

GENETICS

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}, Victoria Mezger^{2,3,4,5}, Christopher L. O'Connor⁶, Anya Greenberg^{2,3,4}, Jihoon G. Yoon^{2,7,8}, Sowmya Badina^{2,9}, Laura H. Mariani⁶, Markus Bitzer⁶, Matthew G. Sampson^{2,3,4,*}, Dongwon Lee^{2,3,4,*}

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 sequencing from microdissected glomerular and tubulointerstitial 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 glomeruli (GLOM) than tubulointerstitium (TUBE) compared to controls. Unexpectedly, higher ASE in GLOM than TUBE was significantly associated with improved kidney disease outcomes in the disease cohort. Differential gene expression analysis suggested that this was the result of an active, protective transcriptional response, including ribosome and adenosine 5'-triphosphate synthesis up-regulation, 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 affected by nongenetic factors (3, 4). Thus, ASE analysis is a versatile method to assess genomic and epigenetic phenomena, including noncoding cis-regulatory variants, genomic imprinting, and alternative splicing (5).

As an intraindividual 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 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, 11).

Prior studies for ASE analysis have often relied solely on RNA sequencing (RNA-seq) or on single-nucleotide polymorphism (SNP)

arrays for genotyping (12, 13). 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 (MAE). 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, 8).

ASE has been extensively studied in diseases such as cancer, neurodevelopmental disorders, and heart disease, with demonstrated instances of causality (5, 14–19). However, ASE evaluation in kidney has previously focused on nondiseased tissue either from tumor nephrectomies (12, 13, 20) or postmortem samples (21). Furthermore, datasets such as Genotype-Tissue Expression (GTEx) (22) do not account for the complexity of the kidney structure, which comprises two distinct functional compartments: the glomeruli (GLOM) and tubulointerstitium (TUBE) (23). As GLOM comprise only a small fraction of the total cell count in the kidney cortex (24), glomerular transcripts are much less abundant in bulk tissue than those of the TUBE. 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 resulting from defects in glomerular filtration (25). The most severe form of proteinuric disease is the nephrotic syndrome, with loss of ≥ 3.5 g of urinary protein daily, resulting in hypoalbuminemia and swelling of the face and extremities. Proteinuric disorders account for up to 90% of end-stage kidney disease (ESKD) worldwide (26). Specifically, a common mechanism is injury to glomerular podocytes (PODs; specialized cells that comprise only $\sim 1\%$ of kidney cortex), and most of the >70 monogenic causes of proteinuria involve genes expressed in this cell type (27). However, many patients with glomerular disorders remain undiagnosed,

¹Division of Nephrology, Brigham and Women's Hospital, Boston, MA 02115, USA.

²Division of Nephrology, Boston Children's Hospital, Boston, MA 02115, USA. ³Harvard Medical School, Boston, MA 02115, USA. ⁴Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. ⁵Medical Faculty, Heidelberg University, Heidelberg 69120, Germany. ⁶Department of Medicine, University of Michigan, Ann Arbor, MI 48109, USA. ⁷Department of Laboratory Medicine, Yonsei University College of Medicine, Seoul 03722, Republic of Korea. ⁸Department of Laboratory Medicine, Gangnam Severance Hospital, Seoul 06273, Republic of Korea. ⁹Centre for Digital Health and Precision Medicine, The Apollo University, Chittoor 517127, India.

*Corresponding author. Email: matthew.sampson@childrens.harvard.edu (M.G.S.); dongwon.lee@childrens.harvard.edu (D.L.)

†These authors contributed equally to this work.

and several known genetic etiologies lack a clear mechanistic pathway to disease (28, 29).

The goal of this study was to find 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) (30). We create a rigorous, generalizable framework for ASE analysis, integrating both available and novel computational tools. Our pipeline includes 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 the GLOM and TUBE of diseased kidneys compared to controls. We then perform extensive investigation into the molecular mechanisms, cell-type specificity, clinical correlations, and differential gene signatures linked to ASE in the human kidney.

RESULTS

Workflow for generation of high-confidence ASE data

We assembled a rigorous pipeline for high-confidence ASE calling in kidney tissue from the NEPTUNE proteinuric kidney disease cohort (see Materials and Methods). Biopsies from these individuals were microdissected into GLOM and TUBE compartments, followed by bulk RNA-seq. We identified 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 Materials and Methods. Here, we highlight three key features introduced in this work: (i) novel QC steps to blacklist variants prone to genotyping errors, (ii) bias reduction from misassigned RNA reads in loci containing overlapping genes, and (iii) 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 (31) and SHAPEIT4 (32). Both tools use read-backed phasing, an advantage of WGS, and SHAPEIT4 allowed us to incorporate haplotype data from ~500,000 participants of the UK Biobank (UKB) (33) to leverage population-level phasing information. As detailed below, our estimated gene-level phasing error (flip) rate was low (~3%), enabling accurate ASE quantification even for genes whose haplotypes were not fully connected by RNA-seq reads.

To ensure high-confidence genotype calling, we also implemented multiple WGS QC measures, excluding variants with (i) inconsistent genotype before and after phasing, (ii) biased DNA allelic fraction, or (iii) abnormal DNA read depth. These variants showed a substantial 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 biallelic RNA expression at homozygous WGS genotypes, suggesting systematically poor read alignment in RNA-seq (fig. S3). Together, 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). Per sample, an average of 302 genes harbored these problematic variants and thus would appear to have significant ASE without this filter. Our QC strategy also identified and excluded 18 problematic samples with an excess number of such problematic variants ($n \geq 1000$) (fig. S2D). Our findings were

further confirmed by verifyBamID analysis (34), 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 (Spliced Transcripts Alignment to a Reference) with WASP (Workflow for Allele-Specific Processing) (9, 35, 36) and confirmed the same benefits in our dataset (fig. S5). WASP-passing reads, taken only from transcripts overlapping heterozygous loci, represented ~9% of total reads per sample. We then performed ASE analysis using phASER (phasing and Allele Specific Expression from RNA-seq) (6), incorporating the individualized blacklists described above. phASER uses read-backed phasing from RNA-seq to optimize phasing of rare variants and accurately phases variants over longer distances resulting from RNA splicing. We did not analyze human leukocyte antigen (HLA) genes in accordance with the phASER pipeline because of the complexities of ASE determination in this gene group. Similarly, we excluded sex chromosome genes to avoid the confounding biological effects of hemizygosity in males and X-chromosome inactivation in females. In addition, phASER reports gene-level ASE as the number of reads mapping to gene haplotypes, rather than the sum of a gene's variant-level reads. This gene-level aggregation is crucial for mitigating noise from individual variants; our evaluation of variant-level data revealed that non-exonic variants exhibited artificial allelic imbalance driven by low RNA-seq read coverage (fig. S6). For downstream analyses, we included genes with sufficient coverage [total haplotypic count ≥ 20 , on the basis of prior literature standards (12, 37)] and designated these as "COV" genes. COV genes without detectable ASE were designated balanced (BAL). Consistently, at lower coverage cutoffs, our data showed inflated ASE and MAE (defined as the expression of only one haplotype) (fig. S7).

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. S8 and Materials and 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. S9). 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. S10 and table S2). Going forward, we kept only "resolved" and "unresolved" genes in downstream analyses—opting to include the latter to avoid missing potential biological signals. However, "unresolved" genes rarely exhibited ASE because of diluted signals from multiple genes.

To quantify the magnitude of ASE, we defined the term "major haplotypic fraction" (MHF): the fraction of reads from the higher-expressed haplotype, with values ranging from 0.5 (balanced) to 1.0 (monoallelic). We defined a gene as having significant ASE as follows: (i) sufficient RNA coverage for high-confidence ASE analysis, i.e., post-TOGA COV genes; (ii) binomial test with $P_{adj} < 0.05$; and

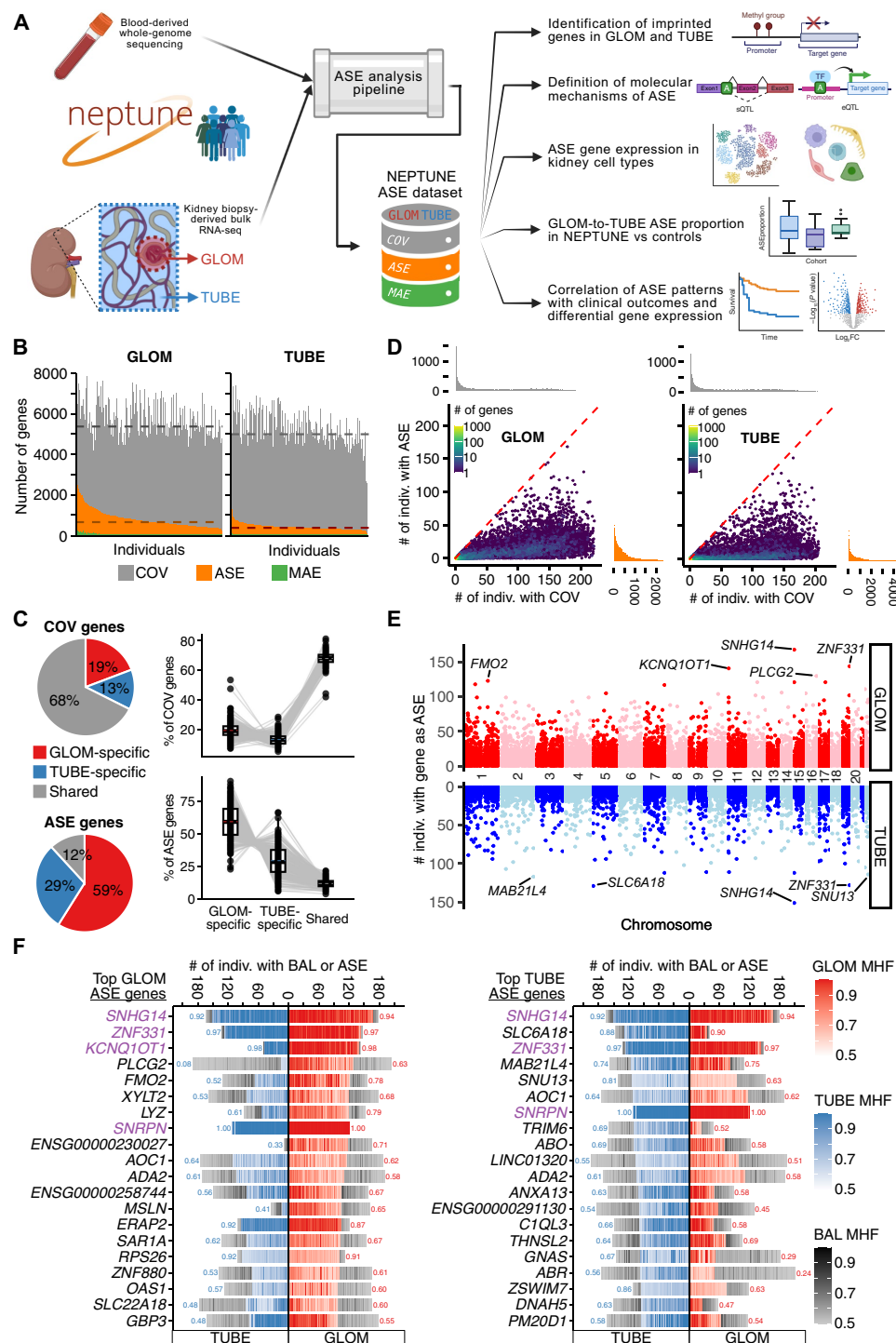


Fig. 1. Study design and overview of ASE in kidney glomerular and tubulointerstitial compartments in NEPTUNE. (A) Schematic overview of our study. Partially created in BioRender. D. Lee (2026), <https://biorender.com/7mg6zwd>. (B) Individual-level data: Each bar represents one individual, ordered by number of ASE genes; stacked bars are cumulative. Dashed lines indicate the median of COV (gray) and ASE (orange) genes per sample. (C) Overlap of COV (top) and ASE (bottom) genes between GLOM and TUBE in paired samples ($n = 206$). Pie charts show the mean proportions of shared and compartment-specific genes across all individuals. Box plots show the distribution of these proportions. (D) Gene-level data: Each dot represents a gene, with the number of individuals of this gene as COV (x axis), compared to that of ASE (y axis). The heatmap represents point density to indicate overlapping data points. Marginal histograms are shown on both axes. (E) The number of individuals sharing an ASE gene (y axis) is plotted (GLOM: top; TUBE: bottom) against its genomic coordinate (x axis). The five most common ASE genes are labeled. (F) The top 20 ASE genes in GLOM (left) and TUBE (right), ordered by number of ASE individuals of this gene. The horizontal axis shows the number of ASE and BAL individuals in GLOM or TUBE. Individual stacked bars have color grading determined by MHF and sorted by the binomial test P value. Imprinted genes (from our imprinting analysis, Fig. 2A) are indicated in purple. The proportion of individuals with ASE among those with COV is shown as the number adjacent to each bar. Indiv, individual.

(iii) MHF ≥ 0.6 (i.e., one haplotype expressed $\geq 50\%$ 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; a similar approach and cutoff have been described previously (3, 7). Figure S11 shows an illustrative ASE analysis plot in one NEPTUNE individual.

Landscape of ASE in human kidneys with proteinuric disease

In our ASE analysis, we achieved extensive coverage of heterozygous variants (table S3), resulting in a median of 5370 genes in GLOM and 4989 in TUBE amenable to ASE analysis (COV genes) per individual (Fig. 1B and table S4). Cumulatively across NEPTUNE GLOM and TUBE, we observed 18,304 COV genes in ≥ 1 individual, of which 14,619 ($\sim 80\%$) were protein-coding—thus, most genes in the human genome were covered. 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 (Fig. 1B and table S4). Moreover, while COV genes were largely shared (mean 68%) between GLOM and TUBE compartments in an individual, only a mean of 12% of ASE genes was similarly shared. Thus, ASE was predominantly compartment-specific, with more than double the proportion of GLOM-specific ASE compared to TUBE-specific ASE (59% versus 29%) (Fig. 1C). MAE genes were rare and similarly distinct, with a median of 14 GLOM and 8 TUBE MAEs per individual. Collectively, we identified 3575 GLOM and 968 TUBE MAEs in ≥ 1 individual (table S4). All ASE genes are reported in table S5.

COV and ASE distribution per gene is shown in Fig. 1D. Most COV genes ($\sim 66\%$ in GLOM and $\sim 67\%$ in TUBE) were shared by $\geq 10\%$ of individuals. However, concordant with a prior study (19), ASE was largely private. Shared ASE genes followed an exponentially decreasing distribution (Fig. 1D and fig. S12), with the majority having ASE in few individuals (≤ 8 in GLOM and ≤ 3 in TUBE). Only 13.7% (GLOM) and 6.7% (TUBE) of ASE genes had ASE in $\geq 10\%$ of individuals, and we later find that these genes share certain molecular mechanisms of ASE, such as imprinting and strong expression quantitative trait locus (eQTL) effects. Furthermore, the proportion of individuals exhibiting ASE per gene was strongly correlated between males and females in both GLOM and TUBE (fig. S13). Overall, this indicates that 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 region, 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). In addition, none of the top 50 most common ASE genes demonstrated significant differences in racial composition, indicating that these common ASE events are not population-specific (fig. S14). 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 often be detected for imprinted genes with sufficient coverage (fig. S15).

We also compared our results with recent kidney multiomic data from Liu *et al.* (20). At the gene level, 58.4% of NEPTUNE TUBE ASE genes overlapped with Liu's ASE dataset (fig. S16A). This proportion further increased when focusing on more commonly shared

ASE genes (fig. S16B). As expected, the vast majority of these shared signals displayed highly consistent effect sizes and directionality across both compartments (fig. S16, C to F). Furthermore, we found that NEPTUNE ASE genes were significantly enriched for upstream bulk allele-specific chromatin accessibility (bASA) peaks [OR (odds ratio) = 1.24, $P < 2.2 \times 10^{-16}$; fig. S17, A and B]. Last, leveraging single-nucleus allele-specific chromatin accessibility (snASA) data as a proxy for cell-type specificity of our ASE signals, we confirmed that cell type-specific ASE contributions remained uniform across individuals (fig. S17C), suggesting that cellular heterogeneity is unlikely to drive variation in ASE events across individuals.

Identification of imprinted genes in human proteinuric kidneys

Given the enrichment of imprinted genes among the top ASE genes (Fig. 1F and fig. S15) and their known tissue specificity (37), we sought to systematically identify imprinted genes within our dataset. To do this, we adapted an established pipeline (37) that assigns an imprinting likelihood to every ASE gene. This is based on characteristics distinct from other regulatory mechanisms: MAE in all individuals where the expressed haplotype varies randomly across individuals because of parent-of-origin effects rather than the genotype (fig. S18). With this pipeline, we initially identified 26 candidate imprinted genes in GLOM (nine not previously identified in any tissue) and 18 candidates in TUBE (five previously unidentified) (figs. S19 to S22).

This method has a potential limitation: the risk of genes with eQTLs of very large effect size (leading to near-MAE) being classified as imprinted. This can occur if the causal noncoding regulatory variant is heterozygous in most individuals and in low linkage disequilibrium with the coding ASE variants. While the prior study concluded that genes meeting these conditions are extremely unlikely (37), we found that our previously unidentified candidate imprinted genes frequently overlapped with strong eQTLs (38). Specifically, seven of nine such candidates in GLOM had significant eQTLs ($P < 10^{-6}$) compared to only 1 of 17 known imprinted genes. A similar pattern was also observed in TUBE. Therefore, we sought to distinguish true imprinting from strong eQTL effects by implementing an additional step that incorporates genotype data and eQTL effect direction to test which mechanism better explained the observed MAE (Materials and Methods). We found that most of these candidates were better explained by strong eQTL effects, including eight GLOM genes (fig. S19) and four TUBE genes (fig. S21). In TUBE, one known imprinted gene, *MEST*, was also better explained by an eQTL (fig. S22).

Adding this step reduced the number of identified imprinted genes to 18 in GLOM and 13 in TUBE (Fig. 2A). Thus, we ultimately found two previously unidentified candidate imprinted genes: *IGLV6-57* (immunoglobulin lambda variable 6-57) in GLOM and *GLYATL2* (glycine-*N*-acetyltransferase like 2) in TUBE. We found many known imprinted genes to be uniquely imprinted in each compartment (seven in GLOM and two in TUBE), but the absence of imprinting in the other compartment was largely due to insufficient ASE data (i.e., low expression), rather than observed BAL expression.

To test the tissue specificity of imprinted GLOM and TUBE genes, we compared them to a list of 299 known human imprinted genes (table S6 and fig. S23), which we curated on the basis of literature review (37, 39). More than 85% of these 299 genes did not have ASE data in GLOM or TUBE because of lack of heterozygous variants, low expression, and/or removal by TOGA. However, most of

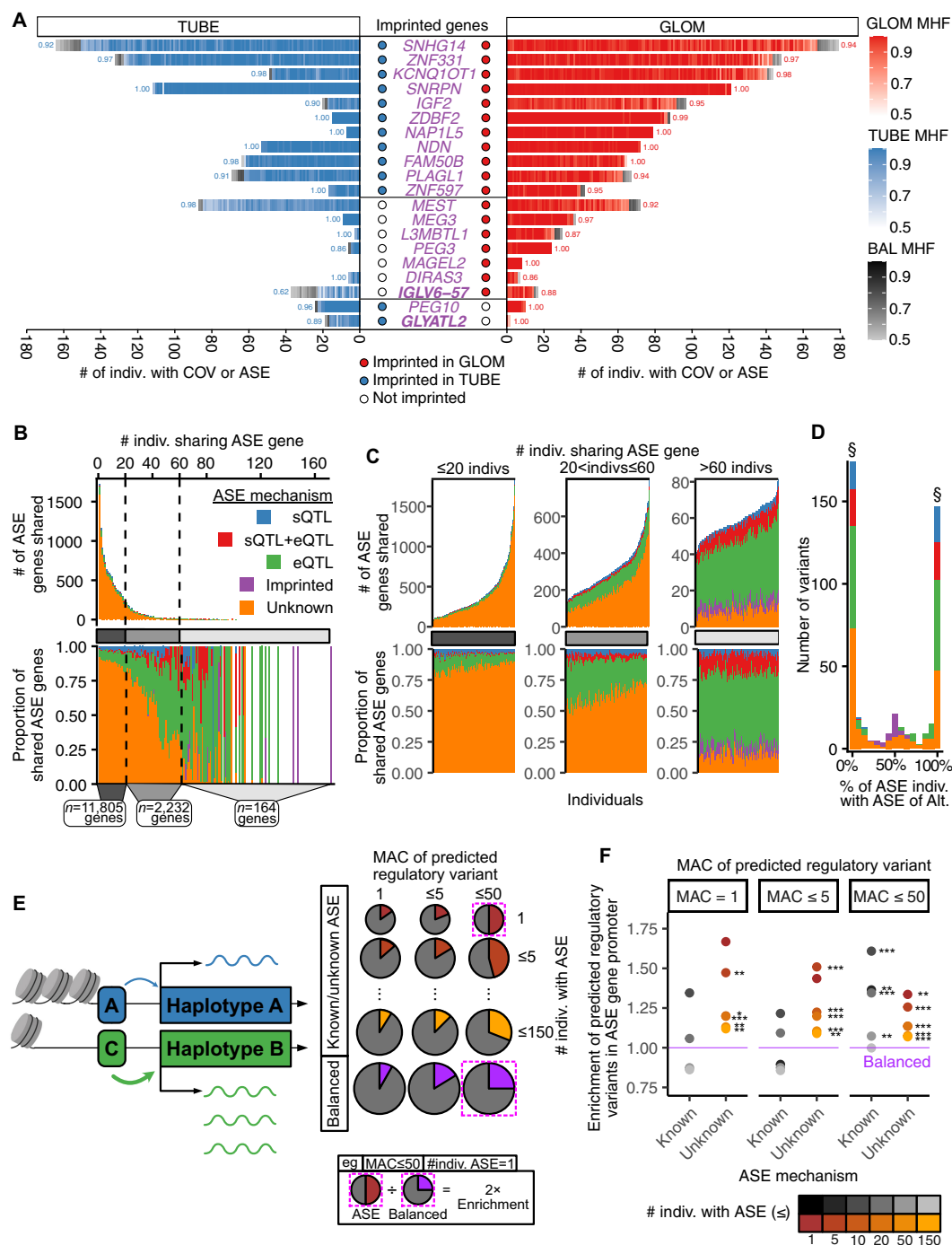


Fig. 2. Molecular mechanisms of ASE in NEPTUNE. (A) Twenty imprinted genes in GLOM and/or TUBE, identified by our imprinting analysis. Similar to Fig. 1F, bars show the number of individuals with BAL or ASE in GLOM or TUBE, with individual stacked bars color graded by MHF. Circles indicate the predicted imprinted status of each gene in the respective compartment, with previously unidentified genes shown in bold. (B) The top panel is a distribution of the ASE gene frequency across the GLOM samples, stratified and colored by their molecular mechanisms, while the bottom panel shows their proportions. Dashed lines split ASE genes into three groups by the number of individuals sharing the ASE gene. The total gene count per group is shown below. (C) Each bar represents one individual, stratified by the three ASE gene groups defined in (B). The top panels show numbers of ASE genes, colored by each molecular mechanism, while the bottom panels show their proportions. (D) Distribution of the alternative allele preference of ASE variants in GLOM, stratified and color coded by their molecular mechanisms. § indicates PAP of Ref. (0%) or Alt. (100%). (E) Schematic of the promoter variant analysis. Partially created in BioRender. Lee, D. (2026) <https://biorender.com/v01e5g9>. (F) Enrichment of predicted promoter regulatory variants of ASE gene-individual pairs, comparing known and unknown ASE mechanisms to BAL gene-individual pairs in GLOM. Each dot represents a group of ASE gene-individual pairs, defined by their mechanism (known/unknown), ASE frequency, and the MAC of the PRVs. The y axis shows enrichment relative to BAL genes (purple line = 1.0). The significance was determined by a one-sided binomial test of the observed count against the expected probability, defined by the proportion of BAL gene-individual pairs with PRVs at the corresponding MAC (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

the known imprinted genes amenable to our ASE-based imprinting analysis still did not demonstrate an imprinting pattern (21 of 38 in GLOM and 30 of 43 in TUBE). Instead, these genes showed biallelic expression (fig. S24), suggesting loss of imprinting in a tissue-specific manner—consistent with a prior study (37).

ASE is primarily sequence-mediated

Alongside imprinting, ASE is often driven by sequence-mediated mechanisms, such as eQTLs and splicing QTLs (sQTLs) (1, 3). Thus, we sought to classify every ASE event in each individual by identifying heterozygous eQTL/sQTLs associated with the corresponding gene (see Materials and Methods) (22, 38). In the absence of imprinting, eQTL, or sQTL effects, we assigned its mechanism as “unknown.” We found two major patterns. First, the causal mechanism for rare ASE genes (e.g., ASE in ≤ 20 GLOM) was overwhelmingly unknown, while sequence-mediated or imprinted mechanism drove most of the common ASE genes (e.g., ASE in > 60 GLOM) (Fig. 2B and fig. S25A). Second, all individuals showed substantially consistent proportions of their ASE explained by each mechanism (Fig. 2C and fig. S25B). Specifically, the vast majority (mean: 86.2%) of the commonly shared ASE genes—which have sufficient power for QTL detection—could be classified into four categories: sQTLs (4.7%), eQTLs (58.4%), sQTLs + eQTLs (14.5%), and imprinting (8.6%) in GLOM. As a representative example, *PLCG2*—one of the most common GLOM ASE genes (Fig. 1F)—demonstrated GLOM-specific, sequence-mediated regulation. In GLOM, *PLCG2* exhibited near-MAE driven by a strong eQTL, rs4243211 (fig. S26). Conversely, in TUBE, *PLCG2* had no significant eQTL and showed BAL expression, underscoring the compartment specificity of certain regulatory mechanisms.

To further validate these findings, we evaluated our ASE data at the single-variant level and assessed the consistency of allelic preference across individuals. We first defined an ASE variant on the basis of (i) variant-level total RNA coverage ≥ 8 , (ii) allelic imbalance $\geq 50\%$, and (iii) $P_{\text{adj}} < 0.05$ (binomial test). We restricted our analysis to common ASE variants (i.e., heterozygous genotype in $\geq 10\%$ of samples, of which $\geq 50\%$ had ASE) to ensure sufficient power for comparison. For each variant, we calculated the proportion of heterozygous individuals preferring the alternative allele (Fig. 2D and fig. S25C). Most ASE variants showed extreme allelic preference; the majority (e.g., 65.3% in GLOM) demonstrated “perfect allelic preference” (PAP), where the same specific allele (reference or alternative) was consistently preferred in every individual (extreme bars in histograms of Fig. 2D and fig. S25C). Conversely, ASE variants within imprinted genes showed random preference, consistent with its parent-of-origin mechanism. We next identified ASE variants that were also significant eQTL/sQTL SNPs for their respective ASE genes, categorizing these intersections as ASE + eQTL (green), ASE + sQTL (blue), or ASE + eQTL + sQTL (red). As expected, variants in these categories were significantly enriched for PAP, accounting for 72.1, 92.7, and 84.8% of variants in GLOM, respectively. Moreover, 61.6% of PAP variants fell into one of these three categories, while none were classified as imprinting (Fig. 2D). Overall, similar results were found in TUBE (fig. S25C). Next, we hypothesized that, if ASE is truly driven by eQTLs, most PAP ASE + eQTL variants should show strong concordance in allele preference between the ASE and eQTL analyses. A total of 96.2% of such variants showed allelic concordance in GLOM. Furthermore, when we compared methylation QTLs (meQTLs) (40) to our common ASE variants, we found that

meQTLs mediated a substantial portion of these variants (42.8% in TUBE and 32.2% in GLOM) (fig. S27, A and B). Notably, meQTLs successfully resolved 29.3% of common ASE variants in TUBE with previously unknown mechanisms. However, these meQTL-driven signals were largely captured by eQTLs with consistent effect sizes and directionality (fig. S27, C and D), indicating a highly overlapping regulatory architecture. Together, these results strongly suggest that ASE is primarily sequence-mediated.

We next hypothesized that genes with unknown mechanism may be driven by rare regulatory promoter variants that eQTL analyses are underpowered to identify. Using the ChromKid model (41), which predicts variant effects on chromatin accessibility in kidney cell types, we analyzed heterozygous promoter variants of all COV genes in each kidney compartment (see Materials and Methods). We first evaluated promoter variants of BAL (non-ASE) genes, generating a baseline likelihood of identifying predicted regulatory variants (PRVs) in a typical promoter. We then compared ASE genes of known/unknown mechanisms with these BAL genes, calculating an enrichment score (Fig. 2E). As expected, genes with an unknown mechanism were significantly enriched for rare [minor allele count (MAC) ≤ 5] PRVs compared to those of known mechanism (Fig. 2F and fig. S25D). This enrichment difference disappeared when including more common (MAC ≤ 50) variants, indicating that rare variants specifically drive unknown ASE. Furthermore, in both mechanistically known and unknown ASEs, rare ASE genes (i.e., observed in few individuals) were particularly enriched for these variants, with decreasing enrichment as we included genes with more common ASE (Fig. 2F and fig. S25D). This suggests that, while ASE is often driven by rare variants, common ASE genes are driven by a broader variety of mechanisms beyond them.

Splicing events and phasing errors affect gene-level ASE consistency

Our prior analysis confirmed that sQTLs are a major mechanism driving ASE events. However, phASER's algorithm represents a potential limitation in identifying sQTL-driven ASE, as it aggregates read counts across all variants of an entire gene haplotype. While this approach works well for eQTLs or imprinting (where the whole transcript is affected uniformly), it masks the complexity of splicing events, in which allelic bias may vary across variants within the same haplotype. This is best illustrated by the example of *OAS1* in Fig. 3A. rs10774671-A (alternative allele) disrupts a canonical splice acceptor site, producing an isoform (*OAS1.2*) that retains the full intron (42). Consequently, this haplotype accumulates reads overlapping intronic variants, such as rs60623134. Conversely, the exonic alleles (such as rs2660) phased with rs10774671-G (reference allele) will have more reads, as the canonical isoform *OAS1.1* is more stable than *OAS1.2*. Therefore, we hypothesize that a unifying characteristic of sQTL-driven ASE is intragenic discordance of allelic imbalance.

To investigate this, we calculated the deviation of variant-level haplotypic fraction from the corresponding gene-level haplotypic fraction using root mean square error (RMSE) (see Materials and Methods). We hypothesized that sGenes (genes with variant-driven alternative splicing, captured by sQTL analysis) would have higher RMSE values because of the allelic discordance within these genes. sGenes had significantly higher RMSE values than genes with other ASE mechanisms (all comparisons, Wilcoxon $P < 0.001$), while imprinted genes had the lowest (Fig. 3B). RMSEs for ASE genes with unknown mechanism fall between sGenes and eGenes, suggesting

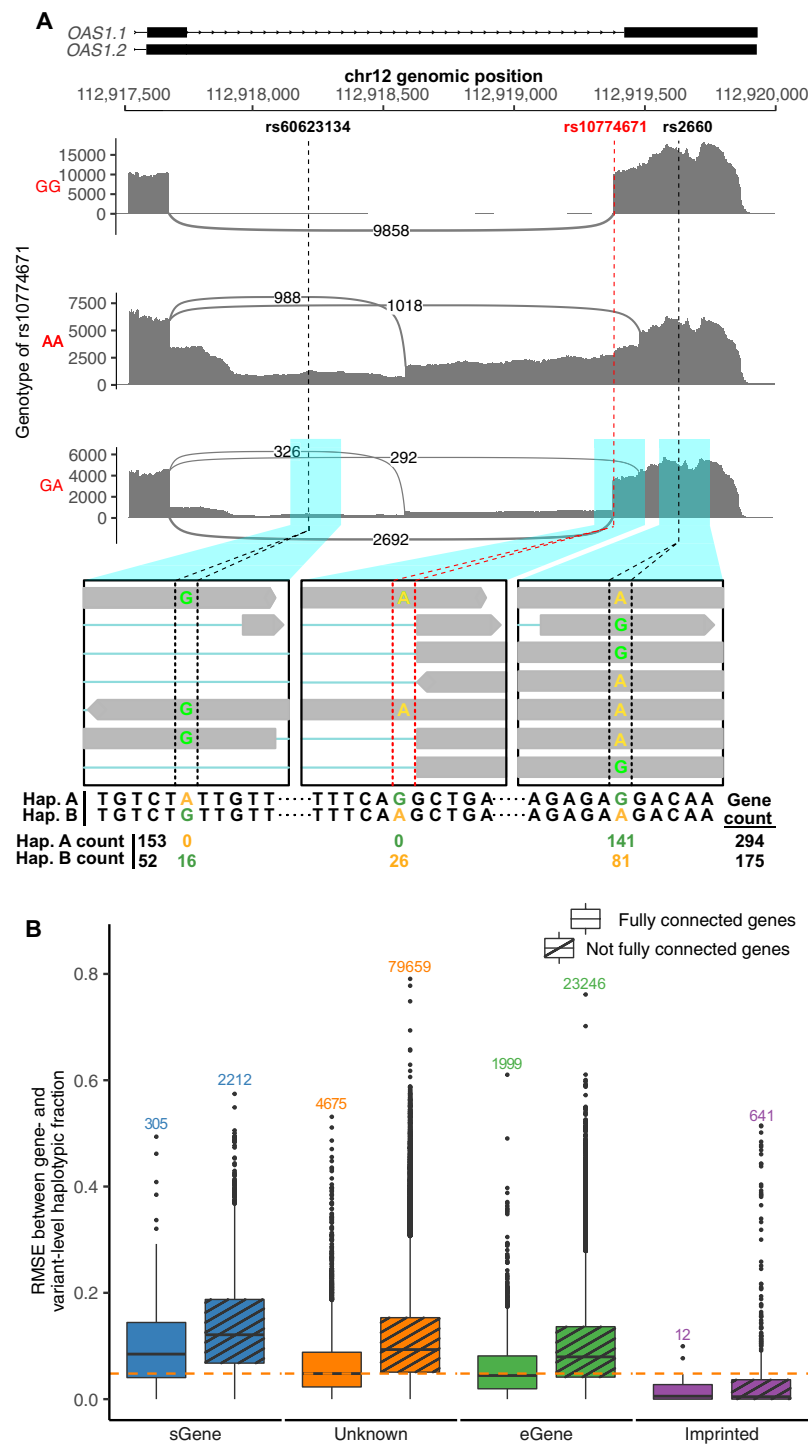


Fig. 3. Alternative splicing and phasing errors confound gene-level ASE analysis. (A) Sashimi plots demonstrate alternative splicing of the 3'-end of *OAS1* based on the sQTL rs10774671 genotype (red), as exemplified in three NEPTUNE individuals: one homozygous reference (GG), one homozygous alternative (AA), and one heterozygous (GA). Vertical dashed lines denote genomic locations of rs10774671 and two other SNPs (rs60623134 and rs2660), whose RNA-seq read alignments in the heterozygous individual are shown below the respective Sashimi plot. Variant-, haplotype-, and gene-level ASE counts are shown below the read alignments. (B) Quantification of discordance in variant- and gene-level haplotypic fractions. The gene-level fraction of haplotype A (a randomly chosen haplotype by phASER) was used as the expected value, and the gene's different variant-level haplotype A fractions were used as the observed values. Genes and respective variants were stratified by the ASE mechanism. Gene-individual pairs were further separated into "fully connected genes," where all the gene's variants were phased by consecutive overlapping RNA reads by phASER, and those that were not, termed "not fully connected genes." Numbers above box plots indicate the number of gene-individual pairs in each category. sGenes whose variants separated by the splicing event were included only. The orange line indicates the median RMSE of the unknown mechanism, fully connected category. Pairwise Wilcoxon tests within each connected group showed that all categories had significant differences in mean RMSE ($P < 0.001$).

that some ASE events of unknown mechanism are driven by sQTLs not found in our dataset. This was largely true for TUBE samples as well (fig. S25E).

Notably, we observed systematically higher RMSE values in genes whose haplotypes were not fully connected by phASER's RNA read-backed phasing (Fig. 3B). 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 expected to have strictly monoallelic expression, the observation of variants with MAE mapping to opposing haplotypes within the same imprinted gene confirms a phasing error. We used this principle to estimate the prevalence of haplotypic flips in our dataset. Using all our imprinted genes per individual (i.e., imprinted gene-individual pairs), we calculated the proportion in which at least two variants with read depth ≥ 8 demonstrated conflicting near-MAE ($\geq 90\%$ allelic bias). In GLOM, we tested 666 imprinted gene-individual pairs with a mean of 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 substantially lower than the previously reported gene-level phasing error rates ($\sim 20\%$) (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 found are transcriptionally coordinated within specific cell types. Using the Kidney Precision Medicine Project (KPMP) single-cell RNA-seq dataset (24), 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 pseudobulk 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" (see Materials and Methods).

We initially analyzed all ASE genes observed in ≥ 1 individual (Fig. 4A). In GLOM, 60.1% of ASE genes were ubiquitous; 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. 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 PODs and parietal epithelial cells (PECs) in GLOM and for proximal tubule cells (PTs) in TUBE (Fig. 4, A and B). Compared to all ASE genes, POD-predominant ASE genes increased from 8.4 to 17.2%, and PEC-predominant genes increased from 13.2 to 21.4%, while ubiquitous genes 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. 1B). We therefore calculated the genome-wide ASE proportion per sample (i.e., a sample's number of ASE genes divided by their number of COV genes). The median ASE proportion in GLOM was significantly higher than TUBE (11.3% versus 7.0%, $P < 2.2 \times 10^{-16}$), with 177 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 additionally analyzed 31 microdissected "control" kidneys with preserved kidney function and without proteinuric disease (see Materials and Methods). As these control samples had only bulk RNA-seq (without paired WGS), we developed an RNA-only ASE pipeline for direct comparison (Materials and Methods and fig. S28A). We note that, in RNA-only ASE analysis, true (near-)MAE is missed as these sites are assumed to be homozygous, resulting in fewer identified ASE events and no MAE compared to the WGS + RNA pipeline (table S4 and fig. S28B). Nonetheless, results from the RNA-only analysis of NEPTUNE were highly consistent with the WGS + RNA pipeline (fig. S28C), demonstrating significantly higher ASE proportion in GLOM versus TUBE across the cohort (median 7.67% versus 5.27%, $P < 2.2 \times 10^{-16}$) (fig. S29A). In contrast, the control cohort showed no significant difference (median ASE proportion 6.48% in GLOM versus 6.96% in TUBE, $P = 0.68$) (fig. S29, B and C). Because of 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 proportion ratio (GTAR). In the RNA-only analysis, the median GTAR in NEPTUNE was significantly higher than the controls (1.40 versus 0.97; $P = 2.2 \times 10^{-6}$) (Fig. 5B).

In light of this difference, we hypothesized that clinical outcomes in proteinuric kidney disease could vary as a function of GTAR. To test this, we dichotomized 194 NEPTUNE participants into a low-GTAR group (GTAR of 0.44 to 1.578, similar to most controls) and a high-GTAR group (1.579 to 6.16), on the basis of WGS + RNA analysis, after excluding 12 participants with missing outcome data (Fig. 5A). Kaplan-Meier survival analysis revealed that the high-GTAR group had slower kidney disease progression [composite end point of ESKD and/or $\geq 40\%$ decline in estimated glomerular filtration rate (eGFR)] compared to the low-GTAR group ($P = 1.57 \times 10^{-3}$) (Fig. 5C). Cox proportional hazards analysis confirmed this finding: Patients with high GTAR were 65% less likely to reach the end point after adjusting for multiple kidney disease-related covariates [hazard ratio (HR) = 0.35, $P = 9.63 \times 10^{-4}$] (Fig. 5D). Notably, there was no difference in survival when stratifying solely on the basis of either GLOM or TUBE ASE proportion (fig. S30).

We next investigated GTAR's correlation with additional clinical outcomes in NEPTUNE (see Materials and Methods). The high-GTAR group achieved complete remission of proteinuria sooner (Cox-adjusted HR = 1.58, $P = 0.023$) (fig. S31A), demonstrating improved renal recovery compared to the low-GTAR group. Similarly, high-GTAR individuals showed a trend toward later and less relapse of active nephrotic syndrome, although the difference was not statistically significant, likely due to limited sample size (fig. S31B). In contrast, the low-GTAR group showed nonsignificantly higher kidney scarring in the forms of global glomerulosclerosis ($P = 0.083$)

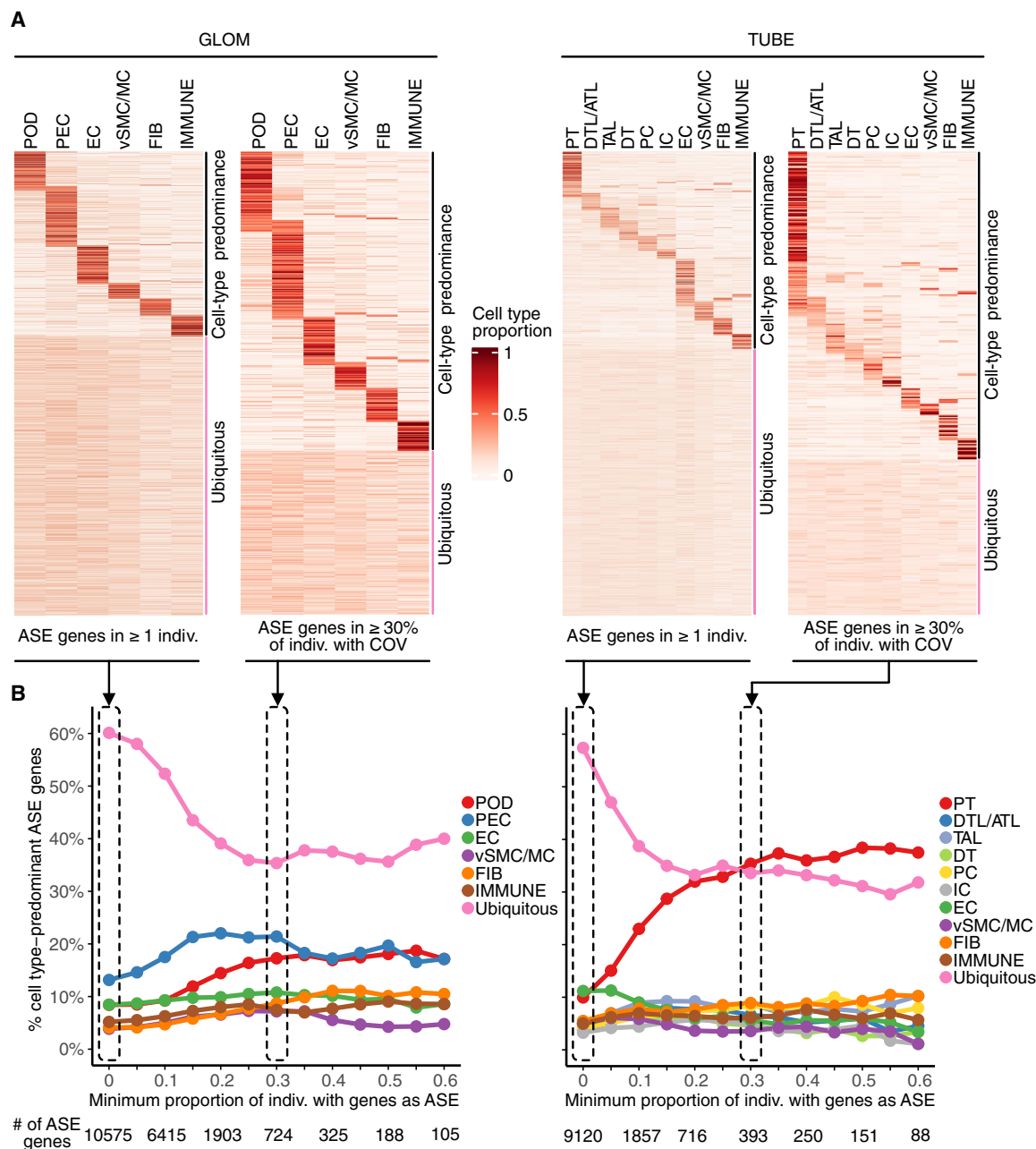


Fig. 4. Cell-type specificity of ASE genes in NEPTUNE kidney compartments. (A) Heatmaps of cell-type proportions for all ASE genes (ASE observed in ≥ 1 individual) and common ASE genes (ASE observed in $\geq 30\%$ individuals with COV) in GLOM (left) and TUBE (right). Rows represent individual ASE genes, columns represent cell types, and color scale indicates cell-type proportion. The right panel of each heatmap indicates whether ASE genes are “cell-type predominance” or “ubiquitous.” All analyses were restricted to genes with COV in $\geq 10\%$ individuals. (B) 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. The dotted black box corresponds to the heatmap in (A). Unique GLOM cell types are PODs and PECs. Unique TUBE cell types are PTs, intercalated cells (ICs), distal convoluted tubule cells (DTs), thick ascending limb cells (TALs), descending/ascending thin limb cells (DTLs/ATLs), and principal cells (PCs). Common cell types include endothelial cells (ECs), vascular smooth muscle/mesangial cells (vSMCs/MCs), fibroblasts (FIBs), and immune cells (IMMUNE).

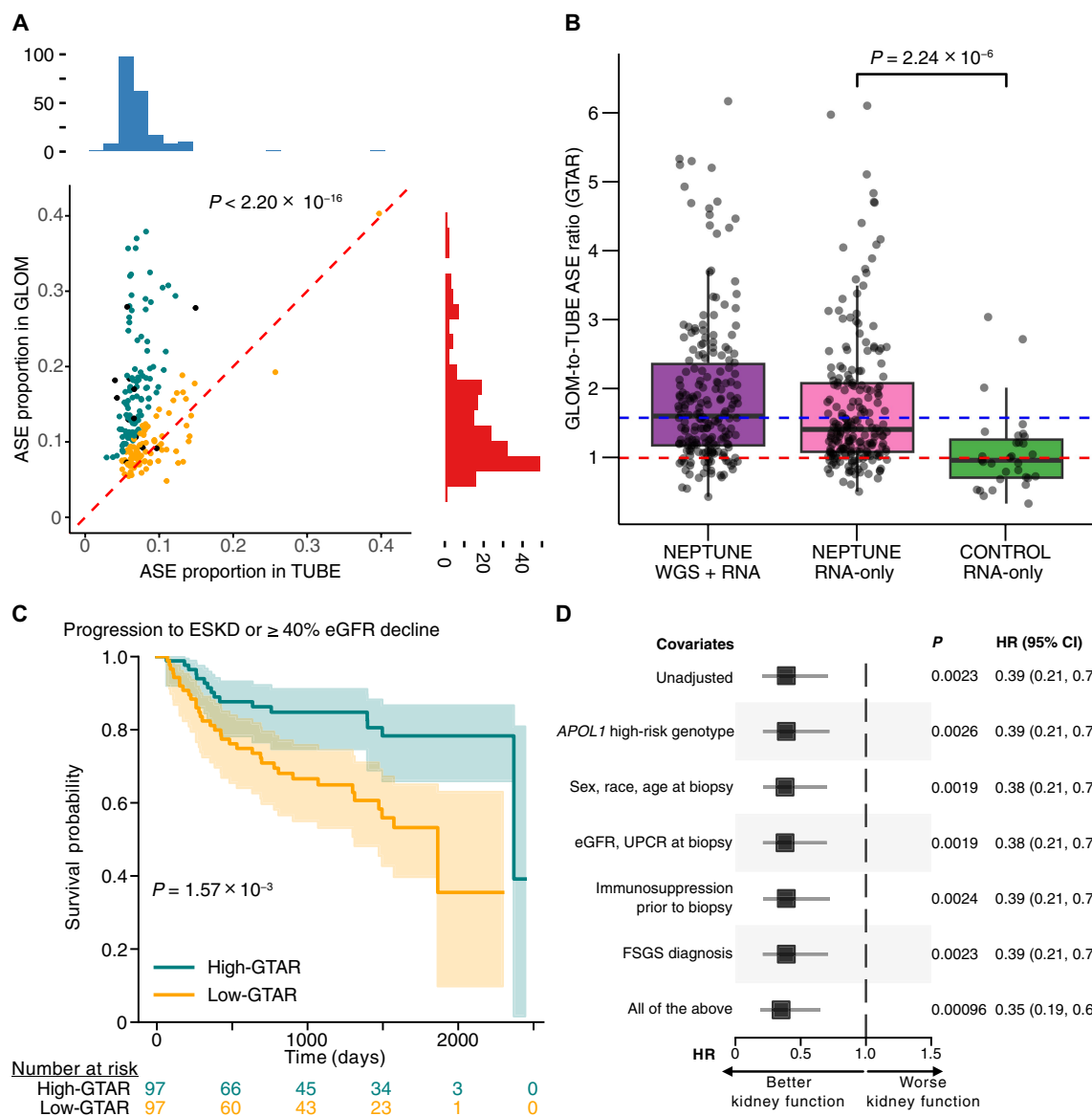


Fig. 5. Genome-wide ASE proportion in GLOM/TUBE in NEPTUNE versus controls and correlation of GTAR with clinical outcomes in NEPTUNE participants.

(A) Scatterplot of genome-wide ASE proportion (ASE/COV per sample) in paired GLOM and TUBE samples. Each dot is a NEPTUNE individual with both GLOM and TUBE ASE data ($n = 206$), plotted by ASE proportion in TUBE (x axis) and in GLOM (y axis). After dichotomizing into two groups for survival analysis, high-GTAR samples ($n = 97$) are indicated in teal and low-GTAR samples ($n = 97$) in orange. Samples without end-point data are in black ($n = 12$). Marginal histograms illustrate the distribution of genome-wide ASE proportion in each compartment. The Wilcoxon signed-rank test P is shown to compare GLOM and TUBE proportions. The red dashed line indicates equal GLOM and TUBE ASE proportions. (B) GTAR per individual, stratified by cohort and ASE analysis pipeline. Each dot is an individual with paired GLOM and TUBE data (NEPTUNE, $n = 206$; control, $n = 31$). The blue dashed line represents the median GTAR for NEPTUNE (WGS + RNA pipeline), and the red dashed line indicates GTAR = 1. (C) Kaplan-Meier survival analysis of NEPTUNE individuals ($n = 194$) comparing the high- and low-GTAR groups. The end point is the composite of ESKD or $\geq 40\%$ eGFR decline. (D) Forest plot of Cox proportional hazards model, with HR in reference to the high-GTAR group and adjusted for multiple kidney disease-related covariates. FSGS, focal segmental glomerulosclerosis; CI, confidence interval.

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 sex or any other demographic, clinical, or histologic parameters. Together, high GTAR was associated with better kidney outcomes.

We also examined whether these clinical and transcriptomic differences were driven by variations in cellular composition, such as a more preserved POD population in the high-GTAR group. Computational

cell-type deconvolution of the GLOM RNA-seq data using MuSiC (43), leveraging a cohort-matched (NEPTUNE) single-nucleus RNA-seq reference (44) (see Materials and Methods), revealed no significant differences in estimated cell-type proportions, including PODs, between the high- and low-GTAR groups (fig. S32A). Furthermore, even after stratifying by disease subtypes [FSGS (focal segmental glomerulosclerosis), MCD (minimal change disease), and MN (membranous nephropathy)], pairwise comparisons revealed no significant

differences in cellular composition between the two groups within any disease subtype (fig. S32B). These results indicate that the elevated global ASE in GLOM and clinical resilience observed in the high-GTAR group are not driven by underlying variations in cellular composition.

Differential gene expression analysis reveals down-regulation of fibrosis genes and up-regulation of ribosome and ATP pathways in the high-GTAR group

We next explored whether differences in gene expression patterns in each GTAR 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 and DESeq2 (see Materials and Methods) (45). GLOM DGE analysis identified 181 significantly differentially expressed genes (DEGs), with $P_{\text{adj}} < 0.05$ and absolute $\log_2$ fold change ($\log_2\text{FC}$) > 0.5 . Of these, 115 genes were up-regulated and 66 down-regulated in the high-GTAR group (Fig. 6A and table S9). In contrast, TUBE analysis demonstrated no significant DEGs. Thus, further analysis was focused on the GLOM DGE results.

Characterization of GLOM DEGs revealed notable biological patterns. First, a high proportion of DEGs (62 of 181; 34.3%) were long noncoding RNAs. The vast majority (58 of 62; 93.5%) were up-regulated, including 20 antisense to protein-coding genes. Five up-regulated long noncoding RNAs were antisense to—thus expected to decrease the expression of—genes known to promote kidney disease (e.g., *AQP5*, *LYZ*, *P2RX7*, *CERS6*, and *NFATC3*—with the latter four specifically causing renal fibrosis) (46–56). Second, the high-GTAR group had up-regulation of several kidney-protective genes and down-regulation of many genes that promote kidney injury—a notable finding given the better kidney clinical outcomes in the high-GTAR group. Specifically, we found up-regulation of *FCGR2B* (Fc γ receptor IIb), an immune inhibitory receptor that suppresses autoimmunity and autoantibody-mediated glomerulonephritis (57–60), and of *NGF* (nerve growth factor), secreted by PODs and a protective factor against glomerular kidney injury (61–63). Conversely, down-regulated DEGs were strongly enriched for genes whose increased expression contributes to kidney dysfunction. Of the 66 down-regulated DEGs, 10 (15.2%) promote kidney fibrosis (*FN1*, *CCL19*, *CCL21*, *ADAMTS4*, *ADAMTS16*, *CEMIP*, *ITGA11*, *NEU4*, and *HK2*) (64–77) or cardiac/pulmonary fibrosis (*ASPN*) (78, 79). Several other down-regulated genes were associated with glomerular disease (e.g., *MRC2*, *ACER2*, *IL1RAP*, and *NEFL*) (80–84). Last, 7 of 115 (6%) up-regulated DEGs were ribosome-related, with no ribosomal genes among down-regulated 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 the 181 GLOM DEGs with overrepresentation analysis (ORA). For up-regulated genes, “ribosome biogenesis in eukaryotes” [Kyoto Encyclopedia of Genes and Genomes (KEGG)] was the single significant concept ($P_{\text{adj}} < 0.05$). For down-regulated genes, we identified 57 significant terms (many redundant), including “dendritic cell antigen processing and presentation,” “response to corticosterone,” “cell adhesion mediated by integrin,” “ligand-gated sodium channel activity,” “collagen binding,” and “immunoglobulin complex” (Fig. 6B and table S11).

To comprehensively identify molecular pathways, we also performed gene set enrichment analysis (GSEA) on the full GLOM transcriptome using clusterProfiler (85), ranking all genes

by signed($\log_2\text{FC}$) $\times -\log_{10}(P)$ (table S11). Consistent with the ORA results, the top two up-regulated terms (disregarding highly redundant terms) related to ribosome and oxidative phosphorylation in all five functional databases (Fig. 6, C and D). The most strongly down-regulated pathways involved cell adhesion and extracellular matrix (ECM) 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.

We next identified genes exhibiting “differential ASE” by using 2×2 contingency tables to compare the number of individuals with ASE versus BAL expression between the high- and low-GTAR groups. Among 10,872 genes analyzed in GLOM, 166 genes showed significant differential ASE ($P_{\text{adj}} < 0.05$; fig. S33 and table S12). All 166 genes exhibited a higher frequency of ASE in the high-GTAR group, with no genes showing the reverse trend. However, downstream enrichment analyses using ORA and GSEA did not yield any significantly enriched biological pathways or Gene Ontology (GO) terms.

Last, we evaluated whether ASE genes are enriched in DEGs between the high- and low-GTAR groups. From all genes used in our original DEG analysis ($n = 27,006$), we restricted our analysis to COV genes in $\geq 10\%$ of individuals in both GTAR groups, resulting in 11,051 genes. We then calculated the difference in the proportion of individuals exhibiting ASE for each of these genes between the two GTAR groups (fig. S34A). As expected, most genes showed a greater proportion of ASE in the high-GTAR group, regardless of their differential expression status (fig. S34B). However, ASE genes were more strongly enriched among down-regulated DEGs than up-regulated DEGs (fig. S34C). These results suggest that elevated ASE in the high-GTAR group reflects contributions from both transcriptional amplification and repression, with repression being the stronger mechanism.

A non-ASE expression signature captures GTAR-associated prognostic value

Building on the DEG results, we explored their translational potential by deriving a simplified, non-ASE prognostic score based purely on standard expression levels, bypassing the complex haplotypic quantification. Using a rank-based single-sample enrichment method, sing-score (86), to contrast the expression of these protective [ribosome/adenosine 5'-triphosphate (ATP)] and injury-related (immune/matrix) pathways, we generated a composite expression signature. This approach was able to discriminate high-GTAR from low-GTAR groups [AUC (area under the curve) = 0.77] (fig. S35, A and B) and demonstrated a trend toward better kidney survival ($P = 0.08$) (fig. S35C). Although its prognostic capacity was less pronounced and failed to fully replicate the performance of the original GTAR metric, this expression-only signature represents a highly promising avenue for future clinical translation.

DISCUSSION

In this study, we present a comprehensive, genome-wide characterization of ASE in the human kidney, leveraging a unique dataset of paired WGS and microdissected bulk RNA-seq from the NEPTUNE cohort (30). 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, proteinuric

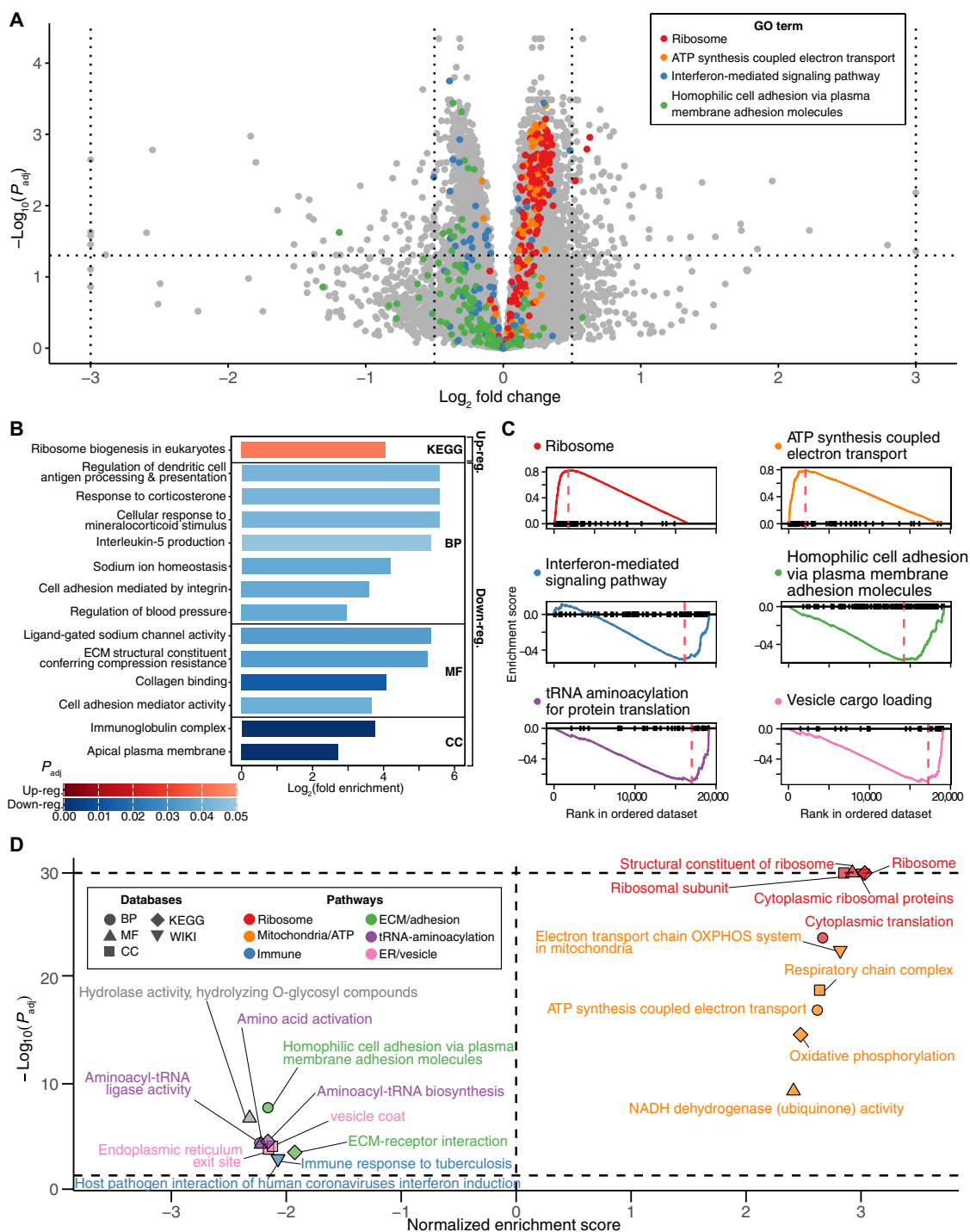


Fig. 6. Differential glomerular gene expression between high- and low-GTAR groups and GSEA. (A) Volcano plot of DEGs with the low-GTAR group as a reference. The horizontal dashed line indicates the P_{adj} threshold (0.05), and the inner and outer vertical dashed lines indicate $\log_2 FC$ cutoffs of ± 0.5 and ± 3 , respectively. Genes exceeding ± 3 are aligned with the outer vertical lines. Significantly up-regulated GO terms related to ribosome and mitochondrial/ATP synthesis (red and orange) and down-regulated immune-related and membrane-associated pathways (blue and green) are highlighted. The corresponding GO terms are shown in (C) with matching colors. **(B)** ORA of significantly up-regulated ($\log_2 FC > 0.5$, $P_{adj} < 0.05$) and down-regulated ($\log_2 FC < -0.5$, $P_{adj} < 0.05$) genes across multiple databases. Bar color scales reflect P_{adj} from ORA (up-regulated: red; down-regulated: blue). **(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 up- and down-regulated GO terms ranked by normalized enrichment score across multiple databases. Terms with $-\log_{10}(P_{adj}) > 30$ are truncated and aligned at the top horizontal dashed line. The bottom dashed line indicates the threshold for statistical significance ($P_{adj} = 0.05$). The vertical dashed line shows where there is no enrichment. Colors correspond to the six categories in (C), and the gray color denotes GO terms not included in these categories. Shapes indicate the database source.

kidneys exhibit a significantly higher proportion of genome-wide ASE in GLOM compared to TUBE. We quantified this discrepancy using a new metric, the 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 that this imbalance is not a marker of dysregulation or injury. Instead, our DGE 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 up-regulation of pathways related to ribosome biogenesis and oxidative phosphorylation (ATP synthesis) as well as down-regulation of injury-related (immune/matrix) pathways in GLOM. This suggests that, rather than ASE being a causal mechanism driving a protective transcriptional state, the observed increase in ASE is a by-product of adaptive 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 up-regulation of known kidney-protective genes and concordant down-regulation of a suite of profibrotic 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). We provide an explanation for individual-specific ASE, previously speculated as stochastic or random effects. By integrating deep learning models (41), we revealed that genes with an “unknown” mechanism of ASE were significantly enriched for rare, heterozygous PRVs 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 raises 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?

An intriguing secondary question is whether ASE signals are also associated with genetic risk factors identified by large-scale kidney function GWAS (40, 41). Upon exploring this overlap, we observed that shared ASE variants were significantly enriched for eGFR GWAS loci, particularly within TUBE (fig. S36). Although this remains a secondary observation, this finding provides a link between molecular phenotypes and broad clinical traits, suggesting that a subset of GWAS risk variants may exert their effects via context-specific allelic imbalances under disease conditions.

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 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 accurate ASE quantification.

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 (32) 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 to 3%) in known imprinted genes. Second, our primary analysis was performed on bulk RNA-seq from microdissected glomerular and tubulointerstitial compartments. While this provides a crucial compartmental resolution, it lacks the granularity of single-cell analysis. We began to address this by integrating KPMP single-cell data (24), which revealed that frequently shared ASE genes tend to be cell type-specific. Furthermore, while cellular heterogeneity or immune infiltration could theoretically generate apparent allelic imbalances, our cell-type deconvolution analysis revealed no significant differences in cell-type proportions across samples, making cellular heterogeneity an unlikely explanation for the observed signals. Nonetheless, alternative biological phenomena such as cell-state diversity, localized clonal expansions, or transcriptional bursting cannot be fully ruled out without true single-cell or spatial ASE quantification. Third, our comparison to the control cohort required using an RNA-only pipeline resulting from data availability, although the stark difference in GTAR was robust. Fourth, the GTAR metric itself requires further validation in independent cohorts, and its sensitivity to technical parameters and sample processing must be fully evaluated; however, because of the scarcity of relevant, independent datasets, we were unable to perform this validation. Last, 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 a comprehensive landscape of ASE 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.

MATERIALS AND METHODS

Nephrotic Syndrome Study Network (NEPTUNE)

NEPTUNE (30) is a multicenter, longitudinal cohort of participants with proteinuric kidney disease, including nephrotic syndrome. Participants are enrolled at the time of their first clinically indicated kidney biopsy, during which an extra core is collected for research. All NEPTUNE participants provided written informed consent before enrollment (parental consent and child assent for minors). This research biopsy core undergoes histopathologic characterization (87–89) and is also microdissected into glomerular and tubulointerstitial

compartments for transcriptomic analysis. WGS data derived from blood and RNA-seq data from GLOM and TUBE samples were obtained and analyzed. Race and ethnicity were self-reported or reported by parents of children. Reporting race and ethnicity in the NEPTUNE study was mandated by the US National Institutes of Health (NIH), consistent with the Inclusion of Women, Minorities, and Children policy. For analysis, the “Other” self-reported racial group comprises individuals with Unknown and Not Available (NA) categories.

Processing of NEPTUNE WGS

CRAM files were created from raw WGS data from the NEPTUNE cohort, provided as FASTQ files, by following the Functional Equivalence pipeline (90). Subsequent processing was performed using the WARP WholeGenomeGermlineSingleSample WDL pipeline (version 2.3.3; <https://github.com/broadinstitute/warp/>), which was run on Cromwell version 52. Individual-level gVCF files were produced using HaplotypeCaller from the GATK version 4.2.0 suite, with default settings. Joint genotyping was then performed across the cohort using the JointGenotyping WDL pipeline from the Broad Institute with the default parameter set. Following joint calling, variant sites tagged as “ExcessHet” were filtered out. Last, VQSR (Variant Quality Score Recalibration) filters were applied using the thresholds of VQSR Tranches SNP 99.7 and INDEL 99.0 to create the final VCF files.

Phasing of NEPTUNE WGS

Here, we used a hybrid approach combining read-backed and population-based phasing to maximize phasing accuracy, which is a critical preliminary step for the downstream calculation of haplotypic counts by phASER (6). NEPTUNE WGS VCF files were initially processed using the WhatsHap tool, which performs read-backed phasing. This approach uses sequencing reads covering two or more heterozygous variants to assemble haplotype blocks, called “phase sets,” which are recorded in the VCF file with a unique PS tag attached to the genotype (31). QC was then performed to filter out low-quality variants, which could be incorrectly imputed during population phasing. We kept variants with genotype missingness rate <5% and Hardy-Weinberg equilibrium exact test $P > 1 \times 10^{-10}$. SHAPEIT4 (32) was subsequently used for population-based phasing of the NEPTUNE WGS data in a two-step approach. We initially ran SHAPEIT4 with the option `--scaffold` to generate haplotype scaffolds for NEPTUNE original VCFs using the UKB haplotype panel as a reference. To create this reference panel, the BGEN format per-chromosome genotype files for ~500,000 UKB participants (33) were first converted to VCF format using the bgenix tool (<https://enkre.net/cgi-bin/code/bgen/doc/trunk/doc/wiki/bgenix.md>), with VCF header correction using bcftools. UKB VCF files were then lifted over from hg19 to hg38 using the GATK LiftoverVcf tool with the option `--RECOVER_SWAPPED_REF_ALT`. These scaffolds retain variants that are present in the original dataset and in the reference panel, representing a sparser set of variants that incorporate prior phase information of some heterozygous genotypes. We then ran SHAPEIT4 again using the NEPTUNE VCFs with read-backed phase sets as input and incorporating the previously generated individual scaffolds. With this approach, scaffold variants remain unchanged, and all additional variants from each NEPTUNE VCF are phased onto their respective scaffold, resulting in the final population-phased VCF.

Processing of NEPTUNE bulk RNA-seq

We analyzed paired-end RNA-seq data from 222 microdissected GLOM samples and 206 paired TUBE samples from the NEPTUNE cohort. QC of individual sequencing reads was performed using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), assessing parameters such as adapter content, GC distribution, and per-base sequence quality. To detect potential contamination and off-target RNA, we applied FastQ Screen (91), which maps a random subset of reads from each sample to 16 different genomes and known artifacts. RNA-seq reads were subsequently aligned to the hg38 reference genome using STAR version 2.7.11b (35). The WASP algorithm (9) was applied during this process using the `--waspOutputMode` option, which flags reads exhibiting reference allele mapping bias with a vW tag in the SAM output. Following alignment, duplicate reads were marked using samtools markdup version 1.19.2. This tool designates which of the duplicate reads to keep according to base quality scores rather than read mapping scores, as the latter may also contribute to reference bias (92). Gene quantification for subsequent analyses was then performed using RSEM version 1.3.1 (93), a robust tool for estimating transcript abundances. As inputs, RSEM was provided with the hg38 reference genome, GENCODE version 45 primary assembly for transcript and gene annotations (94), and BAM files generated using STAR's `--quantMode TranscriptomeSAM` option.

Sample identity and contamination verification of NEPTUNE WGS and RNA-seq

We used verifyBamID (34) to assess sample integrity, specifically checking for cross-sample contamination and potential sample swaps. This tool evaluates the concordance between sequencing reads and known genotypes to estimate contamination arising from sample mixtures. For the WGS data, verifyBamID was provided with a VCF file containing exonic variants filtered using GENCODE version 45 annotations (94). For the RNA-seq data, the STAR-aligned, duplicate-marked BAM files were used as input. This analysis was performed on all 240 GLOM and 218 TUBE samples. The results identified strong evidence of sample mismatches between WGS and RNA-seq data for 13 GLOM and 7 TUBE samples, indicating probable sample swaps. Because of the uncertainty of whether the mismatch originated from the WGS or RNA-seq identifier, we conservatively excluded these samples from both the NEPTUNE GLOM and TUBE cohorts for all downstream analyses.

Generation of individualized phASER blacklists for ASE analysis

High-confidence heterozygous (het) genotyping calls are necessary for valid ASE analysis, as erroneous heterozygous calls can lead to false-positive MAE. We therefore implemented two new QC strategies to improve ASE calling by filtering suspicious variants beyond the default phASER blacklists (6).

First, we performed individual-level WGS QC to remove low-quality heterozygous variants. For all variants called heterozygous after population phasing, we calculated the fraction of sequencing reads supporting the alternative allele in both WGS and RNA-seq, denoted as DNA and RNA alt fraction. To calculate the RNA alt fraction, we ran phASER using only the default variant blacklists (6). Variants were added to an individual-specific blacklist as follows. First, we identified variants with discordant genotype between pre- and post-population phasing. Second, we identified variants with

extreme sequencing depths (WGS sequencing depth <10 or ≥ 80). Third, we identified variants with extreme allelic imbalance at the DNA level (DNA alt fraction <0.2 or >0.8). On average, this individual-level filter blacklisted 178 variants per GLOM sample, of which 5.52% had a mismatched genotype, 88.1% were alt fraction fails, and 44.7% were DNA depth fails. Similarly, an average of 184 per TUBE sample were blacklisted, of which 5.02% had mismatched genotypes, 86.2% were alt fraction fails, and 56.7% were DNA depth fails. Note that some variants failed to pass multiple criteria. Furthermore, individuals with >1000 such QC-failed variants ($n = 18$) were removed from the analysis entirely.

Second, we implemented a cohort-level filter to identify and blacklist variants susceptible to systematic mapping errors, which cause RNA/DNA discordance. For each variant, individuals were grouped by their WGS-derived genotype: homozygous reference (HOM_REF), homozygous alternative (HOM_ALT), or heterozygous (HET). We then used samtools mpileup on the RNA-seq BAM files (before WASP correction) to calculate the average RNA alt fraction for all individuals within each genotype group. Variants were added to a blacklist if the average RNA alt fraction was substantially discordant with the WGS genotype, that is, if the HOM_REF group had an average RNA alt fraction >0.1 or the HOM_ALT group had an average RNA alt fraction <0.9 . This filter removed an additional average of 363 variants per GLOM sample and 361 per TUBE sample. After applying all blacklists (default and custom), the average individual had 62,395 high-confidence unique variants available for ASE analysis in GLOM and 65,614 in TUBE.

TOGA: Transcript reassignment for overlapping genes in ASE

TOGA assigns reads overlapping multiple genes to their most likely gene of origin by integrating total RNA-seq expression data and gene exon/intron annotations. The detailed pipeline is described in fig. S8. Specifically, from the set of overlapping COV genes within an individual, we first excluded nonexpressed genes [transcripts per million (TPM) = 0] on the basis of RSEM (93). We then selected only genes that shared variants within overlapping regions and grouped the genes on the basis of the union of overlapping variants. That is, a gene was included in a group if at least one of its variants overlapped with variants from other genes within the same group. Genes sharing no variants were considered “resolved.”

We first define V as a set of variants, $\{v_1, v_2, \dots, v_M\}$, and G as a set of genes, $\{g_1, g_2, \dots, g_N\}$, where M is the number of unions of overlapping variants and N is the number of unions of overlapping genes. We assumed that an allelic count of an overlapping variant is distributed proportionally on the basis of the genes' expression level and the variants' gene annotation status (e.g., exonic or intronic). We assumed that exonic variants contribute substantially more to the count than intronic ones. Thus, raw weights (w) were assigned as follows: 0.99 for exonic variants (as most RNA reads originate from exons), 0.01 for intronic variants, and 0 for variant-gene pairs that do not exist.

The final weight, $W(v_i, g_j)$, representing the probability that variant v_i originates from gene g_j , was normalized by the total weight of all genes covering that variant, as formulated in Eq. 1

$$W(v_i, g_j) = \frac{w(v_i, g_j)e(g_j)}{\sum_{k=1}^N w(v_i, g_k)e(g_k)} \quad (1)$$

where v_i is the i -th overlapping variant, g_j is the j -th overlapping gene, $e(g_j)$ is g_j 's TPM, and $w(v_i, g_j)$ is the raw weight function defined above. Next, we summed the adjusted allelic counts of all overlapping variants on the basis of W , followed by normalization across all overlapping genes. We defined this metric as CPOV (contribution proportion of overlapping variants) for g_j in Eq. 2, which represents the proportion of total overlapping reads contributed to a given gene

$$\text{CPOV}_j = \frac{\sum_{i=1}^M C(v_i)W(v_i, g_j)}{\sum_{j=1}^N \sum_{i=1}^M C(v_i)W(v_i, g_j)} \quad (2)$$

where $C(v_i)$ is an original allelic count of v_i . For example, if there were a total of 100 RNA reads in a genomic interval where two genes, denoted X and Y, overlapped and our TPM- and exon/intron-based weighting algorithm estimated that 80 reads came from gene X and 20 from gene Y, then the CPOV would be 0.8 and 0.2, respectively.

To further estimate the likelihood of incorrect read assignment, we defined another metric, IRPG (incorrectly assigned read proportion in gene)

$$\text{IRPG}_j = \frac{(1 - \text{CPOV}_j)C(v_o)}{C(v_o) + C(v_{no})} \quad (3)$$

where $C(v_o)$ and $C(v_{no})$ are the read counts for overlapping and non-overlapping variants in a gene, respectively. Following the previous example, if gene X had 60 non-overlapping reads, then its IRPG would be $(1 - 0.8) \times 100/(100 + 60) = 20/160 = 0.125$, indicating that an estimated 12.5% of reads assigned to gene X likely originated from the other overlapping genes. As detailed in fig. S8, we implemented a stepwise classification algorithm on the basis of these metrics. First, genes with $\text{IRPG} < 0.1$ were considered unaffected by read misassignment and classified as “resolved.” For genes with $\text{IRPG} \geq 0.1$, we computed the adjusted haplotypic total count (Cov_{adj}) by redistributing reads from overlapping variants based on CPOV, as shown in Eq. 4

$$\text{Cov}_{\text{adj}} = \text{CPOV} \times C(v_o) + C(v_{no}) \quad (4)$$

Subsequently, genes with $\text{Cov}_{\text{adj}} < 20$ were excluded. Among the remaining genes, those with $\text{CPOV} < 0.5$ were considered to have a low contribution of overlapping reads; thus, we opted to use only non-overlapping reads for their ASE analysis, contingent on non-overlapping coverage ≥ 20 and classified as “resolved by low CPOV.” For the genes with $\text{CPOV} \geq 0.5$, given the higher contribution of overlapping reads, we opted to use all originally assigned reads for ASE analysis but classified these genes as “unresolved,” given the difficulty in accurately ascertaining haplotype-specific overlapping read counts. A detailed example illustrating the entire TOGA workflow is provided in fig. S9. In ASE analysis, we calculated adjusted binomial P values using the “resolved,” “resolved by low CPOV,” and “unresolved” genes only. This method was consistently applied to the NEPTUNE WGS + RNA analysis, the NEPTUNE RNA-only analysis, and the analysis of our control samples.

Quantification of gene-level ASE using phASER

We used the phASER suite of tools for ASE analysis (6). This software has optimal performance with population-phased WGS data using the high-confidence phase of common variants for phase anchoring. As input for phASER, we used the population-phased WGS VCF files, WASP-filtered RNA BAM files, the standard phASER blacklist (containing sites with known allelic mapping bias), the phASER HLA blacklist with corrections, the individualized variant blacklists created by our QC pipeline above, and the recently released GENCODE version 45 comprehensive gene annotation (fig. S1). The initial phASER output is ASE data at the variant and haplotype levels, which is later aggregated by the phASER Gene AE tool to produce gene-level ASE quantification. We applied read-filtering criteria as follows: Only reads with a mapping quality ≥ 255 and a base quality ≥ 10 at the heterozygous variant were considered, and insertions and deletions (indels) were excluded. In addition, duplicate reads were removed by phASER on the basis of samtools flags, and any reads that did not map as a proper pair were also removed. Using the phASER-generated haplotypic reads, we used phASER Gene AE to generate gene-level haplotypic counts for all NEPTUNE samples. GENCODE version 45 (comprehensive gene annotation with primary assembly regions) was used in phASER Gene AE as the reference for gene annotations (94).

To determine which instances of unequal haplotypic expression consisted of significant ASE—i.e., deviated sufficiently from the expected random distribution of allelic counts—we took into account statistical and biological considerations. We only considered genes with total RNA coverage of 20 reads (COV) on the basis of the prior literature reporting high ASE accuracy (12, 37). To quantify the magnitude of ASE, we defined the term MHF, calculated as [preferentially expressed (major) haplotypic read count/total haplotypic read count per gene], thus always ranging from 0.5 (balanced expression) to 1.0 (MAE), similar to Baran *et al.* (37). We performed binomial testing using actual counts to compare the MHF to the expected null distribution (0.5), followed by adjustment for multiple comparisons with the Benjamini-Hochberg method. The significance level of P_{adj} was set at <0.05 . To address the overdispersion of ASE, we instituted an allelic imbalance threshold of one haplotype being expressed at least 50% more than the other, corresponding to an MHF of ≥ 0.6 . A similar approach and cutoff have been described in other ASE studies (3, 7).

Integration and comparison with independent kidney multiomic datasets

To validate the reliability and cellular robustness of our ASE pipeline, we integrated external datasets from Liu *et al.* (20), which included bulk RNA-seq ASE from microdissected GLOM and TUBE (bASE), kidney cortex bASA, and snASA.

Bulk ASE gene-level comparison

We evaluated the proportion of genes exhibiting ASE in at least one NEPTUNE individual that were also identified as ASE genes by RASQUAL (95) in GLOM and TUBE tissues (bASE). Consistency was evaluated across varying stringencies by filtering for increasingly “common” NEPTUNE ASE genes (defined by individual recurrence thresholds). Concordance of genetic effects was verified by correlating the RASQUAL feature SNP effect sizes with the corresponding NEPTUNE eQTL beta values.

Upstream bulk ASA enrichment

To evaluate whether ASE genes were enriched for upstream regulatory accessibility, we assessed intergenic regions upstream of the transcription start site using three window sizes (1, 20, and 50 kb).

Because bASA SNPs are of cohort level and map to open chromatin peaks rather than specific genes, we modeled enrichment using a logistic regression framework. For a given NEPTUNE individual and ASE gene pair, an upstream peak was considered a match if (i) the gene harbored an open chromatin peak in its upstream region and (ii) the specific individual was heterozygous for the cohort-level bASA SNP associated with that peak's accessibility. The logistic regression model evaluated the presence of these matched bASA-SNP peaks in ASE genes versus BAL genes, adjusting for tissue compartment and window size.

Single-nucleus ASA cell-type specificity analysis

To evaluate the potential confounding effect of cell-type heterogeneity on bulk tissue ASE signals, we extended the upstream peak overlap framework to the snASA dataset. Individual-ASE gene pairs in NEPTUNE were matched with upstream heterozygous snASA-SNP peaks. The cell-type specificity profiles of these overlapping peaks were used as a proxy to estimate the contribution of distinct cell types to an individual's overall ASE signals. Because of sample size and power constraints in the single-nucleus data, this cell-type heterogeneity evaluation was restricted to the TUBE compartment.

Identification of imprinted genes

A list of known imprinted genes was curated using the union of two sources: Tucci *et al.* (39), who identified 313 human imprinted genes through an extensive literature review, and Baran *et al.* (37), who used statistical modeling to identify 42 tissue-specific imprinted candidate genes in 48 human adult tissues from the GTEx dataset. The union of these two datasets resulted in 316 known imprinted genes. We then corrected discrepancies in gene names between the two datasets and GENCODE version 45 annotation (94) and compiled a final list of 299 known imprinted genes (table S6).

To identify imprinted genes in our NEPTUNE samples, we incorporated the Baran pipeline, which assigns an imprinting likelihood to a gene on the basis of ASE data. It uses a statistical model to label a variant in an individual as monoallelic, imbalanced, or biallelic expression on the basis of the allelic counts. These variants in a gene are then aggregated across all individuals to compute a generalized likelihood ratio for imprinting (impgrl) for the gene. It uses a unique characteristic of imprinted genes, which is that the MAE pattern of a variant is agnostic to allelic identity because of the parent-of-origin mechanism of imprinting. Thus, an imprinted gene is expected to contain variants with MAE of the reference allele in some individuals and MAE of the alternative allele in others.

As input to the Baran pipeline, we used variant-level allelic counts from phASER. We only analyzed variants with RNA coverage ≥ 8 in COV genes. In addition, we applied a Hardy-Weinberg equilibrium $P \geq 0.001$ following the Baran pipeline. This resulted in an average of $10,007 \pm 1571$ variants per individual in GLOM and 8944 ± 1415 variants per individual in TUBE. Genes containing such variants ($n = 15,963$ GLOM; $n = 14,886$ TUBE) were then required to pass two additional criteria to be labeled as imprinted, as outlined by Baran *et al.* (37). First, the gene required sufficient data to consider its imprinted status. This included having at least two distinct variants and at least five individuals with a heterozygous variant in the gene. In addition, the gene was required to contain a variant with monoallelic reference expression in one individual and monoallelic alternative expression in another (a unique signature of parent-of-origin mechanism). We removed heterogeneously imprinted genes using the pipeline's hetlr metric, defined as the likelihood ratio of heterogeneity of imprinting across the population, compared to other imprinting

statuses. Genes with $\text{hetlr} \geq 1$ were defined as heterogeneously imprinted and removed. The remaining genes were amenable to imprinting analysis ($n = 2419$ GLOM; $n = 2619$ TUBE) and were tested against the second criterion: a sufficient likelihood of imprinting. A gene was labeled as imprinted if its generalized likelihood ratio for imprinting (impplr) > 40 and $\geq 80\%$ of its variant-individuals within the gene showed MAE.

Following this initial analysis, we found that our “previously unidentified” imprinted candidates were significantly enriched for strong eQTLs ($P < 1 \times 10^{-6}$) compared to “known” imprinted genes. This suggested that a strong eQTL effect could mimic an imprinting signal. Therefore, we developed a subsequent filter using three binomial tests to assess whether an eQTL provided a better explanation for the observed ASE. For this correction, we gathered each gene’s significant eQTLs ($P < 1 \times 10^{-6}$) from the NEPTUNE eQTL data (38); if no significant eQTL was present, we used the variant with the lowest eQTL P . For each eQTL in a gene, we stratified the gene’s coding variant-individual pairs into three groups. The “homozygous eQTL group” held all coding variant-individuals where the individual was homozygous for the eQTL. The “heterozygous alternative group” encompassed all coding variant-individuals where the individual was heterozygous for the eQTL and the alternative allele of the coding variant was phased with the alternative allele of the eQTL (purple group in figs. S19 to S22). Last, the “heterozygous reference group” encompassed all coding variant-individuals where the individual was heterozygous for the eQTL and the reference allele of the coding variant was phased with the alternative allele of the eQTL (red group in figs. S19 to S22). We designed the three binomial tests such that the expected result was imprinting, while the alternative hypothesis represented an eQTL-driven ASE mechanism. Any gene initially labeled as an imprinting candidate that showed a significant P value (supporting the eQTL hypothesis) in ≥ 2 of these three tests was relabeled as an eQTL (figs. S19 to S22).

1) Homozygous eQTL group: For an imprinted gene, expression should remain monoallelic. For an eQTL, expression should become BAL. Therefore, a “success” (imprinting) was defined as near-MAE ($\geq 90\%$ RNA reference or alternative allele ratio). We set an expected success rate of 90%. A $P_{\text{hom}} < 0.05$ supported an eQTL mechanism.

2) Heterozygous alternative and reference groups: For an imprinted gene, expression should be 50% monoallelic reference and 50% monoallelic alternative, regardless of phasing. A binomial “success” (imprinting) was defined as an RNA reference allele ratio $\geq 50\%$, with an expected success rate of 50%. The alternative hypothesis (eQTL) posited that the allelic imbalance direction would be consistent with the eQTL allelic direction. Note that we have two tests here as we perform the test for each allele (reference and alternative), separately.

Characterization of gene- and variant-level ASE mechanisms

To assign a molecular mechanism to each ASE event at the gene-individual level, we first filtered for resolved or unresolved (as determined by TOGA) ASE gene-individual pairs with RNA read coverage ≥ 20 (COV genes). Each ASE gene-individual event was then assigned as follows. First, if a gene was identified as “imprinted” by the pipeline described above, that label took precedence over other mechanisms. If not imprinted, gene-individual pairs were labeled as “eQTL” if the individual was heterozygous for a variant with an eQTL nominal $P < 1 \times 10^{-6}$ associated with that gene within 1 Mb of the gene using NEPTUNE eQTL results from our prior

study (38). For sQTLs, we used GTEx version 10 sQTL-sGene pairs from the kidney cortex (22), as determined by LeafCutter (96). If the individual was heterozygous for a significant sQTL (i.e., any variant in the `sQTLs.signif_pairs.tsv` file from GTEx), the associated gene-individual pair was labeled as “sQTL.” ASE gene-individual pairs containing both a significant sQTL and eQTL were labeled “sQTL + eQTL.”

We also evaluated ASE at the variant level to assess allelic preference across individuals. We focused on variants with RNA coverage ≥ 8 within COV genes. Similar to gene-level ASE, an “ASE variant” was defined as a variant where one allele’s expression was 50% greater than the other, with an adjusted binomial $P < 0.05$ (expected = 0.5). On average, $3.83 \pm 2.85\%$ of variants were ASE variants per individual in GLOM and $1.76 \pm 0.94\%$ in TUBE.

At the variant level, we calculated a variant’s allelic preference across the population as the proportion of individuals with ASE of the alternative allele, relative to the total number of individuals with ASE for that variant. We filtered for variants where $> 10\%$ of individuals were heterozygous and $\geq 50\%$ of those heterozygous individuals had ASE. This was applied because variants with ASE in only a few individuals are limited to discrete, nongranular preference values (e.g., 0, 50, or 100% for two individuals). Subsequently, we assessed the overlap between these ASE variants and both eQTLs and sQTLs to determine whether these mechanisms were enriched among variants with high directional consistency. To label eQTLs and sQTLs, we applied the same criteria as gene-level ASE classifications as above, except that the ASE variant should meet those criteria. Among the 55,423 variants with ASE in ≥ 1 individual in NEPTUNE GLOM, 2485 (4.48%) were eQTLs and 2786 (5.03%) were sQTLs.

To overlap ASE variants with meQTLs, we used the Human Kidney meQTL Atlas (https://susztaklab.com/Kidney_meQTL/Download.php) (40). An ASE variant was labeled as a meQTL mechanism if it was associated with methylation that ultimately drove its own expression. Specifically, an ASE variant was labeled meQTL-driven if (i) it was a significant meQTL [FDR (false discovery rate) < 0.01] for a CpG in the file “Kidney_meQTL_S443_Significant.q0.01.txt.gz,” and (ii) that same CpG was a significant eQTM (FDR < 0.05) for the gene that contained the ASE variant, found in the file “Kidney.eQTM.S414.Significant.CpGlevel.FDR0.05.txt.gz.” We identified meQTL-driven ASE among common ASE variants in GLOM/TUBE (i.e., heterozygous genotype in $\geq 10\%$ of samples, of which $\geq 50\%$ had ASE) and explored overlap with other mechanisms, eQTL effect size concordance, and ASE directionality.

Prediction of regulatory variants in promoters

To explore the effect of promoter regulatory variants on ASE, we compared the prevalence of such variants in the promoters of ASE genes against a baseline established from BAL (non-ASE) genes. We conducted the following analysis on all COV genes in all individuals: We first identified the open chromatin region containing the gene’s transcription start site and gathered all heterozygous variants within that region. We then used the ChromKid model developed by Loeb *et al.* (41) to score each heterozygous variant’s likelihood of affecting chromatin accessibility. To score variants in GLOM, we computed the average of the variant’s absolute individual score in immune cells, T cells, stromal cells, podocytes, and endothelial cells. For TUBE, we computed the average of the variant’s absolute score in immune cells, collecting duct, T cells, endothelial cells, distal

tubule, proximal tubule, and loop of Henle. Variants that scored in the top 20% (≥ 28.08 in GLOM; ≥ 21.72 in TUBE) were labeled as PRVs.

We subsequently split the COV gene-individual pairs into three groups: BAL expression, ASE with a known mechanism, and ASE with an unknown mechanism. We computed the proportion of BAL gene-individual pairs with a PRV in their promoters, thus generating a baseline likelihood of identifying PRVs in a typical gene's promoter. We stratified these groups by the rarity (MAC) of the PRV (e.g., proportion where the PRV has an MAC = 1, MAC ≤ 5 , etc.).

For known/unknown ASE, we first grouped the gene-individual pairs by the number of individuals with ASE for that gene, generating subsets of increasing ASE gene commonality (e.g., ASE observed in 1, ≤ 5 , ≤ 10 , ..., ≤ 150 individuals). As with the BAL group, in each ASE commonality subset, we calculated the proportion of gene-individual pairs that had a promoter PRV, stratifying these groups by the PRV rarity. In summary, ASE gene-individual pairs were grouped by ASE mechanism (known/unknown), ASE gene commonality, and MAC of the promoter PRV.

Last, we calculated the enrichment of PRVs of a given MAC in each ASE group by normalizing the proportions to the corresponding BAL gene group. For instance, consider a subset defined by private ASE events (number of individuals with ASE = 1) of an unknown ASE mechanism. If 8% of these gene-individual pairs have a singleton PRV (MAC = 1), this value was compared to the baseline proportion of BAL gene-individual pairs carrying a singleton PRV in their promoters (e.g., 4%). The resulting PRV enrichment in this group would be 2, indicating a twofold enrichment of singleton PRVs in private ASE genes of unknown mechanism.

RMSE and haplotypic flip analysis

To measure overall discordance across heterozygous allele counts on the same haplotype, we calculated the RMSE of the haplotype A fraction for each ASE gene. This was computed using the gene-level haplotype A fraction from phASER Gene AE as the expected value, summed over the variant-level haplotype A fractions for all variants within that gene as follows

$$\text{RMSE}_{\text{gene}} = \sqrt{\sum_i^{n_{\text{var}}} \frac{(\mu_{\text{var},i} - \mu_{\text{gene}})^2}{n_{\text{var}}}} \quad (5)$$

where n_{var} is the number of variants in the gene, $\mu_{\text{var},i}$ is the haplotype A fraction of the i -th variant, and μ_{gene} is the haplotype A fraction of the gene. We included only variants with RNA coverage ≥ 8 in COV genes with ≥ 2 such variants. ASE genes were labeled by their mechanism, as described previously.

Our hypothesis was that sGenes would exhibit higher RMSE values because of variable inclusion/exclusion of variants within a gene. To accurately measure this, no gene-level haplotypic flips should occur within the gene. A haplotypic flip is a technical error where variants within a gene are incorrectly phased, resulting in the assignment of variant allelic counts to the wrong haplotype. In phASER, this can occur when two adjacent haplotype blocks for a gene are not connected by an RNA read or "edge." When this occurs, phASER falls back on WGS phasing to connect the blocks. While this WGS-based connection is often correct, it is less reliable and prone to error, particularly for rare variants. Therefore, for the RMSE analysis, we only used "connected" gene-individual pairs, defined as those where all

haplotype blocks had an edge to adjacent blocks, making no gene-level haplotypic flips possible. While connected genes represented only 8.53% of all gene-individual pairs, this filtering was critical for an accurate RMSE analysis.

The importance of this filter was confirmed using imprinted genes. Before filtering, some imprinted gene-individual pairs displayed spuriously high RMSE values (Fig. 3B and fig. S25E). Further investigation revealed that these were caused by gene-level haplotypic flips. We determined the prevalence of these flips by identifying imprinted genes where two variants within the same individual showed MAE ($\geq 90\%$ allelic ratio) toward opposite haplotypes. Filtering for "connected" genes successfully removed these high-RMSE imprinted genes, resulting in the expected low RMSE values for this category (Fig. 3B and fig. S25E). We note that the Baran pipeline (37) for labeling imprinted genes was not affected by these flips, as it directly uses variant-level allelic counts. Last, our hypothesis of higher RMSE values in sGenes was specific to sGenes with heterozygous variants located both within and outside the alternatively spliced intron; we termed these "Mixed sGenes." Many sGenes had heterozygous variants located only within the alternatively spliced region or only outside of it, leading all measured variants to have a similar effect. We termed these "Uniform sGenes." As expected, the "Mixed sGene" category displayed substantially higher RMSE values than any other ASE mechanism, including "Uniform sGenes." Therefore, the "sGene" category presented in Fig. 3B and fig. S25E consists exclusively of these "Mixed sGenes."

Calculating cell-type proportion of ASE genes using single-cell RNA-seq

To calculate the cell-type proportion of ASE genes, we used single-cell RNA-seq data generated by the KPMP consortium (24). This dataset, generated using the 10x Genomics platform, consisted of 79 samples (14 AKI, 37 CKD, and 28 healthy), comprising 31,496 genes and 225,177 cells in total. We systematically classified the 54 distinct cell types from KPMP into GLOM and TUBE compartments and subsequently regrouped them for analysis (table S7). Unique cell types within the GLOM compartment included PODs and PECs. The TUBE compartment uniquely comprised PTs, intercalated cells (ICs), distal convoluted tubule cells (DTs), thick ascending limb cells (TALs), descending/ascending thin limb cells (DTLs/ATLs), and principal cells (PCs). Cell types present in both compartments included FIBs, IMMUNE, ECs, and vSMCs/MCs.

For each compartment, normalized expression of each gene was quantified using a pseudobulk approach. Specifically, raw counts from individual cells were summed within each grouped cell type per gene. These counts were then normalized by the total expression of that cell type and scaled by a factor of 10,000. The cell-type proportion for a given gene was subsequently calculated as the proportion of its normalized expression within a specific cell type, divided by the total normalized expression for that gene across all grouped cell types in the corresponding compartment.

To define the cell-type predominance of ASE genes, we considered an ASE gene to be enriched in a particular cell type if its cell-type fraction exceeded twice the expected value under uniform expression. For example, in GLOM, with six grouped cell types, a gene was considered cell type-enriched if its fraction was $>(1/6) \times 2 = 0.33$. For TUBE, with 10 grouped cell types, the threshold was $>(1/10) \times 2 = 0.2$.

Cohort of “control” kidneys

The control kidney cohort comprised 31 samples: 23 samples of nonaffected kidney tissue from tumor nephrectomies in the PRECISE cohort (97) and 8 living kidney donor biopsies from the University of Michigan (table S1). All control samples were microdissected into GLOM and TUBE following the same protocol as NEPTUNE; only paired-end bulk RNA-seq data were available for these samples. Participants in the PRECISE cohort had clinically preserved kidney function ($\text{eGFR} \geq 60 \text{ ml/min}$) without proteinuria (urine dipstick with negative or trace protein – within normal range). PRECISE samples were surgical sections instead of needle biopsy. Living donors met standard screening criteria for kidney donation, including age >18 years, preserved glomerular filtration rate (GFR), and the absence of hypertension, diabetes, or proteinuria (specific demographic and clinical details were not available for this subgroup). Ancestry and sex were predicted by Peddy (98) on the basis of RNA-seq–derived genotypes. These study cohorts have been approved by UMICH IRB (PRECISE: HUM00165536, HUM00052918; Living donors: HUM0002468). Written informed consent was obtained from all control cohort participants (PRECISE and living donors) before inclusion.

ASE analysis pipeline using only RNA-seq data

To analyze our control cohort where only RNA-seq data are available, we developed a parallel optimized ASE analysis pipeline, which we termed the “RNA-only” ASE pipeline (fig. S28A). This workflow starts with RNA BAM files aligned by STAR (version 2.7.11b) and duplicate marked by samtools (version 1.19.2). With the BAM file as input, genotype calling was then performed using the GATK variant calling workflow (version 4.5.0.0; <https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels>). We applied SplitNCigarReads to segment paired-end reads with Ns in their CIGAR string. BaseRecalibrator was then used to correct base quality scores with inputs of known polymorphic sites (SNPs and indels) from the GATK Resource Bundle and reference genome from GENCODE version 45 primary assembly (94). After applying quality recalibration with ApplyBQSR, variants were called using HaplotypeCaller using a minimum phred-scaled confidence threshold of 20. Last, hard filtering was performed with VariantFiltration, retaining variants with Fisher Strand ≤ 30 and Quality by Depth ≥ 2 . Only autosomal heterozygous variants were retained in the final VCF files.

An initial ASE analysis using this RNA-derived VCF and WASP-filtered RNA BAMs revealed a median of six monoallelic genes per individual, a strong indication of false-positive heterozygous calls, as the variants supporting these monoallelic genes should have been called homozygous. We identified two primary causes: (i) the use of non-WASP-filtered BAMs in the variant calling step, which allowed reference-biased reads to be miscalled as heterozygous, and (ii) low-quality reads contributing to erroneous calls. To address this, we refined the pipeline by rerunning the entire GATK variant calling workflow using the WASP-filtered BAMs and removed all variants that exhibited MAE. To further increase confidence, we excluded variants not present in the 1000 Genomes (1KG) project (99) and variants listed in the standard phASER blacklist. The subsequent steps for ASE calling were identical to those used in the WGS + RNA pipeline, including running phASER, applying TOGA, and calling ASE genes.

Survival analysis

To assess the association between disease progression and an elevated GLOM ASE proportion, we performed survival analysis. For the primary analysis, the event variable was defined as the occurrence of ESKD or a $\geq 40\%$ decline in eGFR from the time of biopsy (eGFR40). The time variable was the number of days from biopsy to the ESKD or eGFR40 event. Right-censoring was applied to individuals without events, and samples without eGFR information were excluded, resulting in a total of 194 samples for this analysis. We defined kidney failure as two consecutive eGFR $<15 \text{ ml/min}$ or ESKD requiring kidney replacement therapy.

We also analyzed time to complete remission among patients with active disease at biopsy. This analysis was restricted to patients with a urine protein-to-creatinine ratio (UPCR) $> 0.3 \text{ g/g}$ at biopsy, indicating active disease. Time was calculated from the biopsy date to either complete remission or the last follow-up. The primary end point was the achievement of complete remission, defined as UPCR $\leq 0.3 \text{ g/g}$. Patients were censored at their last follow-up if remission had not been achieved. Using the same groups as our primary analysis, we excluded samples without remission data, resulting in a final cohort of 172 samples (84 in the high-GTAR group and 88 in the low-GTAR group).

Last, we assessed time to first relapse among patients who were in remission at the time of biopsy. Time was calculated from the biopsy date to either the first relapse or the last follow-up. The end point was disease relapse, defined as UPCR $> 3.5 \text{ g/g}$. Patients were censored at their last follow-up if they remained relapse-free. This analysis included 21 patients (12 in the high-GTAR group and 9 in the low-GTAR group).

Cell-type deconvolution analysis of NEPTUNE bulk RNA-seq

We estimated broad cell-type proportions in NEPTUNE GLOM bulk RNA-seq samples using MuSiC (43) and a single-nucleus RNA-seq reference from the NEPTUNE cohort (44). Broad cell-type groups were defined separately for each compartment. GLOM groups comprised PODs, PECs, ECs, vSMCs/MCs, FIBs, and IMMUNE. To balance the reference data across broad cell-type groups, we capped the number of nuclei at 3000 per group and 150 per donor-group combination, yielding 18,000 GLOM reference nuclei from 93 donors. Bulk RNA-seq expression matrices were supplied to MuSiC as non-log-transformed, nonnegative input matrices without marker-gene restriction. Reference and bulk matrices were harmonized by gene symbol, and MuSiC was run using all shared genes with nonzero expression, comprising 20,936 GLOM genes. Deconvolution was performed with music_prop. We used Est.prop.weighted as the primary estimate of cell-type proportions. MuSiC returned weighted proportion estimates for 194 GLOM samples, and estimated proportions summed to 1 within floating-point tolerance.

DGE and GSEA

We performed DGE analysis to identify genes differentially expressed between the high- and low-GTAR groups, as defined in the survival analysis. Three low-GTAR samples without complete information for sex, age, race, or RNA-seq batch were excluded, resulting in 97 samples in the high-GTAR group and 94 samples in the low-GTAR group. DGE analysis was performed using DESeq2 (45) with RNA counts quantified by RSEM. The model adjusted for sex, age, race, and RNA-seq batch as covariates. Only autosomal genes with

at least 10 counts in a minimum of three samples were retained for the analysis to exclude low-count, high-variability genes.

GSEA was subsequently performed using clusterProfiler (85), with genes ranked by the product of $\text{signed}(\log_2\text{FC}) \times -\log_{10}(P)$ from the DGE analysis. Genes with null-adjusted P values from DESeq2 were excluded. The gene groups evaluated were from the GO databases, including Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) terms, and pathway databases (KEGG and WikiPathways). P values were adjusted using the Benjamini-Hochberg correction within each database, with significant terms defined as those with an adjusted $P < 0.05$. As a complementary approach, ORA was also performed on the sets of significantly up-regulated ($\log_2\text{FC} > 0.5$, $P_{\text{adj}} < 0.05$) and down-regulated ($\log_2\text{FC} < -0.5$, $P_{\text{adj}} < 0.05$) genes identified by DESeq2.

Identification of differential ASE genes in GLOM between high- and low-GTAR groups

To identify differential ASE genes between the high- and low-GTAR groups, we constructed a 2×2 contingency table of ASE versus BAL (balanced expression = COV but not ASE) counts across the two GTAR groups for each ASE gene and applied a one-sided Fisher's exact test. We only included genes that had ASE in GLOM in at least one individual and further restricted this set to genes with COV in $\geq 10\%$ individuals within each of the high- and low-GTAR groups, resulting in 10,872 genes. For multiple testing correction, we applied the Benjamini-Hochberg method and used $P_{\text{adj}} < 0.05$ as a cutoff to determine significant differential ASE. Subsequent ORA using 166 significant genes and GSEA using all 10,872 genes ranked by $-\log_{10}(P)$ were performed to find significant GO terms ($P_{\text{adj}} < 0.05$).

Construction of the non-ASE expression-based prognostic signature

To derive a simplified prognostic signature using standard bulk tissue expression data, we selected core enrichment genes from the top up- and down-regulated GO terms identified in the high-GTAR versus low-GTAR GSEA (table S11). The up-regulated gene set (G_{UP} , $n = 216$ genes) was constructed by aggregating core enrichment genes from ribosomal, mitochondrial, and ATP-synthesis pathways. The down-regulated gene set (G_{DOWN} , $n = 191$ genes) was compiled from pathways associated with immune response, ECM/adhesion, tRNA-aminoacylation, and endoplasmic reticulum (ER)/vesicle transport. This gene selection mirrors the primary pathways shown in Fig. 6D, with two modifications: The generic term “hydrolase activity, hydrolyzing O-glycosyl compounds” was excluded, and the “type I interferon-mediated signaling pathway” was appended to G_{DOWN} to comprehensively capture immune activation signals. Single-sample gene set enrichment scoring was performed using the rank-based “singscore” R package (86), which quantifies expression phenotypes within an individual sample without requiring cross-sample normalization. For each individual i , all genes were ranked on the basis of their expression values in TPM ($\text{TPM}_{g,i}$, representing the expression of specific gene g in individual i), where N represents the total number of genes evaluated in the sample

$$\text{TPM}_i = (\text{TPM}_{1,i}, \text{TPM}_{2,i}, \dots, \text{TPM}_{N,i})$$

Tied expression values were assigned the minimum shared rank. Normalized enrichment scores for the up-regulated set (S_{UP}) and the down-regulated set (S_{DOWN}) were calculated relative to the empirical distribution of all gene ranks within that individual, yielding

stable scores bound between 0 and 1. A higher score signifies a higher median rank of the target gene set relative to the background transcriptome of that sample. Last, a directional composite total score ($S_{\text{TOTAL},i}$) was computed for each individual i as the difference between the two enrichment scores

$$S_{\text{TOTAL},i} = S_{\text{UP},i} - S_{\text{DOWN},i}$$

Supplementary Materials

The PDF file includes:

Supplementary Text

Figs. S1 to S36

Legends for tables S1 to S12

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S12

REFERENCES

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. Dermizakis, T. Lappalainen, Allelic mapping bias in RNA-sequencing is not a major confounder in eQTL studies. *Genome Biol.* **15**, 467 (2014).
11. S. Mattevi, F. Mazzarotto, P. Martini, Allele-specific expression analysis: Pipelines, applications, challenges, and unmet needs. *Comput. Biol. Med.* **196**, 110890 (2025).
12. J. Fan, J. Hu, C. Xue, H. Zhang, K. Susztak, M. P. Reilly, X. Xiao, M. Li, ASEP: Gene-based detection of allele-specific expression across individuals in a population by RNA sequencing. *PLoS Genet.* **16**, e1008786 (2020).
13. 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).
14. 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. Chandraratnam, M. F. Berger, B. S. Taylor, Widespread selection for oncogenic mutant allele imbalance in cancer. *Cancer Cell* **34**, 852–862.e4 (2018).
15. 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).
16. M. Langmyhr, S. P. Henriksen, C. Cappelletti, W. D. J. van de Berg, L. Pihlström, M. Toft, Allele-specific expression of Parkinson's disease susceptibility genes in human brain. *Sci. Rep.* **11**, 504 (2021).
17. 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).

18. M. I. Sigurdsson, L. Saddic, M. Heydarpour, T.-W. Chang, P. Shekar, S. Aranki, G. S. Couper, S. K. Shernan, J. G. Seidman, S. C. Body, J. D. Muehlschlegel, Allele-specific expression in the human heart and its application to postoperative atrial fibrillation and myocardial ischemia. *Genome Med.* **8**, 127 (2016).
19. 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).
20. H. Liu, A. Abedini, E. Ha, Z. Ma, X. Sheng, B. Dumoulin, C. Qiu, T. Aranyi, S. Li, N. Ditttrich, H.-C. Chen, R. Tao, D.-C. Tarnag, 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).
21. S. E. Castel, F. Aguet, P. Mohammadi, GTEx Consortium, K. G. Ardlie, T. Lappalainen, A vast resource of allelic expression data spanning human tissues. *Genome Biol.* **21**, 234 (2020).
22. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
23. F. A. Garza, S. W. Leslie, "Anatomy, abdomen and pelvis: Kidneys" in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2025); www.ncbi.nlm.nih.gov/books/NBK482385/.
24. 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. Hodgins, 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).
25. I. Ogoburo, F. Tuma, "Physiology, renal" in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2025); www.ncbi.nlm.nih.gov/books/NBK538339/.
26. K. Meliandro, J. C. He, K. N. Campbell, Podocyte-targeted therapies—Progress and future directions. *Nat. Rev. Nephrol.* **20**, 643–658 (2024).
27. N. Mann, H. Sun, A. J. Majmundar, Mechanisms of podocyte injury in genetic kidney disease. *Pediatr. Nephrol.* **40**, 1523–1538 (2025).
28. E. E. Groopman, M. Marasa, S. Cameron-Christie, S. Petrovski, V. S. Aggarwal, H. Milo-Rasouly, Y. Li, J. Zhang, J. Nestor, P. Kithivasan, 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. Bridges, C. D. Malone, C. M. Mebane, N. Dagoanar, 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).
29. 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. Cawley, 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).
30. C. A. Gadegebeku, 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. Catran, 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. K. 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).
31. 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. *bioRxiv* 085050 [Preprint] (2016); <https://doi.org/10.1101/085050>.
32. 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).
33. 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).
34. 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).
35. 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).
36. R. Asimwe, D. Alexander, STAR+WASP reduces reference bias in the allele-specific mapping of RNA-seq reads. *bioRxiv* 576391 [Preprint] (2024); <https://doi.org/10.1101/2024.01.21.576391>.
37. 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).
38. 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. Burt, 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).
39. 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).
40. H. Liu, T. Doke, D. Guo, X. Sheng, Z. Ma, J. Park, H. M. T. Vy, G. N. Nadkarni, A. Abedini, Z. Miao, M. Palmer, B. F. Voight, H. Li, C. D. Brown, M. D. Ritchie, Y. Shu, K. Susztak, Epigenomic and transcriptomic analyses define core cell types, genes and targetable mechanisms for kidney disease. *Nat. Genet.* **54**, 950–962 (2022).
41. G. B. Loeb, P. Kathail, R. W. Shuai, R. Chung, R. 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).
42. V. Nembaware, B. Lupindo, K. Schouest, C. Spillane, K. Scheffler, C. Seoghe, Genome-wide survey of allele-specific splicing in humans. *BMC Genomics* **9**, 265 (2008).
43. X. Wang, J. Park, K. Susztak, N. R. Zhang, M. Li, Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* **10**, 380 (2019).
44. P. J. McCown, C. Boys, S. Datta, V. Nair, E. Otto, D. Fermin, J. Hartman, F. Annesse, F. Alakwaa, K. Latt, C. Smith, B. Godfrey, M. Larkina, A. E. Ashley-Koch, M. E. Garrett, S. Moore, K. Soldano, M. Alaba, M. T. McNulty, D. Lee, M. G. Sampson, L. H. Mariani, A. S. Naik, R. Fallegger, S. Eddy, J. Saez-Rodriguez, O. A. Olabisi, M. Kretzler, Nephrotic Syndrome Study Network (NEPTUNE), A Human Glomerular Disease Atlas defines the APOL1-JAK-STAT feed forward loop in focal segmental glomerulosclerosis. *medRxiv* 25335572 [Preprint] (2025). <https://doi.org/10.1101/2025.09.12.25335572>.
45. 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).
46. 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).
47. 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).
48. 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).
49. 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.* **12**, 774436 (2021).
50. P. Pauksakon, B. Najafian, C. E. Alpers, A. B. Fogo, AJKD atlas of renal pathology: Lysozyme-associated nephropathy. *Am. J. Kidney Dis.* **86**, e1–e3 (2025).
51. R. I. Menzies, F. W. Tam, R. J. Unwin, M. A. Bailey, Purinergic signaling in kidney disease. *Kidney Int.* **91**, 315–323 (2017).
52. 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).
53. S. 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).
54. 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).

55. 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).
56. 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).
57. 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).
58. 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).
59. H. A. Niederer, M. R. Clatworthy, L. C. Willcocks, K. G. C. Smith, FcγmαRIIB, FcγmαRIIIB, and systemic lupus erythematosus. *Ann. N. Y. Acad. Sci.* **1183**, 69–88 (2010).
60. 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 Fcγmα receptor 2b deficient mice, but not mice with conditional deletion of Fcγr2b on either B cells or myeloid cells alone. *Mol. Immunol.* **50**, 49–56 (2012).
61. 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).
62. C. E. Alpers, K. L. Hudkins, M. Ferguson, R. J. Johnson, G. C. Schattelman, M. Bothwell, Nerve growth factor receptor expression in fetal, mature, and diseased human kidneys. *Lab. Invest.* **69**, 703–713 (1993).
63. 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).
64. Y. Liu, Cellular and molecular mechanisms of renal fibrosis. *Nat. Rev. Nephrol.* **7**, 684–696 (2011).
65. 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).
66. 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).
67. Y. Sato, M. Yanagita, Resident fibroblasts in the kidney: A major driver of fibrosis and inflammation. *Inflamm. Regen.* **37**, 17 (2017).
68. 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., Exp. Biol.* **37**, e22938 (2023).
69. T. Wada, N. Sakai, K. Matsushima, S. Kaneko, Fibrocytes: A new insight into kidney fibrosis. *Kidney Int.* **72**, 269–273 (2007).
70. 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).
71. 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).
72. 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* **12**, 648 (2022).
73. 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).
74. 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).
75. 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).
76. 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. Weinb. Baden-Wurt. Ger.* **11**, e2406936 (2024).
77. 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).
78. 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).
79. 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).
80. 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).
81. 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).
82. 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).
83. J.-D. Lu, K. Kadir, 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).
84. 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).
85. 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).
86. M. Foroutan, D. D. Bhuvra, R. Lyu, K. Horan, J. Cursons, M. J. Davis, Single sample scoring of molecular phenotypes. *BMC Bioinformatics* **19**, 404 (2018).
87. L. Barisoni, C. C. Nast, J. C. Jennette, J. B. Hodgins, A. M. Herzenberg, K. V. Lemley, C. M. Conway, J. B. Kopp, M. Kretzler, C. Lienczewski, C. Avila-Casado, S. Bagnasco, S. Sethi, J. Tomaszewski, A. H. Gasim, S. M. Hewitt, Digital pathology evaluation in the multicenter Nephrotic Syndrome Study Network (NEPTUNE). *Clin. J. Am. Soc. Nephrol.* **8**, 1449–1459 (2013).
88. C. C. Nast, K. V. Lemley, J. B. Hodgins, S. Bagnasco, C. Avila-Casado, S. M. Hewitt, L. Barisoni, Morphology in the digital age: Integrating high-resolution description of structural alterations with phenotypes and genotypes. *Semin. Nephrol.* **35**, 266–278 (2015).
89. L. Barisoni, J. P. Troost, C. Nast, S. Bagnasco, C. Avila-Casado, J. Hodgins, M. Palmer, A. Rosenberg, A. Gasim, C. Liensziewski, L. Merlino, H.-P. Chien, A. Chang, S. M. Meehan, J. Gaut, P. Song, L. Holzman, D. Gibson, M. Kretzler, B. W. Gillespie, S. M. Hewitt, Reproducibility of the NEPTUNE descriptor-based scoring system on whole-slide images and histologic and ultrastructural digital images. *Mod. Pathol.* **29**, 671–684 (2016).
90. A. A. Regier, Y. Farjoun, D. E. Larson, O. Krasheninnina, H. M. Kang, D. P. Howrigan, B.-J. Chen, M. Kher, E. Banks, D. C. Ames, A. C. English, H. Li, J. Xing, Y. Zhang, T. Matise, G. R. Abecasis, W. Salerno, M. C. Zody, B. M. Neale, I. M. Hall, Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nat. Commun.* **9**, 4038 (2018).
91. S. W. Wingett, S. Andrews, FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res.* **7**, 1338 (2018).
92. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
93. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
94. J. M. Mudge, S. Carbonell-Sala, M. Diekhans, J. G. Martinez, T. Hunt, I. Jungreis, J. E. Loveland, C. Arnan, I. Barnes, R. Bennett, A. Berry, A. Bignell, D. Cerdán-Vélez, K. Cochran, L. T. Cortés, C. Davidson, S. Donaldson, C. Dursun, R. Fatima, M. Hardy, P. Hebbard, Z. Hollis, B. T. James, Y. Jiang, R. Johnson, G. Kaur, M. Kay, R. J. Mangan, M. Maquedano, L. M. Gómez, N. Mathlouthi, R. Merritt, P. Ni, E. Palumbo, T. Perteghella, F. Pozo, S. Raj, C. Sisú, E. Steed, D. Sumathipala, M.-M. Suner, B. Szczyszka-Ratajczak, E. Wass, Y. T. Yang, D. Zhang, R. D. Finn, M. Gerstein, R. Guigó, T. J. P. Hubbard, M. Kellis, A. Kundaje, B. Paten, M. L. Tress, E. Birney, F. J. Martin, A. Frankish, GENCODE 2025: Reference gene annotation for human and mouse. *Nucleic Acids Res.* **53**, D966–D975 (2025).
95. N. Kumasaka, A. J. Knights, D. J. Gaffney, Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat. Genet.* **48**, 206–213 (2016).
96. Y. I. Li, D. A. Knowles, J. Humphrey, A. N. Barreira, S. P. Dickinson, H. K. Im, J. K. Pritchard, Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).
97. J. A. Schaub, C. L. O'Connor, J. Shi, R. C. Wiggins, K. Shedden, J. B. Hodgins, M. Bitzer, Quantitative morphometrics reveals glomerular changes in patients with infrequent segmentally sclerosed glomeruli. *J. Clin. Pathol.* **75**, 121–127 (2022).
98. B. S. Pedersen, A. R. Quinlan, Who's who? Detecting and resolving sample anomalies in human DNA sequencing studies with Peddy. *Am. J. Hum. Genet.* **100**, 406–413 (2017).

99. The 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

Acknowledgments: This study has benefited from useful comments from members of Sampson lab and Lee lab. We 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. The 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 are 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 KPMP (accessed 26 June 2025; 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 acknowledge the essential contributions of our patient participants and the support of the American public through their tax dollars. The 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 UKB Resource under application number 64945.

Funding: This work was supported by the following: National Institutes of Health grant F32DK122766 (to A.C.O.-W.); American Society of Nephrology KidneyCure Sharon Anderson Award (to A.C.O.-W.); NephCure Kidney International-NEPTUNE Pilot Award (to A.C.O.-W.); Brigham and Women's Hospital Hearst Young Investigator Award (to A.C.O.-W.); National Institutes of Health grant R01HG012871 (to D.L.); NEPTUNE Career Enhancement Award (to D.L.); Boston Children's Hospital OFD/BTREC/CTREC Faculty Development Fellowship Award (to D.L.); National Institutes of Health grant R01DK100449 (to M.B.); National Institutes of Health

grant R21DK126329 (to M.B.). PRECISE has been supported by the University of Michigan and NIH/NIDDK. **Author contributions:** Conceptualization: A.C.O.-W., D.L., and M.G.S. Data curation: A.C.O.-W., A.G., D.L., E.D.S., J.S., and M.B. Formal analysis: A.C.O.-W., D.L., E.D.S., J.S., M.T.M., and V.M. Methodology: A.C.O.-W., A.G., C.L.O., D.L., E.D.S., J.G.Y., J.S., L.H.M., M.B., M.G.S., M.T.M., and V.M. Investigation: A.C.O.-W., C.L.O., D.L., E.D.S., J.S., M.B., M.G.S., M.T.M., and S.B. Visualization: A.C.O.-W., D.L., E.D.S., J.S., and M.T.M. Funding acquisition: A.C.O.-W., D.L., M.B., and M.G.S. Project administration: A.C.O.-W., D.L., and M.G.S. Resources: A.C.O.-W., D.L., M.B., and M.G.S. Software: A.C.O.-W., D.L., E.D.S., J.G.Y., J.S., and V.M. Validation: A.C.O.-W., D.L., E.D.S., J.S., M.B., and M.G.S. Supervision: A.C.O.-W., D.L., and M.G.S. Writing—original draft: A.C.O.-W., D.L., E.D.S., and J.S. Writing—review and editing: A.C.O.-W., A.G., C.L.O., D.L., E.D.S., J.S., J.G.Y., L.H.M., M.B., M.T.M., M.G.S., S.B., and V.M. **Competing interests:** M.G.S. is on the scientific advisory board of Natera. A.C.O.-W. is a volunteer, unpaid member of the Medical Advisory Committee of the Alport Syndrome Foundation, a patient-led organization. L.H.M. reports consulting for Boehringer-Ingelheim, Dimerix Limited, Calliditas Therapeutics, Vera Therapeutics, Travere Therapeutics, Novartis, Chinook, Apellis Pharmaceuticals, and Otsuka. The other authors declare that they have no competing interests. **Data, code, and materials availability:** All data and code needed to evaluate and reproduce the conclusions in the paper are present in the paper and/or the Supplementary Materials. Raw WGS and RNA-seq data from GLOM and TUBE and the processed snRNA-seq data are available through NEPTUNE (www.neptune-study.org/ancillary-studies). CONTROL GLOM and TUBE datasets are available through M.B. and NEPTUNE. Single-cell RNA-seq data used to infer cell-type proportions of ASE genes can be accessed from the KPMP repository (<https://atlas.kpmp.org/repository>). eQTL data are available through NephQTL2 (www.nephqtl2.org). The GTEx sQTL data used for the analyses described here, GTEx_Analysis_v10_sQTL.tar, were obtained from www.gtexportal.org/home/downloads/adult-gtex/ctl on 3 January 2025. All code used for the analyses has been archived in a permanent, publicly accessible repository on Zenodo (<https://doi.org/10.5281/zenodo.20737256>); the corresponding development version is additionally available on GitHub (https://github.com/dongwonlee-lab/ASE_NEPTUNE). This study did not generate new physical materials or reagents.

Submitted 10 February 2026

Accepted 28 August 2026

Published 2 October 2026

10.1126/sciadv.aeg3462

The landscape of allele-specific expression in human kidneys

Ana C. Onuchic-Whitford, Junmo Sung, Eric D. Sakkas, Michelle T. McNulty, Victoria Mezger, Christopher L. O'Connor, Anya Greenberg, Jihoon G. Yoon, Sowmya Badina, Laura H. Mariani, Markus Bitzer, Matthew G. Sampson, and Dongwon Lee

Sci. Adv. **12** (40), eaeg3462. DOI: 10.1126/sciadv.aeg3462

View the article online

<https://www.science.org/doi/10.1126/sciadv.aeg3462>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).