly, we first performed WGS-based genotyping using the GATK best practices pipeline, followed by genotype phasing using WhatsHap (*27*) and SHAPEIT4 (*28*) with ~500,000 participants of the UK Biobank (*29*) to leverage both read-level and population-level phasing information. As detailed below, our estimated gene-level phasing error (flip) rate is low (~3%), enabling accurate ASE quantification even for genes whose haplotypes are not fully connected by RNA-seq reads.

To ensure high-confidence genotype calling, we also implemented multiple WGS QC measures, excluding variants with inconsistent genotype before and after phasing, biased DNA allelic fraction, or abnormal DNA read depth. These variants showed a significant deviation of DNA (or RNA) allelic fraction from the expected 0.5, suggesting poor genotyping or structural variation (**Fig. S2**). In addition, we blacklisted variants with bi-allelic RNA expression at homozygous WGS genotypes, suggesting systematic poor read alignments in RNA-seq (**Fig. S3**). Overall, we removed an average of 543 variants per sample. These blacklisted variants showed a substantially higher proportion of variant-level ASE than the QC-passed variants, consistent with false-positive ASE from erroneous genotyping (**Fig. S4**). This translates to an average of 302 genes per sample harboring these problematic variants. Without this filter, many would appear to have significant ASE. This QC strategy also identified and excluded 18 problematic samples with excess error rates (**Fig. S2D**). Our findings were further confirmed by verifyBamID analysis (*30*), which flagged 13 of the 18 same samples as having DNA-RNA mismatches. This resulted in a final number of 222 GLOM and 206 paired TUBE samples (**Table S1**).

Next, to correct for allelic mapping bias in ASE analyses, we performed RNA-seq read alignment using STAR with the WASP (Workflow for Allele-Specific Processing) (*9*, *31*, *32*) and confirmed the same benefits in our dataset (**Fig. S5**). These WASP-passed reads, taken only from those overlapping heterozygous loci, represented ~9% of total reads per sample. We then used phASER (*6*) that incorporates read-backed phasing from RNA-seq to optimize phasing of

rare variants. It accurately phases variants over longer distances due to RNA splicing and reports gene-level ASE by counting the number of RNA reads mapping to entire haplotypes. For downstream analyses, we restricted our analysis to genes with sufficient coverage (a total haplotypic count ≥ 20)—designated as "COV" genes, based on prior literature standards (*11*, *33*). Consistently, our data also showed inflated ASE and monoallelic expression (MAE), defined as those with only one haplotype expressed, at lower cutoffs (**Fig. S6**).

A major limitation in standard ASE pipelines is that RNA reads aligned to heterozygous variants are assigned to all genes spanning that genomic position, affecting ~48% of expressed genes in our dataset. To address this, we developed TOGA (Transcript reassignment for Overlapping Genes in ASE), a novel pipeline that resolves these overlapping genes (**Fig. S7; Methods**). Briefly, we first estimated the proportion of allelic counts originating from the target gene relative to overlapping ones; if interference were negligible, the target gene was designated as "resolved." If not, we recalculated haplotypic counts using reads from non-overlapping regions only. If sufficient coverage remained after this filtration, the gene was also resolved; otherwise, it was "excluded" or remained "unresolved" with its original haplotypic count, depending on its coverage (**Fig. S8**). This strategy successfully resolved 75% of expressed genes per individual (with only 4% unresolved, and 21% excluded), substantially reducing false-positive ASE and elucidating the gene responsible for ASE in overlapping regions (**Fig. S9; Table S2**). We note that we kept "unresolved" genes to avoid excluding potential biological signals. However, those genes rarely exhibited ASE due to diluted signals from multiple genes. Going forward, we only considered genes classified as "resolved" and "unresolved" by TOGA.

To quantify the magnitude of ASE, we defined the term "Major Haplotypic Fraction" (MHF), the fraction of reads from the more highly-expressed haplotype (ranging from 0.5 [balanced] to 1.0 [monoallelic]). We defined a gene as having significant ASE as follows: (1) sufficient RNA coverage for high-confidence ASE analysis, i.e., TOGA-resolved COV genes; (2) binomial test with $P_{adj} < 0.05$; and (3) MHF ≥ 0.6 (i.e., one haplotype being expressed ≥50% more than the other). This MHF threshold was implemented to address known overdispersion in ASE data, which is not managed by the binomial test alone, and a similar approach and cutoff were described previously (*3*, *7*). An illustrative ASE analysis plot in one NEPTUNE individual is shown in **Fig. S10.**

The landscape of ASE in human kidneys in proteinuric disease

In our ASE analysis, we achieved extensive coverage of heterozygous variants (**Table S3**), resulting in a median of 5,370 genes in GLOM and 4,989 in TUBE amenable to ASE analysis (COV genes) per individual; cumulatively, we observed 18,304 COV genes, of which ~83% were protein-coding and in ≥1 individual. Thus, most genes in the human genome were covered (**Fig. 1B, Table S4**). While the number of COV genes was similar between compartments, the ASE patterns were distinct. We identified a median of 643 GLOM ASE genes per individual, nearly double the median of 361 TUBE ASE genes (**Table S4**). Moreover, while COV genes were largely shared (mean 68%), only a mean of 12% of ASE genes were shared between compartments in an individual. Thus, ASE was predominantly compartment-specific, with more than double the proportion of GLOM-specific ASE compared to TUBE-specific (59% vs. 29%) (**Fig. 1C**). Monoallelic expression (MAE) genes were rare and similarly biased, with a median of 14 genes in GLOM and 8 in TUBE. Collectively, 3,575 genes in GLOM and 968 in TUBE had

MAE in ≥1 individual, suggesting that MAEs are rare and largely individual-specific (**Table S4**). All ASE genes are reported in **Table S5**.

COV and ASE distribution per gene is shown in **Fig. 1D**. Most COV genes (~66% in GLOM, and ~67% in TUBE) were shared by ≥10% of individuals. However, concordant with prior studies (*34*), we found that ASE was largely private. Shared ASE genes followed an exponentially decreasing distribution (**Fig. 1D**, **Fig. S11**), with the majority having ASE in very few individuals (≤ 8 in GLOM, and ≤3 in TUBE). In fact, only 13.7% (GLOM) and 6.7% (TUBE) of ASE genes had ASE in ≥10% of individuals, and we later find that these genes share certain molecular mechanisms of ASE, such as imprinting and strong eQTL effects. Overall, this indicates kidney ASE patterns are predominantly individual-specific.

ASE events were identified throughout the genome in both GLOM and TUBE without statistical enrichment in any specific genomic regions, except for depletion of the excluded chromosome 6 HLA regions (**Fig. 1E**). The 20 most common ASE genes in GLOM and TUBE were largely distinct, as only 5 were shared (**Fig. 1F**). Known imprinted genes were among the most frequent ASE genes and tended to have very high MHF, as imprinting involves epigenetic silencing of one allele. When ranking ASE genes by the proportion of ASE among COV individuals, even more imprinted genes were found among the top 20, indicating that ASE will consistently be detected for imprinted genes with sufficient coverage (**Fig. S12**).

Identification of imprinted genes in human proteinuric kidneys

Given the enrichment of imprinted genes among our top ASE genes (**Fig. 1E; Fig. S12**) and their known tissue specificity (*33*), we sought to systematically identify imprinted genes within our dataset. To do this, we applied an established pipeline (*33*) that assigns an imprinting likelihood to every ASE gene based on characteristics distinct from other regulatory mechanisms: monoallelic expression in all individuals, but with the expressed haplotype varying randomly across individuals (driven by parent-of-origin effects rather than genotype) **(Fig. S13**). This pipeline initially identified 26 candidate imprinted genes in GLOM (nine not previously identified in any tissue) and 18 candidates in TUBE (five novel) (**Figs. S14-17**).

A potential limitation of this method is the risk of misclassifying genes with very large effect size eQTLs (leading to near mono-allelic expression) as imprinted. This can occur if the causal non-coding regulatory variant is heterozygous in most individuals and in low linkage disequilibrium with the (coding) ASE variants. While the previous work concluded that a gene meeting these conditions are extremely unlikely (*33*), we found that our novel candidates frequently overlap with strong eQTLs. Specifically, 7 out of 9 novel candidates in GLOM had significant eQTLs ($P < 10^{-6}$), compared to only 1 out of 17 of known imprinted genes. Similarly, 4 out of 5 novel candidates in TUBE and 1 out of 13 known imprinted genes had such eQTLs. Therefore, we sought to distinguish true imprinting from strong eQTL effects by implementing an additional algorithm that incorporates genotype data and eQTL direction of effect to test which mechanism better explained the observed monoallelic expression (**Methods**). We found that the vast majority of novel candidates were indeed better explained by strong eQTL effects, driving 8 genes in GLOM (**Fig. S14**) and 4 genes in TUBE (**Fig. S16**). In TUBE, one known imprinted gene, *MEST*, was also better explained by an eQTL (**Fig. S17**).

Making this correction reduced the number of imprinted genes to 18 in GLOM and 13 in TUBE. We identified two novel candidate imprinted genes: *IGLV6-57* (immunoglobulin lambda

variable 6-57) in GLOM (**Fig. S18**) and *GLYATL2* (glycine-N-acyltransferase like 2) in TUBE (**Fig. S19**). We found many known imprinted genes to be uniquely imprinted in each compartment (7 in GLOM and 2 in TUBE), but the absence of imprinting in the other compartment for these genes was largely due to insufficient ASE data (i.e., low expression), rather than true balanced expression.

To test the tissue specificity of these imprinted GLOM and TUBE genes, we compared them to a curated list of 299 known imprinted genes (**Table S6; Fig. S20**) (*33*, *35*). Most of these 299 genes did not have ASE data in GLOM or TUBE due to a lack of heterozygous variants, low expression, and/or removal by TOGA. However, the majority of the known imprinted genes amenable to imprinting analysis still did not show imprinting pattern in kidney (21 out of 38 in GLOM, and 30 out of 43 in TUBE). Instead, these genes showed balanced/biallelic expression (**Fig. S21**), suggesting that these known imprinting genes lost their imprinting status in a tissue-specific manner.

ASE is primarily sequence-mediated

Alongside imprinting, ASE is often driven by other sequence-mediated mechanisms, such as eQTLs and splicing QTLs (sQTLs) (*36*, *37*). Thus, we sought to classify each ASE event in each individual by identifying nearby heterozygous eQTL/sQTLs associated with the ASE gene (**Methods**). We found two major patterns: 1) the mechanisms for rare ASE genes were largely unknown, while most of commonly shared ASE genes (i.e., genes with ASE in >30% of individuals) were sequence-mediated or imprinted; and 2) individuals showed remarkably consistent proportions of their ASE explained by each mechanism (**Fig. 2A,B**). Specifically, the vast majority of commonly shared ASE genes (mean of 83.9%)—which have sufficient power for QTL detection—could be explained by one of three molecular mechanisms: 4.7% by sQTLs, 58.4% by eQTLs, 14.5% by both sQTLs and eQTLs, and 8.6% by imprinting (**Fig. 2B**). These results were largely replicated in TUBE (**Fig. S22A,B**). As a representative example, *PLCG2*—the most common ASE gene after excluding imprinted genes (**Fig. 1F**)—demonstrated GLOM-specific, sequence-mediated regulation. In GLOM, *PLCG2* exhibited near mono-allelic expression driven by a strong eQTL (**Fig. S23**). Conversely, in TUBE, the gene showed balanced expression with no eQTL, underscoring the compartment specificity of these regulatory mechanisms.

To further validate these findings, we evaluated our ASE data at the single variant level to compare the directional consistency between ASE and eQTL/sQTL effects. We hypothesized that, if truly sequence-mediated, most ASE variants overlapping with eQTL/sQTL should show strong concordance in allelic preference. Similar to an ASE gene, an ASE variant was defined as having a total RNA coverage of ≥ 8, ≥50% allelic imbalance, and $P_{adj} < 0.05$ from a binomial test. We further restricted our analysis to common ASE variants (≥10% of heterozygous samples, ≥50% showing ASE) to ensure sufficient statistical power for comparison with QTL data. Indeed, we found that most ASE variants showed extreme allelic bias, as the majority (65.3%) demonstrated "perfect allelic preference" where the same specific allele was consistently preferred in every individual carrying the ASE variant (**Fig. 2C**). As expected, the majority (61.6%) of variants with perfect allelic preference were primarily driven by sequence-mediated mechanisms. Conversely, among all sequence-mediated variants overlapping ASE, the vast majority showed perfect allelic preference in ASE: 92.7% of sQTLs, 72.1% of eQTLs, and 84.8% of both. We noted that a non-trivial proportion of ASE variants with strong allelic

preference had an unknown mechanism, hinting that limited power in QTL analysis might have missed a substantial proportion of them. We found similar results in TUBE (**Fig. S22C**). Overall, these results strongly suggest that ASE is primarily sequence-mediated.

We next hypothesized that some genes with an unknown ASE mechanism are driven by rare regulatory variants in their promoters that eQTL analysis is underpowered to identify. We used the ChromKid model (*38*), which predicts the effect of variants on chromatin accessibility to score heterozygous promoter variants in all COV genes. We compared promoter variants in ASE genes of known and unknown mechanism to those of non-ASE genes, which served as our baseline control. This gave us the likelihood to find predicted regulatory variants in the promoter of typical genes (**Fig. 2D**). As expected, the promoters of genes with an unknown ASE mechanism were significantly enriched for rare predicted regulatory variants—those with minor allele count (MAC) ≤ 5—compared to ASE genes of known mechanism (**Fig. 2E**). This difference between known and unknown ASE mechanism disappeared when including more common predicted regulatory variants (MAC ≤ 50), showing that rare variants specifically drive unknown ASE. In both mechanistically known and unknown ASE, rare ASE genes (i.e., observed in few individuals) were particularly enriched for predicted regulatory variants, with decreasing enrichment as we included genes with more common ASE (**Fig. 2E**). This suggests that while ASE is often driven by rare promoter variants, common ASE genes are driven by a larger variety of mechanisms beyond promoter regulatory variants.

Splicing events and phasing errors affect gene-level ASE consistency

The above analysis confirmed that sQTLs are one of the major mechanisms driving ASE events. However, phASER's gene-level aggregation has a potential limitation in identifying sQTL-driven ASE as it sums all reads originating from the same haplotype. While this approach works well for eQTLs or imprinting (where the whole transcript is affected uniformly), it masks the complexity of splicing events where allelic bias may vary across variants within the same haplotype. This is best illustrated by the example in **Fig. 3A**. The alternative allele "A" of rs10774671 disrupts a canonical splicing acceptor site, producing an isoform (*OAS1.2*) which retains the full intron. Consequently, this haplotype accumulates reads at intronic variants, such as rs60623134. Conversely, alleles of exonic variants phased with the reference allele of rs10774671, such as rs2660, have more reads as the canonical isoform (*OAS1.1*) is more stable than *OAS1.2*. Therefore, a unique characteristic of sQTL-driven ASE is intra-genic allelic discordance, where the gene harbors variants with allelic imbalance towards haplotype A, others towards haplotype B, and others with balanced expression.

To confirm this, we quantified the deviation of variant-level haplotype fractions from the corresponding gene-level haplotype fractions using root mean square error (RMSE) (**Methods**). We hypothesized that sGenes (genes with alternative splicing driven by variants) would have higher RMSE values due to the allelic discordance within the genes. As expected, sGenes had significantly higher RMSE values than genes driven by other ASE mechanisms (all comparisons Wilcoxon $P < 0.001$; median RMSE = 0.085) while imprinted genes had the lowest (median RMSE = 0.0019) (**Fig. 3B**). ASE genes of unknown mechanism had a median RMSE of 0.048, falling between sGenes and eGenes (median RMSE = 0.044), suggesting that some ASE genes with unknown mechanism are driven by sQTLs not found in our dataset. This was largely true of TUBE samples as well, although Unknown and eGenes were not significantly different (Wilcoxon $P = 0.34$) (**Fig. S22D**).

Of note, we observed systematically higher RMSE values in genes that were not fully connected by phASER's RNA-backed phasing (**Fig. 3B**, **Fig. S24)**. This is indicative of "haplotypic flips" arising from errors in population-based DNA phasing. Imprinted genes provide a ground truth to estimate this error rate: as they are strictly expected to have monoallelic expression, the observation of monoallelic expression mapping to opposing haplotypes within the same gene confirms a phasing error. We used this principle to estimate the prevalence of haplotypic flips in our dataset. We calculated the proportion of imprinted gene-individuals where at least two variants with read depth ≥ 8 demonstrated conflicting monoallelic expression (≥90% allelic bias). In GLOM, we tested 666 imprinted gene-individuals (mean 3.16 variants per gene) and identified a haplotypic flip rate of only 3%. Similarly, we tested 400 imprinted gene-individuals in TUBE (mean 3.75 variants) and identified a 2.5% rate. This is significantly lower than the previously reported gene-level phasing error rates (*8*), probably due to our enhanced pipeline, further confirming that our gene-level ASE is of high-quality.

Common ASE genes exhibit cell-type specific expression

We next hypothesized that the ASE genes we discovered are transcriptionally coordinated within specific cell types. Using the Kidney Precision Medicine Project (KPMP) single cell RNA-seq dataset (*20*), we classified 54 cell types by compartment and regrouped them into 12 major categories: two unique to GLOM, six unique to TUBE, and four shared between compartments (**Table S7**). We then performed pseudo-bulk aggregation of single-cell RNA-seq data, followed by normalization, to calculate the proportion of each ASE gene's RNA expression attributable to each cell type within the compartment, termed the "cell-type proportion". Genes with a cell-type proportion exceeding a defined threshold (i.e., 0.33 for GLOM and 0.2 for TUBE) in one cell type relative to others were classified as having "cell-type predominance" Genes whose expression was broadly distributed across all cell types, without predominance in any single type, were classified as "ubiquitous" (**Methods**).

We initially analyzed all ASE genes observed in at least one individual (**Fig. 4A, Fig. S25**). In GLOM, 60.1% of ASE genes showed ubiquitous expression; in TUBE, this proportion was 57.4%. However, when we restricted our analysis to more common ASE genes, the proportion of ubiquitous ASE genes decreased while cell-type predominant ASE genes increased. Notably, this increase in cell-type predominance was most pronounced for cell types unique to each compartment. For example, among ASE genes observed in at least 30% of individuals, we found increased cell-type predominance for podocytes (POD) and parietal epithelial cells (PEC) in GLOM and proximal tubule cells (PT) in TUBE (**Fig. 4B**). Compared to all ASE genes, the percentage of POD-predominant ASE genes increased from 8.4% to 17.2%, and PEC-predominant genes increased from 13.2% to 21.4%, while ASE genes with ubiquitous expression decreased from 60.1% to 35.4%. In TUBE, PT-predominant ASE genes increased from 10.0% to 35.3%, and ubiquitous genes decreased from 57.4% to 33.6%. These findings demonstrate that while rare ASE genes are expressed across all cell types, common ASE genes tend to exhibit more compartment- and cell-type-specific expression patterns within GLOM and TUBE.

Proteinuric kidneys have more genome-wide ASE in GLOM than TUBE

A notable observation in NEPTUNE was that GLOM samples had more ASE genes than TUBE (**Fig. 2**). We therefore calculated the ASE proportion per sample (i.e., a sample's number of ASE genes over their number of COV genes). The median ASE proportion in GLOM was indeed significantly higher than TUBE (11.3% vs 7.0%, $P<2.2 \times 10^{-16}$), with 177 out of 206 (85.9%) individuals showing a higher ASE proportion in GLOM (**Fig. 5A**).

To determine whether this discrepancy was specific to proteinuric disease or a normal phenomenon of kidney subcompartments, we analyzed 31 microdissected "control" kidneys without proteinuric disease (**Methods**). As these control samples had only bulk RNA-seq (without paired WGS), we developed an RNA-only ASE pipeline for direct comparison (**Methods; Fig. S26A**). We note that, in RNA-only ASE analysis, true (near)-monoallelic expression is missed as these sites are assumed to be homozygous, resulting in fewer identified ASE events and no monoallelic expression compared to WGS+RNA pipeline (**Table S4; Fig. S26B**). Nonetheless, the RNA-only analysis of NEPTUNE is highly consistent with results from WGS+RNA pipeline (**Fig. S26C**) and consistently demonstrated a significantly higher ASE proportion in GLOM (median 7.67% vs. 5.27%, $P<2.2 \times 10^{-16}$) **(Fig. S27A)**, with a similar number of individuals (173, 84.0%) with higher ASE proportion. In contrast, the control cohort showed no significant difference (median 6.48% vs. 6.96%, P=0.68) (**Fig. S27B, C**).

Due to baseline differences in the NEPTUNE and control cohorts' biology and data (i.e., read depth, ancestry, age, and number of COV genes), we normalized the ASE proportions by calculating a GLOM-to-TUBE ASE ratio (GTAR). In the RNA-only analysis, the median GTAR in NEPTUNE was significantly higher than the controls (1.40 vs 0.97; $P=2.2 \times 10^{-6}$) (**Fig. 5B**). We hypothesized that clinical outcomes in proteinuric kidney disease vary as a function of GTAR. To test this, we dichotomized 194 NEPTUNE participants into a low-GTAR group (GTAR 0.44 – 1.578, similar to most controls) and a high-GTAR group (1.579 – 6.16), based on WGS+RNA analysis, after excluding twelve participants with no clinical data (**Fig. 5A**). Kaplan-Meier survival analysis revealed that the high-GTAR group had slower kidney disease progression (composite of ESKD and/or ≥40% decline in estimated glomerular filtration rate) compared to the low-GTAR group ($P=1.57\times10^{-3}$) (**Fig. 5C**). Cox proportional hazards models also confirmed this: high-GTAR patients were 65% less likely to reach the composite endpoint after adjusting for all covariates (HR=0.35, $P=9.63\times10^{-4}$) (**Fig. 5D**). Of note, we found no difference in survival when stratifying solely based on either GLOM or TUBE ASE proportion. (**Fig. S28**).

We next investigated additional clinical phenotypes (**Methods**). The high-GTAR group also achieved complete remission of proteinuria sooner (adjusted HR=1.58, P=0.023) (**Fig. S29A**) and exhibited a longer time to relapse, although the difference was not statistically significant, likely due to limited sample size (**Fig. S29B**). In contrast, the low-GTAR group showed non-significantly higher global sclerosis (P=0.083) and interstitial fibrosis (P=0.051) (**Table S8**). Moreover, six of the seven participants who developed kidney failure were in the low-GTAR group. We found no significant associations with other demographic, clinical, or histologic parameters. Altogether, a high GTAR is associated with better kidney outcomes.

Differential gene expression analysis reveals downregulation of fibrosis genes and upregulation of ribosome and ATP pathways in the high-GTAR group

We next investigated whether differences in expression patterns in each group could be driving variation in both ASE proportion and disease progression. To this end, we performed differential

gene expression (DGE) analysis between high- and low-GTAR groups, using GLOM or TUBE RNA-seq (separately) and DESeq2 (*39*) (**Methods**). GLOM DGE analysis identified 181 significantly differentially expressed genes (DEGs), with $P_{adj} < 0.05$ and absolute $\log_2$-fold change ($\log_2$FC) > 0.5. Of these, 115 genes were upregulated and 66 downregulated in the high-GTAR group (**Fig. 6A, Table S9**). In contrast, TUBE analysis demonstrated no significant DEGs. We thus focused on the GLOM DGE results.

Characterization of these DEGs revealed striking biological patterns. First, a high proportion of DEGs (62/181; 34.3%) were long non-coding RNAs (lncRNAs). The vast majority (58/62; 93.5%) were upregulated, including 20 antisense to a protein-coding gene, some of which promote kidney injury (e.g., *AQP5*, *LYZ*, *P2RX7*, and *CERS6*) (*40–51*). Second, gene regulation appeared "concordant" with the improved outcome observed in the high-GTAR group: upregulation of kidney-protective genes and downregulation of genes promoting kidney injury. Specifically, we found an upregulation of protein-coding genes with well-described protective roles in the kidney, *FCGR2B* (Fc gamma receptor IIb), an immune inhibitory receptor which suppresses autoimmunity and autoantibody-mediated glomerulonephritis (*52–55*) and *NGF* (nerve growth factor), secreted by podocytes and a protective factor against glomerular kidney injury (*56–58*). Conversely, the 66 downregulated DEGs were strongly enriched for genes that promote kidney injury and fibrosis: 10 (15.2%) of the downregulated genes that promote fibrosis (e.g., *FN1*, *CCL19*, *CCL21*, *ADAMTS4, ADAMTS16, CEMIP*) (*59–74*), a finding consistent with the upregulation of lncRNAs antisense to other fibrosis-inducing genes. Several other downregulated genes were associated with kidney dysfunction (e.g., *MRC2*, *PIGR*, *ACER2, IL1RAP, NEFL*) (*75–80*). Moreover, many DEGs were implicated in kidney disease or function by GWAS studies (e.g., *AKR1B1*, *PIGR, IL1RAP*, *SCNN1G, GRM7, STC*) (*81–87*), or were implicated in monogenic kidney diseases (e.g., *SCNN1B*, *SCNN1G, FN1*). Finally, 8/115 (7%) of the upregulated DEGs were ribosome-related, with none among the downregulated DEGs—a trend strongly supported by subsequent pathway analyses. Detailed characterization of all DEGs in the context of kidney disease biology is summarized in **Table S10**.

We next systematically assessed these 181 DEGs with over-representation analysis (ORA). For upregulated genes, "ribosome biogenesis" (KEGG) was the single significant concept ($P_{adj} < 0.05$). For downregulated genes, we identified 57 significant terms (many redundant), including "antigen processing", "response to corticosterone", "cell adhesion", "sodium channel activity", "collagen binding" and "immunoglobulin complex" (**Fig. 6B, Table S9**).

To comprehensively identify molecular pathways, we also performed gene set enrichment analysis (GSEA) on the full GLOM transcriptome using clusterProfiler, ranking all genes by signed($\log_2$FC)×-$\log_{10}(P)$ (**Table S11**) (*88*). Consistent with the ORA results, the top 2 upregulated terms (disregarding highly redundant terms) related to ribosome and oxidative phosphorylation in all 5 functional databases (**Fig. 6C,D**). The most strongly downregulated pathways involved cell adhesion and extra-cellular matrix interaction, interferon signaling, vesicle transport and endocytosis, and aminoacyl-tRNA biosynthesis. In the TUBE DGE analysis, GSEA did not identify any significantly enriched GO terms.

Discussion

In this study, we present a comprehensive, genome-wide characterization of allele-specific expression in the human kidney, leveraging a unique dataset of paired WGS and microdissected

bulk RNA-seq from the NEPTUNE cohort (*26*). By developing and applying a rigorous computational pipeline, we addressed major technical hurdles in ASE analysis, such as overlapping genes and concealed genotyping errors. This allowed us to uncover an accurate landscape of ASE in proteinuric kidney disease. Our primary finding is that, unlike in control tissue, diseased kidneys exhibit a significantly higher proportion of genome-wide ASE in GLOM compared to TUBE. We quantified this discrepancy using a new metric, the GLOM-to-TUBE ASE ratio (GTAR), and demonstrated that a high GTAR was strongly associated with improved kidney survival and faster disease remission.

A critical implication of our work is the biological meaning behind the elevated GTAR. The association of high GTAR with better clinical outcomes suggests this imbalance is not a marker of dysregulation or injury. Instead, our differential gene expression analysis between high- and low-GTAR groups points to a more active, protective transcriptional state. We found that the high-GTAR (better outcome) group showed significant upregulation of pathways related to ribosome biogenesis and oxidative phosphorylation (ATP synthesis) in GLOM. This suggests that the observed increase in ASE is a byproduct of elevated transcriptional and metabolic activity, which may confer glomerular resilience to proteinuric injury. This is further supported by the specific DEGs identified; the high-GTAR group exhibited upregulation of known kidney-protective genes and concordant downregulation of a suite of pro-fibrotic and injury-related genes, suggesting an active, adaptive glomerular response.

Our study also provides insights into the molecular mechanisms driving ASE. We demonstrated that most ASE events are individual-specific. However, ASE genes shared across many individuals were predominantly driven by known molecular mechanisms (i.e., eQTLs, sQTLs, and genomic imprinting). More importantly, we provide an explanation for individual-specific ASE, previously speculated as stochastic or random effects. By integrating deep learning models (*38*), we showed that genes with an "unknown" mechanism of ASE were significantly enriched for rare, heterozygous predicted regulatory variants in their promoter regions. This finding strongly suggests that a substantial portion of rare, individual-specific ASE may not be stochastic, but rather a deterministic outcome of an individual's unique gene regulatory architecture, which eQTL/sQTL studies are underpowered to detect.

The intersection of elevated, disease-specific glomerular ASE and the deterministic nature of its regulation raise a new set of interesting questions. For example, is it possible that the genetic architecture of certain individuals, perhaps through rare regulatory variants, prevents key genes from making an effective transcriptional response, predisposing them to the low-GTAR (poor outcome) state? This also raises the broader question of whether genetic effects on glomerular disease risk are context-specific, only becoming apparent or exerting their pathogenic effect under the physiological stress of glomerular injury. Consequently, it leads to the following important question: to what extent do studies relying on healthy reference datasets miss these critical, disease-specific regulatory mechanisms?

These biological insights were enabled by the development of a stringent and rigorous bioinformatic pipeline designed to generate high-confidence ASE data. We implemented multiple novel QC steps, including detection of discordant genotypes between pre- and post-phasing and systematic genotyping errors by evaluating homozygous samples. Furthermore, we developed TOGA, a new algorithm to resolve read misassignments from overlapping genes, a problem affecting nearly half of the analyzable genes. This rigorous framework, including the use of WASP (*9*) to correct mapping bias and phASER (*6*) for haplotype-level quantification, was critical for accurately quantifying ASE.

This study has several limitations. First, the use of short-read RNA-seq can fail to resolve haplotypes across long distances, particularly for genes with few heterozygous variants. However, we mitigated this by integrating population-level phasing from SHAPEIT4 (*28*) and RNA read-backed phasing from phASER, which leverages spliced reads to connect distant variants. The high quality of our phasing is supported by our low estimation of gene-level haplotypic flips (2.5-3%) in known imprinted genes. Second, our primary analysis was performed on bulk RNA-seq from microdissected glomerular and tubulointerstitial compartments. While this provides crucial compartmental resolution, it lacks the granularity of single-cell analysis. We began to address this by integrating KPMP single-cell data (*20*), which revealed that frequently shared ASE genes tend to be cell-type specific. Nonetheless, performing ASE analysis at a true single-cell resolution is needed to fully delineate the cell-specific drivers of ASE in kidney disease. Third, our comparison to the control cohort required using an RNA-only pipeline due to data availability, although the stark difference in GTAR was robust. Lastly, our identification of rare variants driving "unknown" ASE relies on computational predictions, and functional validation of these specific promoter variants is necessary.

In conclusion, we provide the first comprehensive landscape of allele-specific expression in kidney tissue in proteinuric disease, revealing widespread transcriptional changes in the glomerular compartments. We establish a new metric, GTAR, which links a higher proportion of glomerular ASE to a protective transcriptional state and improved kidney survival. This study demonstrates that rigorous ASE analysis is a powerful tool for uncovering disease biology and identifying clinically relevant biomarkers in human kidney disease.

References and Notes

1. S. Cleary, C. Seoighe, Perspectives on Allele-Specific Expression. *Annu. Rev. Biomed. Data Sci.* **4**, 101–122 (2021).

2. T. Pastinen, Genome-wide allele-specific analysis: insights into regulatory variation. *Nat Rev Genet* **11**, 533–538 (2010).

3. P. Mohammadi, S. E. Castel, B. B. Cummings, J. Einson, C. Sousa, P. Hoffman, S. Donkervoort, Z. Jiang, P. Mohassel, A. R. Foley, H. E. Wheeler, H. K. Im, C. G. Bonnemann, D. G. MacArthur, T. Lappalainen, Genetic regulatory variation in populations informs transcriptome analysis in rare disease. *Science* **366**, 351–356 (2019).

4. D. A. Knowles, J. R. Davis, H. Edgington, A. Raj, M.-J. Favé, X. Zhu, J. B. Potash, M. M. Weissman, J. Shi, D. F. Levinson, P. Awadalla, S. Mostafavi, S. B. Montgomery, A. Battle, Allele-specific expression reveals interactions between genetic variation and environment. *Nat Methods* **14**, 699–702 (2017).

5. C. D. Robles-Espinoza, P. Mohammadi, X. Bonilla, M. Gutierrez-Arcelus, Allele-specific expression: applications in cancer and technical considerations. *Curr Opin Genet Dev* **66**, 10–19 (2021).

6. S. E. Castel, P. Mohammadi, W. K. Chung, Y. Shen, T. Lappalainen, Rare variant phasing and haplotypic expression from RNA sequencing with phASER. *Nat Commun* **7**, 12817 (2016).

7. S. E. Castel, A. Levy-Moonshine, P. Mohammadi, E. Banks, T. Lappalainen, Tools and best practices for data processing in allelic expression analysis. *Genome Biol* **16**, 195 (2015).

8. Y. Choi, A. P. Chan, E. Kirkness, A. Telenti, N. J. Schork, Comparison of phasing strategies for whole human genomes. *PLoS Genet* **14**, e1007308 (2018).

9. B. van de Geijn, G. McVicker, Y. Gilad, J. K. Pritchard, WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods* **12**, 1061–1063 (2015).

10. N. I. Panousis, M. Gutierrez-Arcelus, E. T. Dermitzakis, T. Lappalainen, Allelic mapping bias in RNA-sequencing is not a major confounder in eQTL studies. *Genome Biol* **15**, 467 (2014).

11. J. Fan, J. Hu, C. Xue, H. Zhang, K. Susztak, M. P. Reilly, R. Xiao, M. Li, ASEP: Gene-based detection of allele-specific expression across individuals in a population by RNA sequencing. *PLoS Genet* **16**, e1008786 (2020).

12. Y. Muto, P. C. Wilson, N. Ledru, H. Wu, H. Dimke, S. S. Waikar, B. D. Humphreys, Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. *Nat Commun* **12**, 2190 (2021).

13. C. M. Bielski, M. T. A. Donoghue, M. Gadiya, A. J. Hanrahan, H. H. Won, M. T. Chang, P. Jonsson, A. V. Penson, A. Gorelick, C. Harris, A. M. Schram, A. Syed, A. Zehir, P. B. Chapman, D. M. Hyman, D. B. Solit, K. Shannon, S. Chandarlapaty, M. F. Berger, B. S.

Taylor, Widespread Selection for Oncogenic Mutant Allele Imbalance in Cancer. *Cancer Cell* **34**, 852-862.e4 (2018).

14. C. Lee, E. Y. Kang, M. J. Gandal, E. Eskin, D. H. Geschwind, Profiling allele-specific gene expression in brains from individuals with autism spectrum disorder reveals preferential minor allele usage. *Nat Neurosci* **22**, 1521–1532 (2019).

15. D. M. McKean, J. Homsy, H. Wakimoto, N. Patel, J. Gorham, S. R. DePalma, J. S. Ware, S. Zaidi, W. Ma, N. Patel, R. P. Lifton, W. K. Chung, R. Kim, Y. Shen, M. Brueckner, E. Goldmuntz, A. J. Sharp, C. E. Seidman, B. D. Gelb, J. G. Seidman, Loss of RNA expression and allele-specific expression associated with congenital heart disease. *Nat Commun* **7**, 12824 (2016).

16. H. Liu, A. Abedini, E. Ha, Z. Ma, X. Sheng, B. Dumoulin, C. Qiu, T. Aranyi, S. Li, N. Dittrich, H.-C. Chen, R. Tao, D.-C. Tarng, F.-J. Hsieh, S.-A. Chen, S.-F. Yang, M.-Y. Lee, P.-Y. Kwok, J.-Y. Wu, C.-H. Chen, A. Khan, N. A. Limdi, W.-Q. Wei, T. L. Walunas, E. W. Karlson, E. E. Kenny, Y. Luo, L. Kottyan, J. J. Connolly, G. P. Jarvik, C. Weng, N. Shang, J. B. Cole, J. M. Mercader, R. Mandla, T. D. Majarian, J. C. Florez, M. E. Haas, L. A. Lotta, Regeneron Genetics Center‡, GHS-RGC DiscovEHR Collaboration§, T. G. Drivas, Penn Medicine BioBank¶, H. M. T. Vy, G. N. Nadkarni, L. K. Wiley, M. P. Wilson, C. R. Gignoux, H. Rasheed, L. F. Thomas, B. O. Åsvold, B. M. Brumpton, S. I. Hallan, K. Hveem, J. Zheng, J. N. Hellwege, M. Zawistowski, S. Zöllner, N. Franceschini, H. Hu, J. Zhou, K. Kiryluk, M. D. Ritchie, M. Palmer, T. L. Edwards, B. F. Voight, A. M. Hung, K. Susztak, Regeneron Genetics Center, GHS-RGC DiscovEHR Collaboration, Penn Medicine BioBank, Kidney multiome-based genetic scorecard reveals convergent coding and regulatory variants. *Science* **387**, eadp4753 (2025).

17. S. E. Castel, F. Aguet, P. Mohammadi, GTEx Consortium, F. Aguet, S. Anand, K. G. Ardlie, S. Gabriel, G. A. Getz, A. Graubert, K. Hadley, R. E. Handsaker, K. H. Huang, S. Kashin, X. Li, D. G. MacArthur, S. R. Meier, J. L. Nedzel, D. T. Nguyen, A. V. Segrè, E. Todres, F. Aguet, S. Anand, K. G. Ardlie, B. Balliu, A. N. Barbeira, A. Battle, R. Bonazzola, A. Brown, C. D. Brown, S. E. Castel, D. F. Conrad, D. J. Cotter, N. Cox, S. Das, O. M. de Goede, E. T. Dermitzakis, J. Einson, B. E. Engelhardt, E. Eskin, T. Y. Eulalio, N. M. Ferraro, E. D. Flynn, L. Fresard, E. R. Gamazon, D. Garrido-Martín, N. R. Gay, G. A. Getz, M. J. Gloudemans, A. Graubert, R. Guigó, K. Hadley, A. R. Hame, R. E. Handsaker, Y. He, P. J. Hoffman, F. Hormozdiari, L. Hou, K. H. Huang, H. K. Im, B. Jo, S. Kasela, S. Kashin, M. Kellis, S. Kim-Hellmuth, A. Kwong, T. Lappalainen, X. Li, X. Li, Y. Liang, D. G. MacArthur, S. Mangul, S. R. Meier, P. Mohammadi, S. B. Montgomery, M. Muñoz-Aguirre, D. C. Nachun, J. L. Nedzel, D. T. Nguyen, A. B. Nobel, M. Oliva, Y. S. Park, Y. Park, P. Parsana, A. S. Rao, F. Reverter, J. M. Rouhana, C. Sabatti, A. Saha, A. V. Segrè, A. D. Skol, M. Stephens, B. E. Stranger, B. J. Strober, N. A. Teran, E. Todres, A. Viñuela, G. Wang, X. Wen, F. Wright, V. Wucher, Y. Zou, P. G. Ferreira, G. Li, M. Melé, E. Yeger-Lotem, M. E. Barcus, D. Bradbury, T. Krubit, J. A. McLean, L. Qi, K. Robinson, N. V. Roche, A. M. Smith, L. Sobin, D. E. Tabor, A. Undale, J. Bridge, L. E. Brigham, B. A. Foster, B. M. Gillard, R. Hasz, M. Hunter, C. Johns, M. Johnson, E. Karasik, G. Kopen, W. F. Leinweber, A. McDonald, M. T. Moser, K. Myer, K. D. Ramsey, B. Roe, S. Shad, J. A. Thomas, G. Walters, M. Washington, J. Wheeler, S. D. Jewell, D. C. Rohrer, D. R. Valley, D. A. Davis, D. C. Mash, M. E. Barcus, P. A. Branton, L. Sobin, L. K. Barker, H. M. Gardiner, M. Mosavel, L. A. Siminoff, P.

Flicek, M. Haeussler, T. Juettemann, W. J. Kent, C. M. Lee, C. C. Powell, K. R. Rosenbloom, M. Ruffier, D. Sheppard, K. Taylor, S. J. Trevanion, D. R. Zerbino, N. S. Abell, J. Akey, L. Chen, K. Demanelis, J. A. Doherty, A. P. Feinberg, K. D. Hansen, P. F. Hickey, L. Hou, F. Jasmine, L. Jiang, R. Kaul, M. Kellis, M. G. Kibriya, J. B. Li, Q. Li, S. Lin, S. E. Linder, S. B. Montgomery, M. Oliva, Y. Park, B. L. Pierce, L. F. Rizzardi, A. D. Skol, K. S. Smith, M. Snyder, J. Stamatoyannopoulos, B. E. Stranger, H. Tang, M. Wang, P. A. Branton, L. J. Carithers, P. Guan, S. E. Koester, A. R. Little, H. M. Moore, C. R. Nierras, A. K. Rao, J. B. Vaught, S. Volpi, K. G. Ardlie, T. Lappalainen, A vast resource of allelic expression data spanning human tissues. *Genome Biol* **21**, 234 (2020).

18. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

19. F. A. Garza, S. W. Leslie, "Anatomy, Abdomen and Pelvis: Kidneys" in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2025; http://www.ncbi.nlm.nih.gov/books/NBK482385/).

20. B. B. Lake, R. Menon, S. Winfree, Q. Hu, R. Melo Ferreira, K. Kalhor, D. Barwinska, E. A. Otto, M. Ferkowicz, D. Diep, N. Plongthongkum, A. Knoten, S. Urata, L. H. Mariani, A. S. Naik, S. Eddy, B. Zhang, Y. Wu, D. Salamon, J. C. Williams, X. Wang, K. S. Balderrama, P. J. Hoover, E. Murray, J. L. Marshall, T. Noel, A. Vijayan, A. Hartman, F. Chen, S. S. Waikar, S. E. Rosas, F. P. Wilson, P. M. Palevsky, K. Kiryluk, J. R. Sedor, R. D. Toto, C. R. Parikh, E. H. Kim, R. Satija, A. Greka, E. Z. Macosko, P. V. Kharchenko, J. P. Gaut, J. B. Hodgin, KPMP Consortium, M. T. Eadon, P. C. Dagher, T. M. El-Achkar, K. Zhang, M. Kretzler, S. Jain, An atlas of healthy and injured cell states and niches in the human kidney. *Nature* **619**, 585–594 (2023).

21. I. Ogobuiro, F. Tuma, "Physiology, Renal" in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2025; http://www.ncbi.nlm.nih.gov/books/NBK538339/).

22. K. Meliambro, J. C. He, K. N. Campbell, Podocyte-targeted therapies - progress and future directions. *Nat Rev Nephrol* **20**, 643–658 (2024).

23. N. Mann, H. Sun, A. J. Majmundar, Mechanisms of podocyte injury in genetic kidney disease. *Pediatr Nephrol* **40**, 1523–1538 (2025).

24. E. E. Groopman, M. Marasa, S. Cameron-Christie, S. Petrovski, V. S. Aggarwal, H. Milo-Rasouly, Y. Li, J. Zhang, J. Nestor, P. Krithivasan, W. Y. Lam, A. Mitrotti, S. Piva, B. H. Kil, D. Chatterjee, R. Reingold, D. Bradbury, M. DiVecchia, H. Snyder, X. Mu, K. Mehl, O. Balderes, D. A. Fasel, C. Weng, J. Radhakrishnan, P. Canetta, G. B. Appel, A. S. Bomback, W. Ahn, N. S. Uy, S. Alam, D. J. Cohen, R. J. Crew, G. K. Dube, M. K. Rao, S. Kamalakaran, B. Copeland, Z. Ren, J. Bridgers, C. D. Malone, C. M. Mebane, N. Dagaonkar, B. C. Fellström, C. Haefliger, S. Mohan, S. Sanna-Cherchi, K. Kiryluk, J. Fleckner, R. March, A. Platt, D. B. Goldstein, A. G. Gharavi, Diagnostic Utility of Exome Sequencing for Kidney Disease. *N Engl J Med* **380**, 142–151 (2019).

25. D. M. Connaughton, C. Kennedy, S. Shril, N. Mann, S. L. Murray, P. A. Williams, E. Conlon, M. Nakayama, A. T. van der Ven, H. Ityel, F. Kause, C. M. Kolvenbach, R. Dai, A. Vivante, D. A. Braun, R. Schneider, T. M. Kitzler, B. Moloney, C. P. Moran, J. S. Smyth, A.

Kennedy, K. Benson, C. Stapleton, M. Denton, C. Magee, C. M. O'Seaghdha, W. D. Plant, M. D. Griffin, A. Awan, C. Sweeney, S. M. Mane, R. P. Lifton, B. Griffin, S. Leavey, L. Casserly, D. G. de Freitas, J. Holian, A. Dorman, B. Doyle, P. J. Lavin, M. A. Little, P. J. Conlon, F. Hildebrandt, Monogenic causes of chronic kidney disease in adults. *Kidney Int* **95**, 914–928 (2019).

26. C. A. Gadegbeku, D. S. Gipson, L. B. Holzman, A. O. Ojo, P. X. K. Song, L. Barisoni, M. G. Sampson, J. B. Kopp, K. V. Lemley, P. J. Nelson, C. C. Lienczewski, S. G. Adler, G. B. Appel, D. C. Cattran, M. J. Choi, G. Contreras, K. M. Dell, F. C. Fervenza, K. L. Gibson, L. A. Greenbaum, J. D. Hernandez, S. M. Hewitt, S. R. Hingorani, M. Hladunewich, M. C. Hogan, S. L. Hogan, F. J. Kaskel, J. C. Lieske, K. E. C. Meyers, P. H. Nachman, C. C. Nast, A. M. Neu, H. N. Reich, J. R. Sedor, C. B. Sethna, H. Trachtman, K. R. Tuttle, O. Zhdanova, G. E. Zilleruelo, M. Kretzler, Design of the Nephrotic Syndrome Study Network (NEPTUNE) to evaluate primary glomerular nephropathy by a multidisciplinary approach. *Kidney Int* **83**, 749–756 (2013).

27. M. Martin, M. Patterson, S. Garg, S. O Fischer, N. Pisanti, G. W. Klau, A. Schöenhuth, T. Marschall, "WhatsHap: fast and accurate read-based phasing" (preprint, Bioinformatics, 2016); https://doi.org/10.1101/085050.

28. O. Delaneau, J.-F. Zagury, M. R. Robinson, J. L. Marchini, E. T. Dermitzakis, Accurate, scalable and integrative haplotype estimation. *Nat Commun* **10**, 5436 (2019).

29. C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, J. Marchini, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

30. G. Jun, M. Flickinger, K. N. Hetrick, J. M. Romm, K. F. Doheny, G. R. Abecasis, M. Boehnke, H. M. Kang, Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* **91**, 839–848 (2012).

31. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

32. R. Asiimwe, D. Alexander, STAR+WASP reduces reference bias in the allele-specific mapping of RNA-seq reads. *bioRxiv*, 2024.01.21.576391 (2024).

33. Y. Baran, M. Subramaniam, A. Biton, T. Tukiainen, E. K. Tsang, M. A. Rivas, M. Pirinen, M. Gutierrez-Arcelus, K. S. Smith, K. R. Kukurba, R. Zhang, C. Eng, D. G. Torgerson, C. Urbanek, GTEx Consortium, J. B. Li, J. R. Rodriguez-Santana, E. G. Burchard, M. A. Seibold, D. G. MacArthur, S. B. Montgomery, N. A. Zaitlen, T. Lappalainen, The landscape of genomic imprinting across diverse adult human tissues. *Genome Res* **25**, 927–936 (2015).

34. D. van Beek, J. Verdonschot, K. Derks, H. Brunner, T. M. de Kok, I. C. W. Arts, S. Heymans, M. Kutmon, M. Adriaens, Allele-specific expression analysis for complex genetic phenotypes applied to a unique dilated cardiomyopathy cohort. *Sci Rep* **13**, 564 (2023).

35. V. Tucci, A. R. Isles, G. Kelsey, A. C. Ferguson-Smith, Erice Imprinting Group, Genomic Imprinting and Physiological Processes in Mammals. *Cell* **176**, 952–965 (2019).

36. S. K. Han, M. T. McNulty, C. J. Benway, P. Wen, A. Greenberg, A. C. Onuchic-Whitford, Nephrotic Syndrome Study Network (NEPTUNE), D. Jang, J. Flannick, N. P. Burtt, P. C. Wilson, B. D. Humphreys, X. Wen, Z. Han, D. Lee, M. G. Sampson, Mapping genomic regulation of kidney disease and traits through high-resolution and interpretable eQTLs. *Nat Commun* **14**, 2229 (2023).

37. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

38. G. B. Loeb, P. Kathail, R. W. Shuai, R. Chung, Reinier J. Grona, S. Peddada, V. Sevim, S. Federman, K. Mader, A. Y. Chu, J. Davitte, J. Du, A. R. Gupta, C. J. Ye, S. Shafer, L. Przybyla, R. Rapiteanu, N. M. Ioannidis, J. F. Reiter, Variants in tubule epithelial regulatory elements mediate most heritable differences in human kidney function. *Nat Genet* **56**, 2078–2092 (2024).

39. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

40. L. Rossi, M. C. Nicoletti, M. Carmosino, L. Mastrofrancesco, A. Di Franco, F. Indrio, R. Lella, L. Laviola, F. Giorgino, M. Svelto, L. Gesualdo, G. Procino, Urinary Excretion of Kidney Aquaporins as Possible Diagnostic Biomarker of Diabetic Nephropathy. *J Diabetes Res* **2017**, 4360357 (2017).

41. C. Gao, W. Zhang, Urinary AQP5 is independently associated with eGFR decline in patients with type 2 diabetes and nephropathy. *Diabetes Res Clin Pract* **155**, 107805 (2019).

42. Y. Ren, M. Yu, D. Zheng, W. He, J. Jin, Lysozyme promotes renal fibrosis through the JAK/STAT3 signal pathway in diabetic nephropathy. *Arch Med Sci* **20**, 233–247 (2024).

43. Y. H. Lee, J.-W. Seo, M. Kim, D. Tae, J. Seok, Y. G. Kim, S.-H. Lee, J. S. Kim, H. S. Hwang, K.-H. Jeong, J.-Y. Moon, Urinary mRNA Signatures as Predictors of Renal Function Decline in Patients With Biopsy-Proven Diabetic Kidney Disease. *Front Endocrinol (Lausanne)* **12**, 774436 (2021).

44. P. Paueksakon, B. Najafian, C. E. Alpers, A. B. Fogo, AJKD Atlas of Renal Pathology: Lysozyme-Associated Nephropathy. *Am J Kidney Dis* **86**, e1–e3 (2025).

45. R. I. Menzies, F. W. Tam, R. J. Unwin, M. A. Bailey, Purinergic signaling in kidney disease. *Kidney Int* **91**, 315–323 (2017).

46. S. R. J. Taylor, C. M. Turner, J. I. Elliott, J. McDaid, R. Hewitt, J. Smith, M. C. Pickering, D. L. Whitehouse, H. T. Cook, G. Burnstock, C. D. Pusey, R. J. Unwin, F. W. K. Tam, P2X7 deficiency attenuates renal injury in experimental glomerulonephritis. *J Am Soc Nephrol* **20**, 1275–1281 (2009).

47. C. M. Turner, F. W. K. Tam, P.-C. Lai, R. M. Tarzi, G. Burnstock, C. D. Pusey, H. T. Cook, R. J. Unwin, Increased expression of the pro-apoptotic ATP-sensitive P2X7 receptor in experimental and human glomerulonephritis. *Nephrol Dial Transplant* **22**, 386–395 (2007).

48. Z. Zhu, Y. Cao, Y. Jian, H. Hu, Q. Yang, Y. Hao, H. Jiang, Z. Luo, X. Yang, W. Li, J. Hu, H. Liu, W. Liang, G. Ding, Z. Chen, CerS6 links ceramide metabolism to innate immune responses in diabetic kidney disease. *Nat Commun* **16**, 1528 (2025).

49. X. Wang, M. Song, X. Li, C. Su, Y. Yang, K. Wang, C. Liu, Z. Zheng, Y. Jia, S. Ren, W. Dong, J. Chen, T. Wang, L. Liu, M. Guan, C. Zhang, Y. Xue, CERS6-derived ceramides aggravate kidney fibrosis by inhibiting PINK1-mediated mitophagy in diabetic kidney disease. *Am J Physiol Cell Physiol* **325**, C538–C549 (2023).

50. Y. Guo, J. Min, B. Chang, L. Liu, J. Zhang, W. Chen, Modulation of podocyte extracellular matrix remodeling in membranous nephropathy by the NFATc3/LRRC55/BK channel pathway. *J Cell Commun Signal* **19**, e70022 (2025).

51. Y. Nie, X. Zhai, J. Li, A. Sun, H. Che, J. W. Christman, G. Chai, P. Zhao, M. Karpurapu, NFATc3 Promotes Pulmonary Inflammation and Fibrosis by Regulating Production of CCL2 and CXCL2 in Macrophages. *Aging Dis* **14**, 1441–1457 (2023).

52. J. D. Ooi, P.-Y. Gan, T. Chen, P. J. Eggenhuizen, J. Chang, M. A. Alikhan, D. Odobasic, S. R. Holdsworth, A. R. Kitching, FcγRIIB regulates T-cell autoreactivity, ANCA production, and neutrophil activation to suppress anti-myeloperoxidase glomerulonephritis. *Kidney Int* **86**, 1140–1149 (2014).

53. P. E. H. Sharp, J. Martin-Ramirez, S. M. Mangsbo, P. Boross, C. D. Pusey, I. P. Touw, H. T. Cook, J. S. Verbeek, R. M. Tarzi, FcγRIIb on myeloid cells and intrinsic renal cells rather than B cells protects from nephrotoxic nephritis. *J Immunol* **190**, 340–348 (2013).

54. H. A. Niederer, M. R. Clatworthy, L. C. Willcocks, K. G. C. Smith, FcgammaRIIB, FcgammaRIIIB, and systemic lupus erythematosus. *Ann N Y Acad Sci* **1183**, 69–88 (2010).

55. P. E. H. Sharp, J. Martin-Ramirez, P. Boross, S. M. Mangsbo, J. Reynolds, J. Moss, C. D. Pusey, H. T. Cook, R. M. Tarzi, J. S. Verbeek, Increased incidence of anti-GBM disease in Fcgamma receptor 2b deficient mice, but not mice with conditional deletion of Fcgr2b on either B cells or myeloid cells alone. *Mol Immunol* **50**, 49–56 (2012).

56. M. C. Caroleo, V. Carito, A. Pingitore, I. D. Perrotta, M. Perri, D. Mancuso, A. Russo, E. Cione, Human kidney podocyte cell population as a novel biological target of nerve growth factor. *Growth Factors* **33**, 14–22 (2015).

57. C. E. Alpers, K. L. Hudkins, M. Ferguson, R. J. Johnson, G. C. Schatteman, M. Bothwell, Nerve growth factor receptor expression in fetal, mature, and diseased human kidneys. *Lab Invest* **69**, 703–713 (1993).

58. L. Aloe, S. Rossi, L. Manni, Altered expression of nerve growth factor and its receptors in the kidneys of diabetic rats. *J Nephrol* **24**, 798–805 (2011).

59. H. Chen, Z. Zhang, L. Zhou, T. Cai, B. Liu, L. Wang, J. Yang, Identification of CCL19 as a Novel Immune-Related Biomarker in Diabetic Nephropathy. *Front Genet* **13**, 830437 (2022).

60. Y. Sato, M. Yanagita, Resident fibroblasts in the kidney: a major driver of fibrosis and inflammation. *Inflamm Regen* **37**, 17 (2017).

61. D. Chen, M. Shao, Y. Song, G. Ren, F. Guo, X. Fan, Y. Wang, W. Zhang, G. Qin, Single-cell RNA-seq with spatial transcriptomics to create an atlas of human diabetic kidney disease. *FASEB J* **37**, e22938 (2023).

62. C.-H. Chen, G.-M. Ke, P.-C. Lin, K.-D. Lin, Therapeutic DNA vaccine encoding CEMIP (KIAA1199) ameliorates kidney fibrosis in obesity through inhibiting the Wnt/β-catenin pathway. *Biochim Biophys Acta Gen Subj* **1865**, 130019 (2021).

63. T. Wada, N. Sakai, K. Matsushima, S. Kaneko, Fibrocytes: a new insight into kidney fibrosis. *Kidney Int* **72**, 269–273 (2007).

64. N. Sakai, T. Wada, H. Yokoyama, M. Lipp, S. Ueha, K. Matsushima, S. Kaneko, Secondary lymphoid tissue chemokine (SLC/CCL21)/CCR7 signaling regulates fibrocytes in renal fibrosis. *Proc Natl Acad Sci U S A* **103**, 14098–14103 (2006).

65. H.-L. Zhou, Y.-T. Wang, T. Gao, W.-G. Wang, Y.-S. Wang, Distribution and expression of fibroblast-specific protein chemokine CCL21 and chemokine receptor CCR7 in renal allografts. *Transplant Proc* **45**, 538–545 (2013).

66. P.-T. Xiao, J.-H. Hao, Y.-J. Kuang, C.-X. Dai, X.-L. Rong, L.-L. Jiang, Z.-S. Xie, L. Zhang, Q.-Q. Chen, E.-H. Liu, Targeting Neuraminidase 4 Attenuates Kidney Fibrosis in Mice. *Adv Sci (Weinh)* **11**, e2406936 (2024).

67. J. Zhao, T. Tian, J. Huang, H. Zha, L. Shi, Y. Yao, A novel role of ADAMTS16 in renal fibrosis through activating TGF-β/Smad signaling. *Cell Signal* **122**, 111347 (2024).

68. I. K. Vojtusek, M. Laganovic, M. Burek Kamenaric, S. Bulimbasic, S. Hrkac, G. Salai, V. Ivkovic, M. Coric, R. Novak, L. Grgurevic, First Characterization of ADAMTS-4 in Kidney Tissue and Plasma of Patients with Chronic Kidney Disease-A Potential Novel Diagnostic Indicator. *Diagnostics (Basel)* **12**, 648 (2022).

69. Y. Liu, Cellular and molecular mechanisms of renal fibrosis. *Nat Rev Nephrol* **7**, 684–696 (2011).

70. S. L. K. Bowers, S. Davis-Rodriguez, Z. M. Thomas, V. Rudomanova, W. C. Bacon, A. Beiersdorfer, Q. Ma, P. Devarajan, B. C. Blaxall, Inhibition of fibronectin polymerization alleviates kidney injury due to ischemia-reperfusion. *Am J Physiol Renal Physiol* **316**, F1293–F1298 (2019).

71. R. Bansal, S. Nakagawa, S. Yazdani, J. van Baarlen, A. Venkatesh, A. P. Koh, W.-M. Song, N. Goossens, H. Watanabe, M. B. Beasley, C. A. Powell, G. Storm, N. Kaminski, H. van

Goor, S. L. Friedman, Y. Hoshida, J. Prakash, Integrin alpha 11 in the regulation of the myofibroblast phenotype: implications for fibrotic diseases. *Exp Mol Med* **49**, e396 (2017).

72. Y. Zhang, J.-H. Zhu, Y. Zhou, Z.-T. Li, H. Liu, R.-X. Ma, Z.-L. Li, Activation of HIF-1α C-terminal transactivation domain promotes tubulointerstitial fibrosis through hexokinase 2-mediated metabolic reprogramming. *Cell Signal* **127**, 111531 (2025).

73. S. Huang, X. Lai, L. Yang, F. Ye, C. Huang, Y. Qiu, S. Lin, L. Pu, Z. Wang, W. Huang, Asporin Promotes TGF-β-induced Lung Myofibroblast Differentiation by Facilitating Rab11-Dependent Recycling of TβRI. *Am J Respir Cell Mol Biol* **66**, 158–170 (2022).

74. L. Medzikovic, L. Aryan, G. Ruffenach, M. Li, N. Savalli, W. Sun, S. Sarji, J. Hong, S. Sharma, R. Olcese, G. Fishbein, M. Eghbali, Myocardial fibrosis and calcification are attenuated by microRNA-129-5p targeting Asporin and Sox9 in cardiac fibroblasts. *JCI Insight* **8**, e168655 (2023).

75. C. Lopez-Silva, A. Surapaneni, I. M. Schmidt, D. Upadhyay, A. Srivastava, R. Palsson, I. E. Stillman, E. P. Rhee, S. S. Waikar, M. E. Grams, Circulating Protein and Metabolite Correlates of Histologically Confirmed Diabetic Kidney Disease. *Kidney Med* **6**, 100920 (2024).

76. K. M. Krawczyk, H. Nilsson, J. Nyström, D. Lindgren, K. Leandersson, K. Swärd, M. E. Johansson, Localization and Regulation of Polymeric Ig Receptor in Healthy and Diseased Human Kidney. *Am J Pathol* **189**, 1933–1944 (2019).

77. K. M. Sas, V. Nair, J. Byun, P. Kayampilly, H. Zhang, J. Saha, F. C. Brosius, M. Kretzler, S. Pennathur, Targeted Lipidomic and Transcriptomic Analysis Identifies Dysregulated Renal Ceramide Metabolism in a Mouse Model of Diabetic Kidney Disease. *J Proteomics Bioinform* **Suppl 14**, 002 (2015).

78. S. V. Parikh, A. Malvar, H. Song, V. Alberton, B. Lococo, J. Vance, J. Zhang, L. Yu, D. Birmingham, B. H. Rovin, Molecular imaging of the kidney in lupus nephritis to characterize response to treatment. *Transl Res* **182**, 1–13 (2017).

79. J.-D. Lu, K. Kadier, D. Dilixiati, B. Qiao, R. Nuer, A. Zebibula, M. Rexiati, K. Li, S. Li, Association between serum neurofilament light chain levels and chronic kidney disease: a cross-sectional population-based study from the National Health and Nutrition Examination Survey (2013-2014 cycle). *Ren Fail* **46**, 2427178 (2024).

80. E. van der Plas, O. Lullmann, L. Hopkins, J. L. Schultz, P. C. Nopoulos, L. A. Harshman, Associations between neurofilament light-chain protein, brain structure, and chronic kidney disease. *Pediatr Res* **91**, 1735–1740 (2022).

81. T. Makiishi, S. Araki, D. Koya, S. Maeda, A. Kashiwagi, M. Haneda, C-106T polymorphism of AKR1B1 is associated with diabetic nephropathy and erythrocyte aldose reductase content in Japanese subjects with type 2 diabetes mellitus. *Am J Kidney Dis* **42**, 943–951 (2003).

82. W. Obara, A. Iida, Y. Suzuki, T. Tanaka, F. Akiyama, S. Maeda, Y. Ohnishi, R. Yamada, T. Tsunoda, T. Takei, K. Ito, K. Honda, K. Uchida, K. Tsuchiya, W. Yumura, T. Ujiie, Y.

Nagane, K. Nitta, S. Miyano, I. Narita, F. Gejyo, H. Nihei, T. Fujioka, Y. Nakamura, Association of single-nucleotide polymorphisms in the polymeric immunoglobulin receptor gene with immunoglobulin A nephropathy (IgAN) in Japanese patients. *J Hum Genet* **48**, 293–299 (2003).

83. M. Gorski, A. Tin, M. Garnaas, G. M. McMahon, A. Y. Chu, B. O. Tayo, C. Pattaro, A. Teumer, D. I. Chasman, J. Chalmers, P. Hamet, J. Tremblay, M. Woodward, T. Aspelund, G. Eiriksdottir, V. Gudnason, T. B. Harris, L. J. Launer, A. V. Smith, B. D. Mitchell, J. R. O'Connell, A. R. Shuldiner, J. Coresh, M. Li, P. Freudenberger, E. Hofer, H. Schmidt, R. Schmidt, E. G. Holliday, P. Mitchell, J. J. Wang, I. H. de Boer, G. Li, D. S. Siscovick, Z. Kutalik, T. Corre, P. Vollenweider, G. Waeber, J. Gupta, P. A. Kanetsky, S.-J. Hwang, M. Olden, Q. Yang, M. de Andrade, E. J. Atkinson, S. L. R. Kardia, S. T. Turner, J. M. Stafford, J. Ding, Y. Liu, C. Barlassina, D. Cusi, E. Salvi, J. A. Staessen, P. M. Ridker, H. Grallert, C. Meisinger, M. Müller-Nurasyid, B. K. Krämer, H. Kramer, S. E. Rosas, I. M. Nolte, B. W. Penninx, H. Snieder, M. Fabiola Del Greco, A. Franke, U. Nöthlings, W. Lieb, S. J. L. Bakker, R. T. Gansevoort, P. van der Harst, A. Dehghan, O. H. Franco, A. Hofman, F. Rivadeneira, S. Sedaghat, A. G. Uitterlinden, S. Coassin, M. Haun, B. Kollerits, F. Kronenberg, B. Paulweber, N. Aumann, K. Endlich, M. Pietzner, U. Völker, R. Rettig, V. Chouraki, C. Helmer, J.-C. Lambert, M. Metzger, B. Stengel, T. Lehtimäki, L.-P. Lyytikäinen, O. Raitakari, A. Johnson, A. Parsa, M. Bochud, I. M. Heid, W. Goessling, A. Köttgen, W. H. L. Kao, C. S. Fox, C. A. Böger, Genome-wide association study of kidney function decline in individuals of European descent. *Kidney Int* **87**, 1017–1029 (2015).

84. A. Durand, C. A. Winkler, N. Vince, V. Douillard, E. Geffard, E. Binns-Roemer, D. K. Ng, P.-A. Gourraud, K. Reidy, B. Warady, S. Furth, J. B. Kopp, F. J. Kaskel, S. Limou, Identification of Novel Genetic Risk Factors for Focal Segmental Glomerulosclerosis in Children: Results From the Chronic Kidney Disease in Children (CKiD) Cohort. *Am J Kidney Dis* **81**, 635-646.e1 (2023).

85. M. Stafford-Smith, Y.-J. Li, J. P. Mathew, Y.-W. Li, Y. Ji, B. G. Phillips-Bute, C. A. Milano, M. F. Newman, W. E. Kraus, M. D. Kertai, S. H. Shah, M. V. Podgoreanu, Duke Perioperative Genetics and Safety Outcomes (PEGASUS) Investigative Team, Genome-wide association study of acute kidney injury after coronary bypass graft surgery identifies susceptibility loci. *Kidney Int* **88**, 823–832 (2015).

86. H. Okuda, K. Okamoto, M. Abe, K. Ishizawa, S. Makino, O. Tanabe, J. Sugawara, A. Hozawa, K. Tanno, M. Sasaki, G. Tamiya, M. Yamamoto, S. Ito, T. Ishii, Genome-wide association study identifies new loci for albuminuria in the Japanese population. *Clin Exp Nephrol* **24**, 1–9 (2020).

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

88. S. Xu, E. Hu, Y. Cai, Z. Xie, X. Luo, L. Zhan, W. Tang, Q. Wang, B. Liu, R. Wang, W. Xie, T. Wu, L. Xie, G. Yu, Using clusterProfiler to characterize multiomics data. *Nat Protoc* **19**, 3292–3320 (2024).

5

Acknowledgments:
This study has benefited from useful comments from members of Sampson lab and Lee lab.

We would like to acknowledge the Boston Children's Hospital High-Performance Computing Resources BCH HPC Clusters, Enkefalos 3 (E3), made available for conducting the research reported in this publication. Software used in the project was installed and configured by BioGrids. We would also like to acknowledge that Fig.1A was created with BioRender.com.

The Nephrotic Syndrome Study Network (NEPTUNE) is part of the Rare Diseases Clinical Research Network (RDCRN), which is funded by the National Institutes of Health (NIH) and led by the National Center for Advancing Translational Sciences (NCATS) through its Division of Rare Diseases Research Innovation (DRDRI). NEPTUNE is funded under grant number U54DK083912 as a collaboration between NCATS and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). Additional funding and/or programmatic support is provided by the University of Michigan, NephCure Kidney International and the Halpin Foundation. RDCRN consortia are supported by the RDCRN Data Management and Coordinating Center (DMCC), funded by NCATS and the National Institute of Neurological Disorders and Stroke (NINDS) under U2CTR002818.

The results here are in whole or part based upon data generated by the Kidney Precision Medicine Project (KPMP). Accessed June 26, 2025. https://www.kpmp.org. The KPMP is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) through the following grants: U01DK133081, U01DK133091, U01DK133092, U01DK133093, U01DK133095, U01DK133097, U01DK114866, U01DK114908, U01DK133090, U01DK133113, U01DK133766, U01DK133768, U01DK114907, U01DK114920, U01DK114923, U01DK114933, U24DK114886, UH3DK114926, UH3DK114861, UH3DK114915, and UH3DK114937. We gratefully acknowledge the essential contributions of our patient participants and the support of the American public through their tax dollars.

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

This research has been conducted using the UK Biobank Resource under Application Number 64945.

Funding:
National Institutes of Health grant F32DK122766 (ACO)
American Society of Nephrology KidneyCure Sharon Anderson Award (ACO)
NephCure Kidney International-NEPTUNE Pilot Award (ACO)
Brigham and Women's Hospital Hearst Young Investigator Award (ACO)
National Institutes of Health grant R01HG012871 (DL)
NEPTUNE Career Enhancement Award (DL)
Boston Children's Hospital OFD/BTREC/CTREC Faculty Development Fellowship Award (DL)
National Institutes of Health grant R01DK100449 (MB)
National Institutes of Health grant R21DK126329 (MB)
PRECISE has been supported by the University of Michigan and NIH/ NIDDK.

Author contributions:
Conceptualization: ACO, DL, MGS
Methodology: ACO, AG, CLO, DL, EDS, JS, JGY, LHM, MB, MTM
Investigation: ACO, CLO, DL, EDS, JS, MB, MTM, SB
Visualization: ACO, DL, EDS, JS, MTM
Funding acquisition: ACO, DL, MB, MGS
Project administration: ACO, DL
Supervision: ACO, DL, MGS
Writing – original draft: ACO, DL, EDS, JS
Writing – review & editing: ACO, AG, CLO, DL, EDS, JS, JGY, LHM, MB, MTM, MGS, SB

Competing interests:
MGS is on the scientific advisory board of Natera. All other authors have declared that they have no competing interests.

Data and materials availability:
Raw WGS and RNA-seq data from GLOM and TUBE are available through NEPTUNE (https://www.neptune-study.org/ancillary-studies). CONTROL GLOM and TUBE datasets are available through Dr. Markus Bitzer and NEPTUNE. Single cell RNA-seq data used to infer cell-type proportions of ASE genes can be accessed from KPMP repository (https://atlas.kpmp.org/repository). eQTL data are available through NephQTL2 (https://www.nephqtl2.org). The GTEx sQTL data used for the analyses described in this manuscript, GTEx_Analysis_v10_sQTL.tar, were obtained from: https://www.gtexportal.org/home/downloads/adult-gtex/qtl on 01/03/25. All code used for study analyses are publicly available in our Github repository (https://github.com/dongwonlee-lab/ASE_NEPTUNE).

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S29

Tables S1 to S11

References (*1–33*)

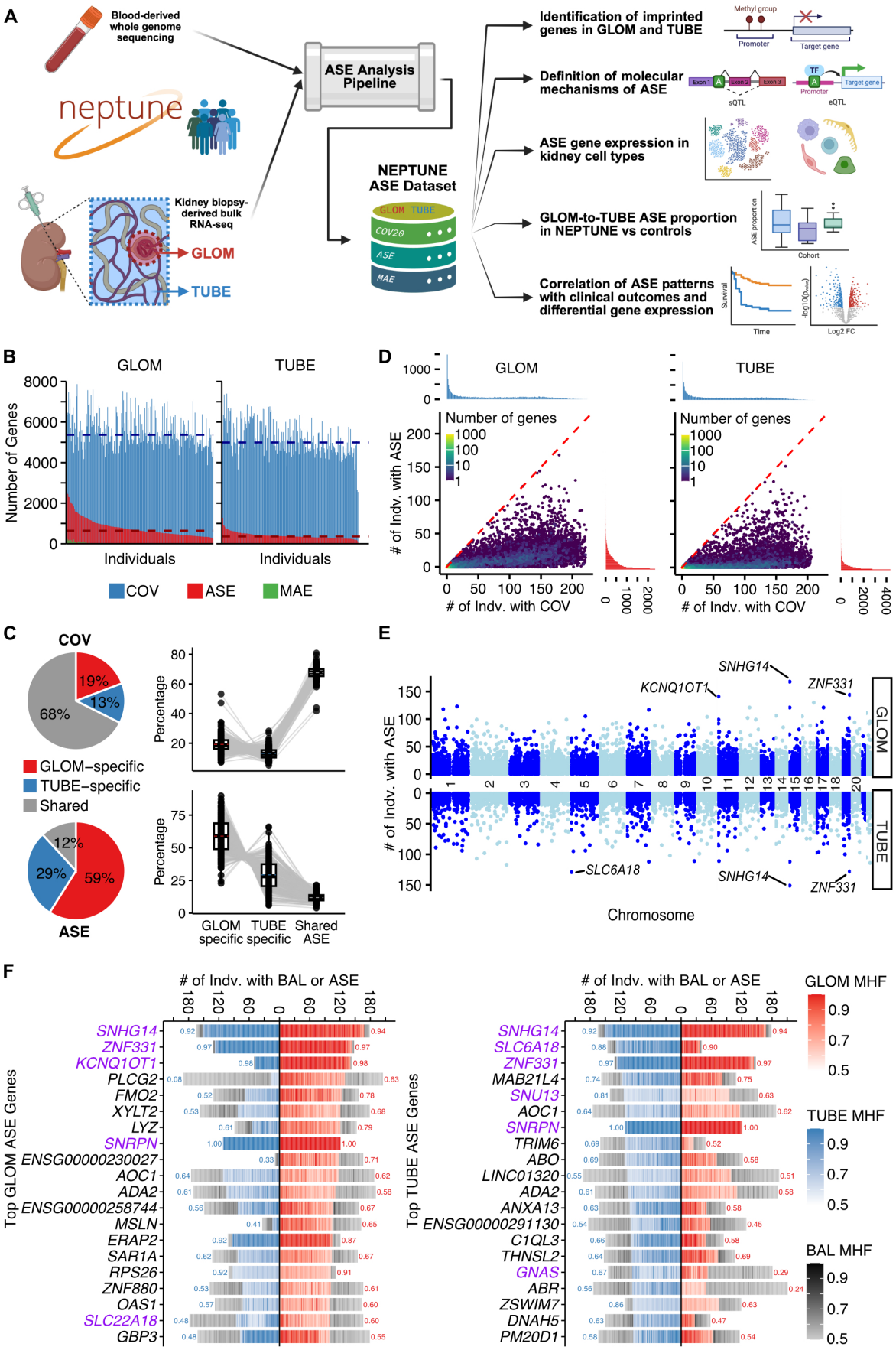
A
Blood-derived whole genome sequencing
neptune
Kidney biopsy-derived bulk RNA-seq
GLOM
TUBE
ASE Analysis Pipeline
NEPTUNE ASE Dataset
GLOM TUBE
COV20
ASE
MAE
Identification of imprinted genes in GLOM and TUBE
Methyl group
Promoter
Target gene
Definition of molecular mechanisms of ASE
Exon 1
Exon 2
Exon 3
sQTL
TF
eQTL
ASE gene expression in kidney cell types
GLOM-to-TUBE ASE proportion in NEPTUNE vs controls
ASE proportion
Cohort
Correlation of ASE patterns with clinical outcomes and differential gene expression
Survival
Time
Log2 FC
B
GLOM
TUBE
Number of Genes
Individuals
COV
ASE
MAE
C
COV
19%
13%
68%
GLOM–specific
TUBE–specific
Shared
12%
29%
59%
ASE
Percentage
GLOM specific
TUBE specific
Shared ASE
D
Number of genes
# of Indv. with ASE
# of Indv. with COV
E
SNHG14
KCNQ1OT1
ZNF331
SLC6A18
GLOM
TUBE
# of Indv. with ASE
Chromosome
F
# of Indv. with BAL or ASE
Top GLOM ASE Genes
SNHG14
ZNF331
KCNQ1OT1
PLCG2
FMO2
XYLT2
LYZ
SNRPN
ENSG00000230027
AOC1
ADA2
ENSG00000258744
MSLN
ERAP2
SAR1A
RPS26
ZNF880
OAS1
SLC22A18
GBP3
Top TUBE ASE Genes
SNHG14
SLC6A18
ZNF331
MAB21L4
SNU13
AOC1
SNRPN
TRIM6
ABO
LINC01320
ADA2
ANXA13
ENSG00000291130
C1QL3
THNSL2
GNAS
ABR
ZSWIM7
DNAH5
PM20D1
GLOM MHF
TUBE MHF
BAL MHF

Fig. 1. Study design and overview of ASE in kidney glomerular (GLOM) and tubulointerstitial (TUBE) compartments in NEPTUNE. (A) NEPTUNE is a unique consortium providing paired blood-derived whole genome sequencing (WGS) and kidney biopsy-derived bulk RNA-seq from microdissected glomerular (GLOM) and tubulointerstitial (TUBE) compartments from individuals with proteinuric kidney disease. By integrating an individual's WGS and RNA-seq through a rigorous, optimized ASE analysis pipeline (detailed in **Fig. S1**), we generated high-confidence genome-wide ASE data for the NEPTUNE cohort. This ASE dataset was used for subsequent molecular, cellular, epigenetic, and clinical analyses. (**B**) Number of genes amenable to analysis (COV, blue), significant allele-specific expression (ASE, red), and monoallelic expression (MAE, green) for each individual in the NEPTUNE GLOM and TUBE datasets. Stacked bars represent cumulative values. Dashed lines indicate the median number of COV (blue) and ASE (red) genes per individual. (**C**) Sharing of COV and ASE genes between GLOM and TUBE within an individual. In each individual, the overlap was defined as genes present in both compartments, while genes only in GLOM or only in TUBE were classified as GLOM-specific or TUBE-specific, respectively. Pie charts show the mean proportions of shared versus compartment-specific genes across individuals and boxplots show the distribution of these proportions. (**D**) The number of individuals with sufficient coverage (COV, x-axis) is compared to the number of individuals with ASE (y-axis). Each dot represents a gene. A heatmap represents point density. Marginal histograms are shown on both axes. (**E**) Analogous to a Manhattan plot, the number of individuals with ASE (y-axis) for an ASE gene is compared to its genomic coordinate (x-axis). The top plot is GLOM and the bottom is TUBE. The top three common ASE genes in each compartment are labeled. (**F**) The top 20 ASE genes in GLOM (left) and TUBE (right), ordered by number of individuals with ASE per cohort. The horizontal axis shows the number of individuals with ASE in GLOM (red) or TUBE (blue), and the number of individuals with balanced expression of that gene (BAL, defined as COV without ASE, and in grayscale). Individual stacked bars have color grading determined by major haplotypic fraction (MHF) and sorted (center to lateral) by the p-value (lowest to highest) from the binomial test. Imprinted genes are indicated in purple. The proportion of individuals with ASE among those with COV is indicated adjacent to each bar.

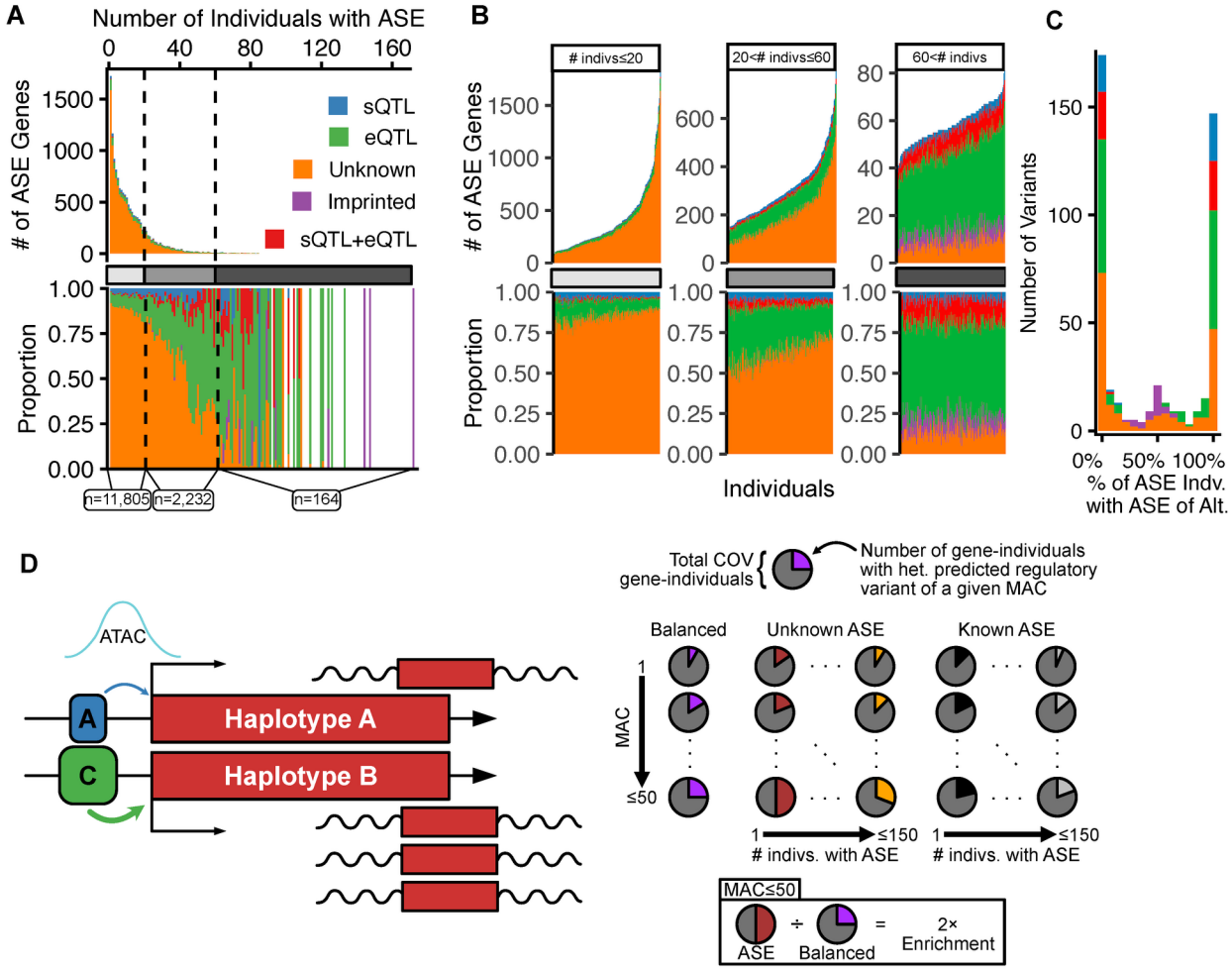
A
Number of Individuals with ASE
0
40
80
120
160
# of ASE Genes
1500
1000
500
0
sQTL
eQTL
Unknown
Imprinted
sQTL+eQTL
Proportion
1.00
0.75
0.50
0.25
0.00
n=11,805
n=2,232
n=164
B
indivs≤20
20<# indivs≤60
60<# indivs
# of ASE Genes
Proportion
Individuals
C
Number of Variants
150
100
50
0
0%
50%
100%
% of ASE Indv. with ASE of Alt.
D
ATAC
A
C
Haplotype A
Haplotype B
Total COV gene-individuals
Number of gene-individuals with het. predicted regulatory variant of a given MAC
Balanced
Unknown ASE
Known ASE
MAC
1
≤50
1
≤150
indivs. with ASE
MAC≤50
ASE
÷
Balanced
=
2×
Enrichment
E
GLOM
TUBE
MAC=1
MAC<5
MAC<50
Enrichment of Predicted Regulatory Variants in Promoter
1.8
1.4
1.0
Known
Unknown
ASE Mechanism
Balanced
Number of individuals with ASE (≤)
1
5
10
20
50
150

Fig. 2. Molecular mechanisms of ASE in NEPTUNE (**A**) Prevalence and mechanisms of ASE genes in NEPTUNE GLOM. Top panel is a histogram of ASE genes where units are single ASE genes, plotting the number of individuals with ASE of that gene (x-axis), and colored by the mechanism driving ASE. The bottom panel is the histogram with bars normalized to 1, showing proportion of the bar explained by each ASE mechanism. Dashed lines split the genes into three groups by number of individuals sharing the ASE gene. The total gene count per group is shown below. (**B**) Each bar represents one NEPTUNE individual, stratified by the three ASE gene groups defined in (A). The top panels show absolute numbers of ASE genes, color coded by each molecular mechanism, while the bottom panels show their proportions. (**C**) Histogram of ASE variants exhibiting their allelic directionality in GLOM, color coded by their molecular mechanisms. The x-axis represents the fraction of individuals with alternative allele preference among all individuals with ASE at that variant. Peaks at 0% and 100% indicate perfect allelic preference. **(D)** Schematic of the promoter variants analysis. The left plot shows the hypothesis that ASE of an unknown mechanism is driven by rare regulatory variants in promoters affecting chromatin accessibility (ATAC). The right plot shows our method for calculating enrichment. For all RNA COV gene-individuals in GLOM, we predicted all potential regulatory variants in their promoters using the ChromKid model. Then, we calculated the proportion of gene-individuals with a predicted regulatory variant of a given minor allele count (MAC), stratifying by ASE gene frequency (number of individuals with ASE of the gene) and Known vs. Unknown ASE mechanisms. This proportion was further normalized by the proportion found in balanced genes at a given MAC to calculate the enrichment. **(E)** Enrichment of predicted regulatory variants in the promoters of ASE gene-individuals, comparing Known and Unknown ASE mechanism to balanced gene-individuals in GLOM (left) and TUBE (right). Each dot represents a group of ASE gene-individuals, defined by their mechanism (Known/Unknown), ASE frequency, and the MAC of the predicted regulatory variants. The y-axis shows enrichment relative to balanced genes (pink line = 1.0). The significance was determined by a binomial test (alternative = greater) of the observed proportion against the expected probability, defined by the proportion of balanced gene-individuals with predicted regulatory variants at the corresponding MAC. ($p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***)

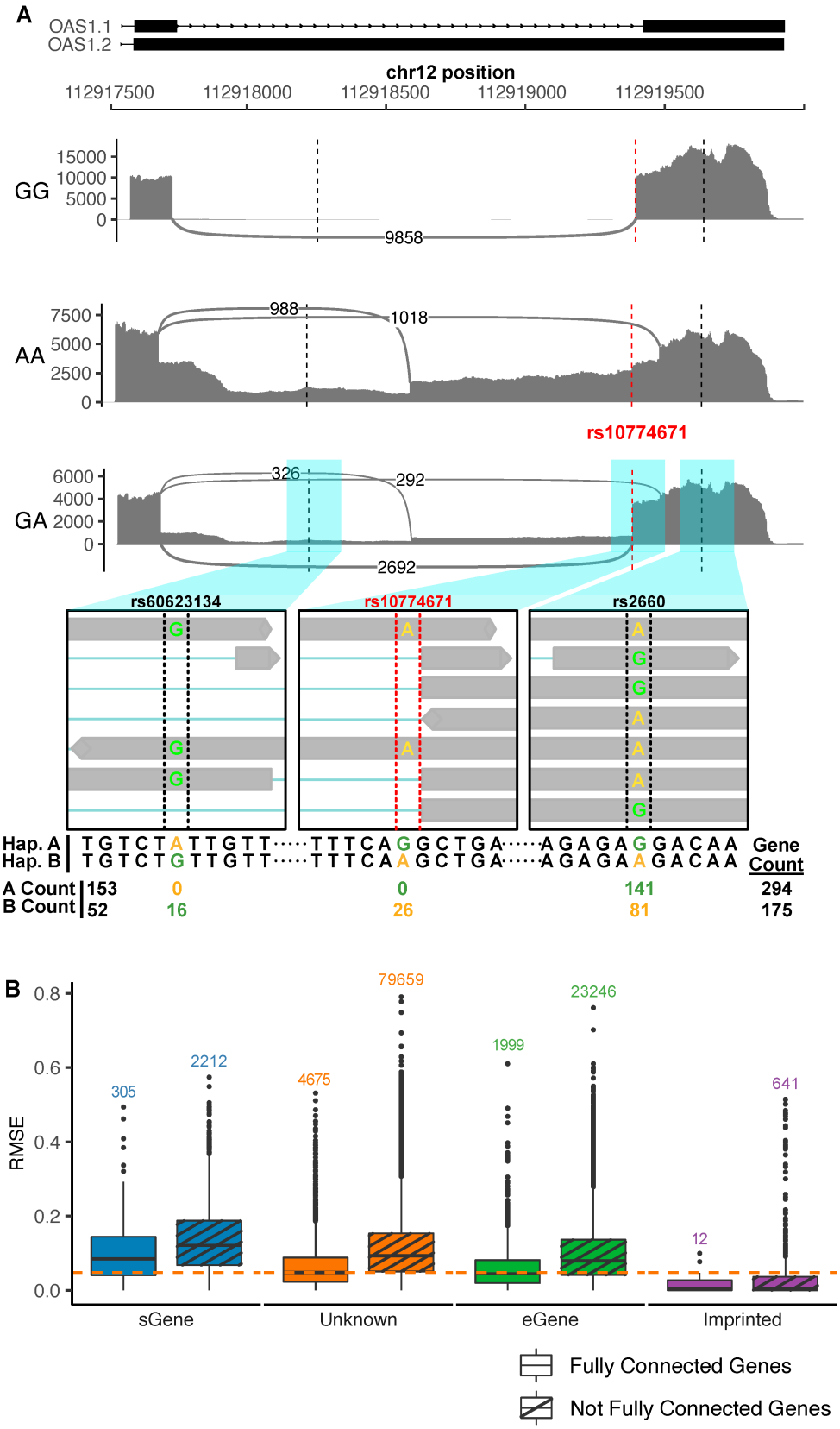
A
OAS1.1
OAS1.2
chr12 position
112917500
112918000
112918500
112919000
112919500
GG
15000
10000
5000
0
9858
AA
7500
5000
2500
0
988
1018
rs10774671
GA
6000
4000
2000
0
326
292
2692
rs60623134
rs10774671
rs2660
Hap. A
Hap. B
Gene Count
A Count
B Count
153
52
0
16
0
26
141
81
294
175
B
RMSE
0.8
0.6
0.4
0.2
0.0
305
2212
4675
79659
1999
23246
12
641
sGene
Unknown
eGene
Imprinted
Fully Connected Genes
Not Fully Connected Genes

Fig. 3. Alternative splicing and phasing errors complicate gene-level ASE analysis (**A**) Sashimi plot showing alternative splicing of the 3'-end of *OAS1*, promoted by the sQTL rs10774671 (red). Vertical dashed lines denote the genomic locations of three SNPs whose RNA-seq alignments in a heterozygous individual are shown below the plot. Variant-, haplotype-, and gene-level ASE counts are shown below the alignment. (**B**) RMSE of the haplotype A fraction (reads aligning to haplotype A / total reads) using variant-level haplotype A fraction as observed values and the gene-level haplotype A fraction as the predicted value, stratified by ASE mechanism. Gene-individuals were separated into "Fully Connected Genes" (where all variants are phased by RNA reads) and "Not Fully Connected Genes" (relying on population phasing). Numbers above boxplots indicate the count of gene-individuals in each category. sGenes were separated into "Mixed" (variants separated by the splicing event) and "Uniform" (all variants on the same side of the splicing event) types, and only Mixed ones are shown. The orange line indicates the median of the Unknown+Connected group. Pairwise Wilcoxon tests showed that all groups had significant mean differences ($P < 0.001$).

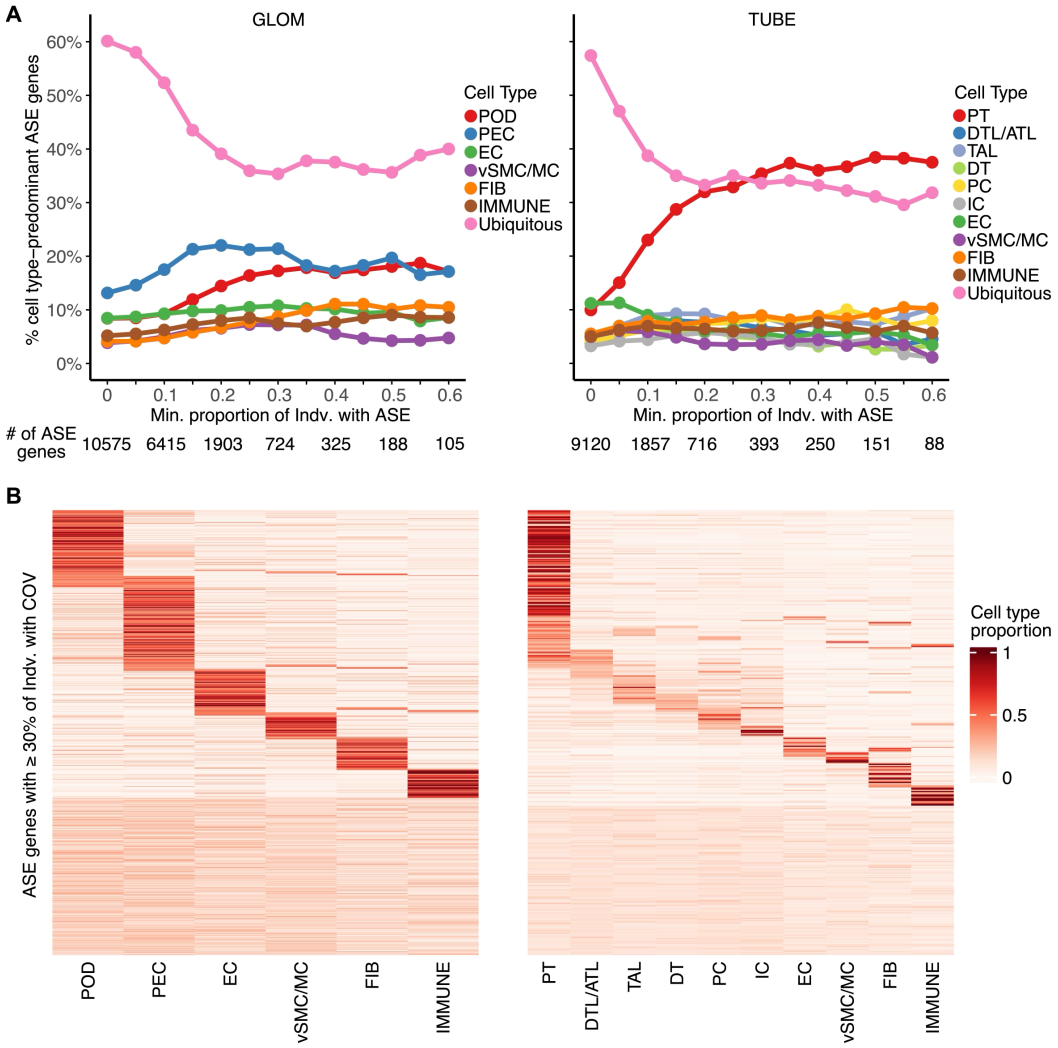

Fig. 4. Cell-type specificity of ASE genes. (**A**) Trends in cell-type predominance of ASE genes in GLOM (left) and TUBE (right) as a function of ASE frequency. The x-axis is the minimum proportion of individuals with ASE among individuals with COV, and the y-axis is the percentage of ASE genes dominantly expressed in a given cell type. Colors represent cell types in each compartment, and numbers below the x-axis denote the total number of ASE genes meeting the corresponding threshold. (**B**) Heatmaps of cell-type proportions for common ASE genes (i.e., ASE observed in ≥30% individuals with COV) in GLOM (left) and TUBE (right). Rows represent individual ASE genes; columns represent cell types; color scale indicates cell-type proportion. All analyses were restricted to genes with COV in ≥10% individuals.

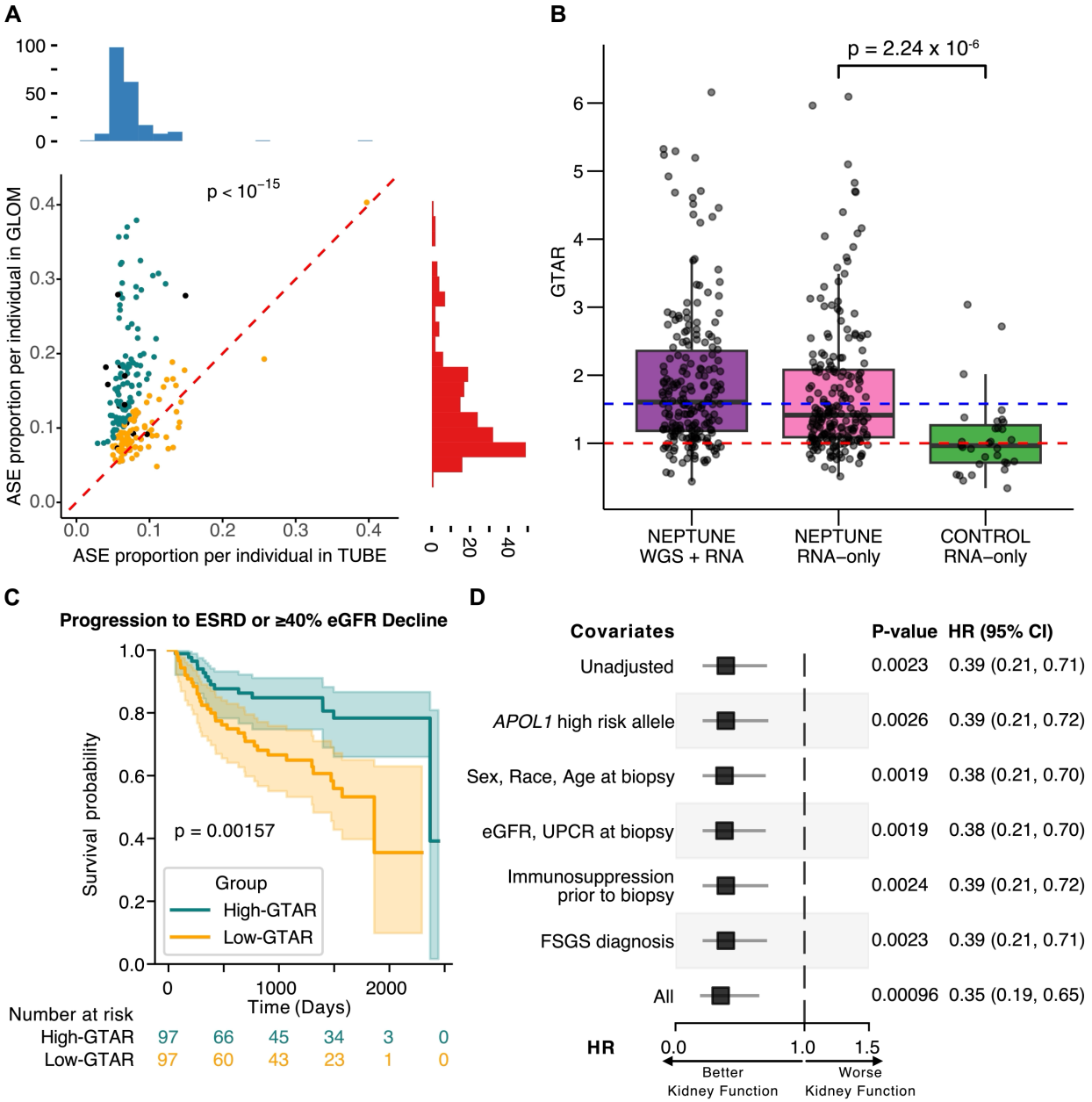

Fig. 5. Comparison of high- and low-GTAR NEPTUNE groups. (**A**) Scatterplot of genome-wide ASE proportion in paired GLOM and TUBE samples. High-GTAR samples (n = 97) are colored in teal and low-GTAR samples (n = 97) are yellow. Samples without endpoint data are gray (n = 12, not included in survival analysis). Marginal histograms illustrate the distribution of genome-wide ASE proportion in each compartment. The Wilcoxon signed-rank test was used to compare GLOM and TUBE proportions. The red dashed line indicates equal GLOM and TUBE ASE proportions. (**B**) GTAR per individual, by cohort and ASE analysis pipeline. Each dot is an individual with paired GLOM and TUBE data. The blue dashed line represents the median GTAR for NEPTUNE (WGS + RNA), and the red dashed line indicates GTAR = 1. (C) Kaplan–Meier survival analysis of NEPTUNE (n = 194), comparing the high- and low-GTAR groups. (D) Forest plot of Cox proportional hazards model, with hazard ratio (HR) in reference to high-GTAR group, adjusted for multiple kidney disease-specific covariates. GTAR, GLOM-to-TUBE ASE ratio.

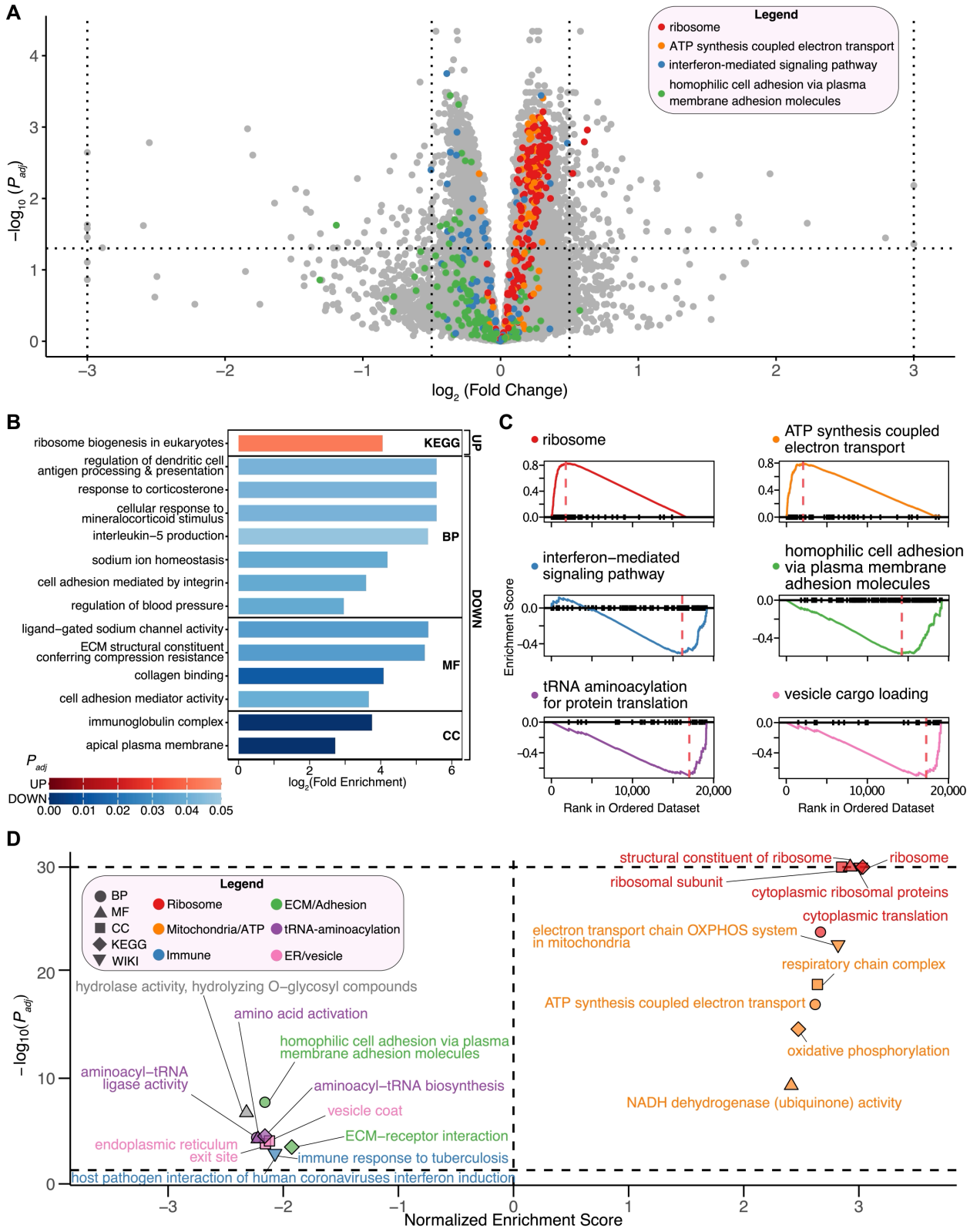
A
Legend
ribosome
ATP synthesis coupled electron transport
interferon-mediated signaling pathway
homophilic cell adhesion via plasma membrane adhesion molecules
−log10 (Padj)
log2 (Fold Change)
B
ribosome biogenesis in eukaryotes
regulation of dendritic cell antigen processing & presentation
response to corticosterone
cellular response to mineralocorticoid stimulus
interleukin−5 production
sodium ion homeostasis
cell adhesion mediated by integrin
regulation of blood pressure
ligand−gated sodium channel activity
ECM structural constituent conferring compression resistance
collagen binding
cell adhesion mediator activity
immunoglobulin complex
apical plasma membrane
KEGG
BP
MF
CC
UP
DOWN
log2(Fold Enrichment)
Padj
UP
DOWN
0.00 0.01 0.02 0.03 0.04 0.05
C
ribosome
ATP synthesis coupled electron transport
interferon−mediated signaling pathway
homophilic cell adhesion via plasma membrane adhesion molecules
tRNA aminoacylation for protein translation
vesicle cargo loading
Enrichment Score
Rank in Ordered Dataset
D
Legend
BP
MF
CC
KEGG
WIKI
Ribosome
Mitochondria/ATP
Immune
ECM/Adhesion
tRNA-aminoacylation
ER/vesicle
structural constituent of ribosome
ribosome
ribosomal subunit
cytoplasmic ribosomal proteins
cytoplasmic translation
electron transport chain OXPHOS system in mitochondria
respiratory chain complex
ATP synthesis coupled electron transport
oxidative phosphorylation
NADH dehydrogenase (ubiquinone) activity
hydrolase activity, hydrolyzing O−glycosyl compounds
amino acid activation
homophilic cell adhesion via plasma membrane adhesion molecules
aminoacyl−tRNA ligase activity
aminoacyl−tRNA biosynthesis
vesicle coat
ECM−receptor interaction
endoplasmic reticulum exit site
immune response to tuberculosis
host pathogen interaction of human coronaviruses interferon induction
−log10(Padj)
Normalized Enrichment Score

Fig. 6. Differential glomerular gene expression between high- and low-GTAR groups and gene set enrichment analyses (GSEA). (**A**) Volcano plot of differentially expressed genes (DEGs) with the low GTAR group as reference. The horizontal dashed line indicates the adjusted p-value threshold (0.05), and the inner and outer vertical dashed lines indicate $\log_2$(fold change) cutoffs of ±0.5 and ±3, respectively. Genes exceeding ±3 are aligned with the outer vertical lines. Significantly upregulated GO terms related to ribosome and mitochondrial/ATP synthesis (red and orange) and downregulated immune-related and membrane-associated pathways (blue and green) are highlighted. The corresponding GO term names are shown in *(C)* with matching colors. (**B**) ORA of significantly upregulated ($\log_2$FC > 0.5, P_{adj} < 0.05) and downregulated ($\log_2$FC < –0.5, P_{adj} < 0.05) genes across multiple databases. Bar colors reflect adjusted p-values from ORA, with red for upregulated and blue for downregulated terms. (**C**) GSEA enrichment plots of significant GO terms, grouped into six functional categories: ribosome (red), mitochondria/ATP (orange), immune (blue), ECM/adhesion (green), tRNA-aminoacylation (purple), and ER/vesicle (pink). (**D**) GSEA summary of the top upregulated and downregulated GO terms ranked by NES across multiple databases. Terms with $-\log_{10}(P_{adj})$ > 30 are truncated and aligned at the horizontal dashed line. Colors correspond to the six categories in *(C)* and grey color denotes GO terms not included in these categories. Shapes indicate the database source.