The landscape of allele-specific expression in human kidneys

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

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

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

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

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

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

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

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

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

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

† These authors contributed equally to this work.

Abstract

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

Introduction

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

As an intra-individual analysis, ASE minimizes batch effects and environmental confounders to uncover true cis-acting regulatory mechanisms and rare variant effects[6]. However, ASE analysis is subject to technical biases requiring rigorous quality control (QC), particularly regarding haplotype phasing, genotyping, and RNA read mapping[7]. Errors or biases in any of these steps can result in substantial inaccuracies 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. 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 non-diseased tissue either from tumor nephrectomies[12,13,20] or post-mortem samples[21]. Furthermore, datasets such as GTEx[22] do not account for the complexity of kidney structure, which comprises two distinct functional

66 compartments: the glomeruli and tubulointerstitium[23]. As glomeruli comprise only a small
67 fraction of the total cell count in the kidney cortex[24], glomerular transcripts are much less
68 abundant in bulk tissue than those of the tubulointerstitium. This makes it challenging to study
69 ASE of glomerular-specific genes. Such compartmental resolution is particularly important when
70 studying proteinuric diseases, characterized by loss of large amounts of protein in the urine due
71 to defects in glomerular filtration[25]. The most severe form of proteinuric disease is the nephrotic
72 syndrome, with loss of ≥3.5 grams of urinary protein daily, resulting in hypoalbuminemia and
73 swelling of the face and extremities. Proteinuric disorders account for up to 90% of end-stage
74 kidney disease (ESKD) worldwide[26]. Specifically, a common mechanism is injury to glomerular
75 podocytes (specialized cells which comprise only ~1% of kidney cortex), and most of the >70
76 monogenic causes of proteinuria involve genes expressed in this cell type[27]. However, many
77 patients with glomerular disorders remain undiagnosed, and several known genetic etiologies
78 lack a clear mechanistic pathway to disease[28,29].

79 The goal of this study was to discover the genome-wide landscape of ASE in the human
80 kidney and its role in the pathogenesis and progression of proteinuric kidney diseases (**Fig. 1A**).
81 We use a unique dataset of WGS with paired RNA-seq from microdissected compartments of
82 kidney biopsy from the Nephrotic Syndrome Study Network (NEPTUNE)[30]. We create a
83 rigorous, generalizable framework for ASE analysis, integrating both available and novel
84 computational tools. Our pipeline includes new steps to address major sources of bias in ASE,
85 including overlapping genes and concealed genotyping errors. We generate a high-confidence
86 ASE dataset and characterize distinct ASE patterns in glomeruli (GLOM) and tubulointerstitium
87 (TUBE) of diseased kidneys, compared to controls. We then perform extensive investigation into
88 the molecular mechanisms, cell-type specificity, clinical correlations, and differential gene
89 signatures linked to ASE in the human kidney.
90

91 **Results**

92 ***A new workflow for generation of high-confidence ASE data***

93 We assembled a rigorous pipeline for high-confidence ASE calling in kidney tissue from the
94 NEPTUNE proteinuric kidney disease cohort (**Methods**). Biopsies from these individuals were
95 microdissected into GLOM and TUBE compartments, followed by bulk RNA sequencing. We

96 identified 240 NEPTUNE individuals with available paired WGS and GLOM RNA-seq, of
97 which 218 also had TUBE RNA-seq.

98 Our ASE analysis workflow is summarized in **Fig. S1** and described in detail in
99 **Methods**. Here, we highlight three key features introduced in this work: 1) novel QC steps to
100 blacklist variants prone to genotyping errors; 2) bias reduction from misassigned RNA reads in
101 loci containing overlapping genes; and 3) ASE definition using multiple criteria to minimize
102 false positive calls.

103 Briefly, we first performed WGS-based genotyping using the GATK best practices
104 pipeline, followed by genotype phasing using WhatsHap[31] and SHAPEIT4[32]. Both tools employ
105 read-backed phasing, an advantage of WGS, and SHAPEIT4 allowed us to incorporate haplotype
106 data from ~500,000 participants of the UK Biobank[33] to leverage population-level phasing
107 information. As detailed below, our estimated gene-level phasing error (flip) rate was low (~3%),
108 enabling accurate ASE quantification even for genes whose haplotypes were not fully connected
109 by RNA-seq reads.

110 To ensure high-confidence genotype calling, we also implemented multiple WGS QC
111 measures, excluding variants with: 1) inconsistent genotype before and after phasing; 2) biased
112 DNA allelic fraction; or 3) abnormal DNA read depth. These variants showed a significant
113 deviation of DNA (or RNA) allelic fraction from the expected 0.5, suggesting poor genotyping
114 or structural variation (**Fig. S2**). In addition, we blacklisted variants with bi-allelic RNA
115 expression at homozygous WGS genotypes, suggesting systematic poor read alignment in RNA-
116 seq (**Fig. S3**). Altogether, we removed an average of 543 variants per sample. These blacklisted
117 variants showed a substantially higher proportion of variant-level ASE than the QC-passed
118 variants, consistent with false-positive ASE from erroneous genotyping (**Fig. S4**). Per sample, an
119 average of 302 genes harbored these problematic variants, and thus would appear to have
120 significant ASE without this filter. Our QC strategy also identified and excluded 18 problematic
121 samples with an excess number of such problematic variants ($n \geq 1{,}000$) (**Fig. S2D**). Our
122 findings were further confirmed by verifyBamID analysis[34], which flagged 13 of the 18 same
123 samples as having DNA-RNA mismatches. This resulted in a final number of 222 GLOM and
124 206 paired TUBE samples (**Table S1**).

125 Next, to correct for allelic mapping bias in ASE analyses, we performed RNA-seq read
126 alignment using STAR (Spliced Transcripts Alignment to a Reference) with WASP (Workflow
127 for Allele-Specific Processing)[9,35,36] and confirmed the same benefits in our dataset (**Fig. S5**).
128 WASP-passing reads, taken only from transcripts overlapping heterozygous loci, represented
129 ~9% of total reads per sample. We then performed ASE analysis using phASER (phasing and
130 Allele Specific Expression from RNA-seq)[6], incorporating the individualized blacklists described
131 above. phASER employs read-backed phasing from RNA-seq to optimize phasing of rare
132 variants, and accurately phases variants over longer distances due to RNA splicing. In addition,
133 phASER reports gene-level ASE as the number of reads mapping to gene haplotypes, rather than
134 the sum of a gene's variant-level reads. For downstream analyses, we included genes with
135 sufficient coverage (total haplotypic count ≥20, based on prior literature standards[12,37]), and
136 designated these as "COV" genes. Consistently, at lower coverage cutoffs, our data showed
137 inflated ASE and monoallelic expression (MAE, defined as expression of only one haplotype)
138 (**Fig. S6**).

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

154 To quantify the magnitude of ASE, we defined the term "Major Haplotypic Fraction"
155 (MHF): the fraction of reads from the higher-expressed haplotype, with values ranging from 0.5

156 (balanced) to 1.0 (monoallelic). We defined a gene as having significant ASE as follows: (1)
157 sufficient RNA coverage for high-confidence ASE analysis, i.e., post-TOGA COV genes; (2)
158 binomial test with $P_{adj} < 0.05$; and (3) MHF ≥ 0.6 (i.e., one haplotype expressed ≥50% more than
159 the other). This MHF threshold was implemented to address known overdispersion in ASE data,
160 which is not managed by the binomial test alone; a similar approach and cutoff have been
161 described previously[3,7]. **Fig. S10** shows an illustrative ASE analysis plot in one NEPTUNE
162 individual.

163

164 ***The landscape of ASE in human kidneys with proteinuric disease***

165 In our ASE analysis, we achieved extensive coverage of heterozygous variants (**Table S3**),
166 resulting in a median of 5,370 genes in GLOM and 4,989 in TUBE amenable to ASE analysis
167 (COV genes) per individual (**Fig 1B, Table S4**). Cumulatively across NEPTUNE GLOM and
168 TUBE, we observed 18,304 COV genes in ≥1 individual, of which 14,619 (~80%) were protein-
169 coding—thus, most genes in the human genome were covered. While the number of COV genes
170 was similar between compartments, the ASE patterns were distinct. We identified a median of
171 643 GLOM ASE genes per individual, nearly double the median of 361 TUBE ASE genes (**Fig**
172 **1B, Table S4**). Moreover, while COV genes were largely shared (mean 68%) between GLOM
173 and TUBE compartments in an individual, only a mean of 12% of ASE genes were similarly
174 shared. Thus, ASE was predominantly compartment-specific, with more than double the
175 proportion of GLOM-specific compared to TUBE-specific ASE (59% vs. 29%) (**Fig. 1C**).
176 Monoallelic expression (MAE) genes were rare and similarly distinct, with a median of 14
177 GLOM and 8 TUBE MAEs per individual. Collectively, we identified 3,575 GLOM and 968
178 TUBE MAEs in ≥1 individual (**Table S4**). All ASE genes are reported in **Table S5**.

179 COV and ASE distribution per gene is shown in **Fig. 1D**. Most COV genes (~66% in
180 GLOM, and ~67% in TUBE) were shared by ≥10% of individuals. However, concordant with
181 prior studies[19], ASE was largely private. Shared ASE genes followed an exponentially
182 decreasing distribution (**Fig. 1D**, **Fig. S11**), with the majority having ASE in few individuals (≤8
183 in GLOM, and ≤3 in TUBE). In fact, only 13.7% (GLOM) and 6.7% (TUBE) of ASE genes had
184 ASE in ≥10% of individuals, and we later find that these genes share certain molecular

185 mechanisms of ASE, such as imprinting and strong eQTL effects. Overall, this indicates kidney
186 ASE patterns are predominantly individual-specific.

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

195

196 ***Identification of imprinted genes in human proteinuric kidneys***

197 Given the enrichment of imprinted genes among the top ASE genes (**Fig. 1F; Fig. S12**) and their
198 known tissue specificity[37], we sought to systematically identify imprinted genes within our
199 dataset. To do this, we adapted an established pipeline[37] that assigns an imprinting likelihood to
200 every ASE gene. This is based on characteristics distinct from other regulatory mechanisms:
201 monoallelic expression in all individuals where the expressed haplotype varies randomly across
202 individuals due to parent-of-origin effects rather than genotype **(Fig. S13A**). With this pipeline,
203 we initially identified 26 candidate imprinted genes in GLOM (nine not previously identified in
204 any tissue) and 18 candidates in TUBE (five novel) (**Figs. S14-17**).

205 This method has a potential limitation: the risk of genes with eQTLs of very large effect
206 size (leading to near mono-allelic expression) being classified as imprinted. This can occur if the
207 causal non-coding regulatory variant is heterozygous in most individuals and in low linkage
208 disequilibrium with the coding ASE variants. While the prior study concluded that genes meeting
209 these conditions are extremely unlikely[37], we found that our novel candidate imprinted genes
210 frequently overlapped with strong eQTLs[38]. Specifically, 7 out of 9 novel candidates in GLOM
211 had significant eQTLs ($P < 10^{-6}$), compared to only 1 out of 17 of known imprinted genes. A
212 similar pattern was also observed in TUBE. Therefore, we sought to distinguish true imprinting
213 from strong eQTL effects by implementing an additional step that incorporates genotype data
214 and eQTL effect direction to test which mechanism better explained the observed monoallelic

215 expression (**Methods**). We found that most novel candidates were indeed better explained by
216 strong eQTL effects, including 8 GLOM genes (**Fig. S14**) and 4 TUBE genes (**Fig. S16**). In
217 TUBE, one known imprinted gene, *MEST*, was also better explained by an eQTL (**Fig. S17**).

218 Adding this step reduced the number of identified imprinted genes to 18 in GLOM and
219 13 in TUBE (**Fig 2A**). Thus, we ultimately discovered two novel candidate imprinted genes:
220 *IGLV6-57* (immunoglobulin lambda variable 6-57) in GLOM and *GLYATL2* (glycine-N-
221 acyltransferase like 2) in TUBE. We found many known imprinted genes to be uniquely
222 imprinted in each compartment (7 in GLOM and 2 in TUBE), but the absence of imprinting in
223 the other compartment was largely due to insufficient ASE data (i.e., low expression), rather than
224 observed balanced expression.

225 To test tissue specificity of imprinted GLOM and TUBE genes, we compared them to a
226 list of 299 known human imprinted genes (**Table S6; Fig. S18**), which we curated based on
227 literature review[37,39]. Over 85% of these 299 genes did not have ASE data in GLOM or TUBE
228 due to lack of heterozygous variants, low expression, and/or removal by TOGA. However, most
229 of the known imprinted genes amenable to our ASE-based imprinting analysis still did not
230 demonstrate an imprinting pattern (21 of 38 in GLOM, and 30 of 43 in TUBE). Instead, these
231 genes showed biallelic expression (**Fig. S19**), suggesting loss of imprinting in a tissue-specific
232 manner—consistent with prior studies[37].

233

234 ***ASE is primarily sequence-mediated***

235 Alongside imprinting, ASE is often driven by sequence-mediated mechanisms, such as eQTLs
236 and splicing QTLs (sQTLs)[1,3]. Thus, we sought to classify every ASE event in each individual
237 by identifying heterozygous eQTL/sQTLs associated with the corresponding gene
238 (**Methods**)[38,40]. In the absence of imprinting, eQTL, or sQTL effects, we assigned its mechanism
239 as "unknown." We discovered two major patterns. First, the causal mechanism for rare ASE
240 genes (e.g., ASE in ≤20 GLOM) was overwhelmingly unknown, while sequence-mediated or
241 imprinted mechanism drove most of the common ASE genes (e.g., ASE in >60 GLOM) (**Fig.
242 2B**, **Fig. S20A**). Second, all individuals showed remarkably consistent proportions of their ASE
243 explained by each mechanism (**Fig. 2C, Fig. S20B**). Specifically, the vast majority (mean
244 86.2%) of the commonly shared ASE genes—which have sufficient power for QTL detection—

245 could be classified into the four categories: sQTLs (4.7%), eQTLs (58.4%), sQTLs+eQTLs
246 (14.5%), and imprinting (8.6%) in GLOM. As a representative example, *PLCG2*—one of the
247 most common GLOM ASE genes (**Fig. 1F**)—demonstrated GLOM-specific, sequence-mediated
248 regulation. In GLOM, *PLCG2* exhibited near-monoallelic expression driven by a strong eQTL,
249 rs4243211 (**Fig. S21**). Conversely, in TUBE, *PLCG2* had no significant eQTL and showed
250 balanced expression, underscoring the compartment-specificity of certain regulatory
251 mechanisms.

252 To further validate these findings, we evaluated our ASE data at the single-variant level
253 and assessed the consistency of allelic preference across individuals. We first defined an ASE
254 variant based on: (1) variant-level total RNA coverage ≥ 8, (2) allelic imbalance ≥ 50%, and (3)
255 $P_{adj} < 0.05$ (binomial test). We restricted our analysis to common ASE variants (i.e.,
256 heterozygous genotype in ≥10% of samples, of which ≥50% had ASE) to ensure sufficient power
257 for comparison. For each variant, we calculated the proportion of heterozygous individuals
258 preferring the alternative allele (**Fig. 2D, Fig. S20C)**. Remarkably, most ASE variants showed
259 extreme allelic preference; the majority (e.g., 65.3% in GLOM) demonstrated "perfect allelic
260 preference" (PAP), where the same specific allele (reference or alternative) was consistently
261 preferred in every individual (extreme bars in histograms of **Fig. 2D** and **Fig. S20C**).
262 Conversely, ASE variants within imprinted genes showed random preference, consistent with its
263 parent-of-origin mechanism. We next identified ASE variants that were also significant
264 eQTL/sQTL SNPs for their respective ASE genes, categorizing these intersections as
265 ASE+eQTL (green), ASE+sQTL (blue), or ASE+eQTL+sQTL (red). As expected, variants in
266 these categories were significantly enriched for PAP, accounting for 72.1%, 92.7%, and 84.8%
267 of variants in GLOM, respectively. Moreover, 61.6% of PAP variants fell into one of these three
268 categories, while none were classified as imprinting (**Fig. 2D**). Overall, similar results were
269 found in TUBE (**Fig. S20C**). Finally, we hypothesized that, if ASE is truly driven by eQTLs,
270 most PAP ASE+eQTL variants should show strong concordance in allele preference between the
271 ASE and eQTL analyses. Indeed, 96.2% of such variants showed allelic concordance in GLOM.
272 Altogether, these results strongly suggest that ASE is primarily sequence-mediated.

273 We next hypothesized that genes with unknown mechanism may be driven by rare
274 regulatory promoter variants that eQTL analyses are underpowered to identify. Using the
275 ChromKid model[41], which predicts variant effects on chromatin accessibility in kidney cell

276 types, we analyzed heterozygous promotor variants of all COV genes, in each kidney
277 compartment (**Methods**). We first evaluated promoter variants of balanced (non-ASE) genes,
278 generating a baseline likelihood of identifying predicted regulatory variants in a typical
279 promoter. We then compared ASE genes of known/unknown mechanisms with these balanced
280 genes, calculating an enrichment score (**Fig. 2E**). As expected, genes with an unknown
281 mechanism were significantly enriched for rare (minor allele count, MAC ≤5) predicted
282 regulatory variants, compared to those of known mechanism (**Fig. 2F, Fig. S20D**). This
283 enrichment difference disappeared when including more common (MAC ≤ 50) variants,
284 indicating that rare variants specifically drive unknown ASE. Furthermore, in both
285 mechanistically known and unknown ASE, rare ASE genes (i.e., observed in few individuals)
286 were particularly enriched for these variants, with decreasing enrichment as we included genes
287 with more common ASE (**Fig. 2F, Fig. S20D**). This suggests that while ASE is often driven by
288 rare variants, common ASE genes are driven by a broader variety of mechanisms beyond them.

289

290 ***Splicing events and phasing errors affect gene-level ASE consistency***

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

303 To investigate this, we calculated the deviation of variant-level haplotypic fraction from
304 the corresponding gene-level haplotypic fraction, using root mean square error (RMSE)
305 (**Methods**). We hypothesized that sGenes (genes with variant-driven alternative splicing,

306 captured by sQTL analysis) would have higher RMSE values due to the allelic discordance
307 within these genes. Indeed, sGenes had significantly higher RMSE values than genes with other
308 ASE mechanisms (all comparisons Wilcoxon $P < 0.001$), while imprinted genes had the lowest
309 (**Fig. 3B**). RMSEs for ASE genes with unknown mechanism fall between sGenes and eGenes,
310 suggesting that some ASE events of unknown mechanism are driven by sQTLs not found in our
311 dataset. This was largely true for TUBE samples as well (**Fig. S20E**).

312 Of note, we observed systematically higher RMSE values in genes whose haplotypes
313 were not fully connected by phASER's RNA read-backed phasing (**Fig. 3B)**. This is indicative of
314 "haplotypic flips" arising from errors in population-based DNA phasing. Imprinted genes
315 provide a ground truth to estimate this error rate: as they are expected to have strictly monoallelic
316 expression, the observation of variants with monoallelic expression mapping to opposing
317 haplotypes within the same imprinted gene confirms a phasing error. We used this principle to
318 estimate the prevalence of haplotypic flips in our dataset. Using all our imprinted genes per
319 individual (i.e., imprinted gene-individual pairs), we calculated the proportion in which at least
320 two variants with read depth ≥ 8 demonstrated conflicting near-monoallelic expression (≥90%
321 allelic bias). In GLOM, we tested 666 imprinted gene-individual pairs with a mean of 3.16
322 variants per gene, and identified a haplotypic flip rate of only 3%. Similarly, we tested 400
323 imprinted gene-individuals in TUBE (mean 3.75 variants) and identified a 2.5% rate. This is
324 significantly lower than the previously reported gene-level phasing error rates (~20%) [8],
325 probably due to our enhanced pipeline, further confirming that our gene-level ASE is of high-
326 quality.

327

328 ***Common ASE genes exhibit cell-type specific expression***

329 We next hypothesized that the ASE genes we discovered are transcriptionally coordinated within
330 specific cell types. Using the Kidney Precision Medicine Project (KPMP) single cell RNA-seq
331 dataset[24], we classified 54 cell types by compartment and regrouped them into 12 major
332 categories: two unique to GLOM, six unique to TUBE, and four shared between compartments
333 (**Table S7**). We then performed pseudo-bulk aggregation of single-cell RNA-seq data, followed
334 by normalization, to calculate the proportion of each ASE gene's RNA expression attributable to
335 each cell type within the compartment, termed the "cell-type proportion". Genes with a cell-type

336 proportion exceeding a defined threshold (i.e., 0.33 for GLOM and 0.2 for TUBE) in one cell
337 type relative to others were classified as having "cell-type predominance" Genes whose
338 expression was broadly distributed across all cell types, without predominance in any single
339 type, were classified as "ubiquitous" (**Methods**).

340 We initially analyzed all ASE genes observed in ≥1 individual (**Fig. 4A**). In GLOM,
341 60.1% of ASE genes were ubiquitous; in TUBE, this proportion was 57.4%. However, when we
342 restricted our analysis to more common ASE genes, the proportion of ubiquitous ASE genes
343 decreased while cell-type predominant ASE genes increased. This increase in cell-type
344 predominance was most pronounced for cell types unique to each compartment. For example,
345 among ASE genes observed in at least 30% of individuals, we found increased cell-type
346 predominance for podocytes (POD) and parietal epithelial cells (PEC) in GLOM, and for
347 proximal tubule cells (PT) in TUBE (**Fig. 4A, B**). Compared to all ASE genes, POD-
348 predominant ASE genes increased from 8.4% to 17.2%, and PEC-predominant genes increased
349 from 13.2% to 21.4%, while ubiquitous genes decreased from 60.1% to 35.4%. In TUBE, PT-
350 predominant ASE genes increased from 10.0% to 35.3%, and ubiquitous genes decreased from
351 57.4% to 33.6%. These findings demonstrate that while rare ASE genes are expressed across all
352 cell types, common ASE genes tend to exhibit more compartment- and cell-type-specific
353 expression patterns within GLOM and TUBE.

354

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

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

361 To determine whether this discrepancy was specific to proteinuric disease or a normal
362 phenomenon of kidney subcompartments, we additionally analyzed 31 microdissected "control"
363 kidneys with preserved kidney function and without proteinuric disease (**Methods**). As these
364 control samples had only bulk RNA-seq (without paired WGS), we developed an RNA-only
365 ASE pipeline for direct comparison (**Methods; Fig. S22A**). We note that, in RNA-only ASE

366 analysis, true (near)-monoallelic expression is missed as these sites are assumed to be
367 homozygous, resulting in fewer identified ASE events and no monoallelic expression compared
368 to WGS+RNA pipeline (**Table S4; Fig. S22B**). Nonetheless, results from the RNA-only analysis
369 of NEPTUNE were highly consistent with the WGS+RNA pipeline (**Fig. S22C**), demonstrating
370 significantly higher ASE proportion in GLOM vs TUBE across the cohort (median 7.67% vs.
371 5.27%, $P < 2.2 \times 10^{-16}$) **(Fig. S23A)**. In contrast, the control cohort showed no significant
372 difference (median ASE proportion 6.48% in GLOM vs. 6.96% in TUBE, $P = 0.68$) (**Fig. S23B,**
373 **C**). Due to baseline differences in the NEPTUNE and control cohorts' biology and data (i.e.,
374 read depth, ancestry, age, and number of COV genes), we normalized the ASE proportions by
375 calculating a GLOM-to-TUBE ASE proportion ratio (GTAR). In the RNA-only analysis, the
376 median GTAR in NEPTUNE was significantly higher than the controls (1.40 vs 0.97; P=2.2 ×
377 10^{-6}) (**Fig. 5B**).

378 In light of this difference, we hypothesized that clinical outcomes in proteinuric kidney
379 disease could vary as a function of GTAR. To test this, we dichotomized 194 NEPTUNE
380 participants into a low-GTAR group (GTAR 0.44 – 1.578, similar to most controls) and a high-
381 GTAR group (1.579 – 6.16), based on WGS+RNA analysis, after excluding 12 participants with
382 missing outcome data (**Fig. 5A**). Kaplan-Meier survival analysis revealed that the high-GTAR
383 group had slower kidney disease progression (composite endpoint of ESKD and/or ≥40% decline
384 in estimated glomerular filtration rate—eGFR) compared to the low-GTAR group ($P = 1.57 \times$
385 10^{-3}) (**Fig. 5C**). Cox proportional hazards analysis confirmed this finding: high-GTAR patients
386 were 65% less likely to reach the endpoint after adjusting for multiple kidney disease-related
387 covariates (HR = 0.35, $P = 9.63 \times 10^{-4}$) (**Fig. 5D**). Of note, there was no difference in survival
388 when stratifying solely based on either GLOM or TUBE ASE proportion (**Fig. S24**).

389 We next investigated GTAR's correlation with additional clinical outcomes in
390 NEPTUNE (**Methods**). The high-GTAR group achieved complete remission of proteinuria
391 sooner (Cox adjusted HR = 1.58, $P = 0.023$) (**Fig. S25A**), demonstrating improved renal
392 recovery compared to low-GTAR group. Similarly, high-GTAR individuals showed a trend
393 towards later and less relapse of active nephrotic syndrome, although the difference was not
394 statistically significant, likely due to limited sample size (**Fig. S25B**). In contrast, the low-GTAR
395 group showed non-significantly higher kidney scarring, in the forms of global glomerulosclerosis
396 ($P = 0.083$) and interstitial fibrosis ($P = 0.051$) (**Table S8**). Moreover, six of the seven

397 participants who developed kidney failure were in the low-GTAR group. We found no
398 significant associations with other demographic, clinical, or histologic parameters. Altogether,
399 high GTAR was associated with better kidney outcomes.

400

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

403 We next explored whether differences in gene expression patterns in each GTAR group could be
404 driving variation in both ASE proportion and disease progression. To this end, we performed
405 differential gene expression (DGE) analysis between high- and low-GTAR groups, using GLOM
406 or TUBE RNA-seq and DESeq2[43] (**Methods**). GLOM DGE analysis identified 181 significantly
407 differentially expressed genes (DEGs), with $P_{adj} < 0.05$ and absolute $\log_2$-fold change ($\log_2$FC) >
408 0.5. Of these, 115 genes were upregulated and 66 downregulated in the high-GTAR group (**Fig.**
409 **6A, Table S9**). In contrast, TUBE analysis demonstrated no significant DEGs. Thus, further
410 analysis was focused on the GLOM DGE results.

411 Characterization of GLOM DEGs revealed striking biological patterns. First, a high
412 proportion of DEGs (62 of 181; 34.3%) were long non-coding RNAs (lncRNAs). The vast
413 majority (58 of 62; 93.5%) were upregulated, including 20 antisense to protein-coding genes.
414 Five upregulated lncRNAs were antisense to—thus expected to decrease expression of—genes
415 known to promote kidney disease (e.g., *AQP5*, *LYZ*, *P2RX7*, *CERS6* and *NFAT3C*—with latter
416 four specifically causing renal fibrosis)[44–54]. Second, the high-GTAR group had upregulation of
417 several kidney-protective genes and downregulation of many genes which promote kidney
418 injury—a notable finding given the better kidney clinical outcomes in the high-GTAR group.
419 Specifically, we found upregulation of *FCGR2B* (Fc gamma receptor IIb), an immune inhibitory
420 receptor which suppresses autoimmunity and autoantibody-mediated glomerulonephritis[55–58], and
421 of *NGF* (nerve growth factor), secreted by podocytes and a protective factor against glomerular
422 kidney injury[59–61]. Conversely, downregulated DEGs were strongly enriched for genes whose
423 increased expression contributes to kidney dysfunction. Of the 66 downregulated DEGs, 10
424 (15.2%) promote kidney fibrosis (*FN1*, *CCL19*, *CCL21*, *ADAMTS4, ADAMTS16, CEMIP,*
425 *ITGA11, NEU4* and *HK2*)[62–75] or cardiac/pulmonary fibrosis (*ASPN*)[76,77]. Several other
426 downregulated genes were associated with glomerular disease (e.g., *MRC2*, *ACER2, IL1RAP,*

427 *NEFL*)[78–82]. Finally, 7 of 115 (6%) upregulated DEGs were ribosome-related, with no ribosomal
428 genes among downregulated DEGs—a trend strongly supported by subsequent pathway
429 analyses. Detailed characterization of all DEGs in the context of kidney disease biology is
430 summarized in **Table S10**.

431 We next systematically assessed the 181 GLOM DEGs with over-representation analysis
432 (ORA). For upregulated genes, "ribosome biogenesis in eukaryotes" (KEGG) was the single
433 significant concept ($P_{adj} < 0.05$). For downregulated genes, we identified 57 significant terms
434 (many redundant), including "dendritic cell antigen processing and presentation", "response to
435 corticosterone", "cell adhesion mediated by integrin", "ligand-gated sodium channel activity",
436 "collagen binding" and "immunoglobulin complex" (**Fig. 6B, Table S11**).

437 To comprehensively identify molecular pathways, we also performed gene set
438 enrichment analysis (GSEA) on the full GLOM transcriptome using clusterProfiler[83], ranking all
439 genes by signed($\log_2$FC) × -$\log_{10}$(P) (**Table S11**). Consistent with the ORA results, the top 2
440 upregulated terms (disregarding highly redundant terms) related to ribosome and oxidative
441 phosphorylation in all 5 functional databases (**Fig. 6D**). The most strongly downregulated
442 pathways involved cell adhesion and extra-cellular matrix interaction, interferon signaling,
443 vesicle transport and endocytosis, and aminoacyl-tRNA biosynthesis. In the TUBE DGE
444 analysis, GSEA did not identify any significantly enriched GO terms.

445

446 **Discussion**

447 In this study, we present a comprehensive, genome-wide characterization of allele-specific
448 expression in the human kidney, leveraging a unique dataset of paired WGS and microdissected
449 bulk RNA-seq from the NEPTUNE cohort[30]. By developing and applying a rigorous
450 computational pipeline, we addressed major technical hurdles in ASE analysis, such as
451 overlapping genes and concealed genotyping errors. This allowed us to uncover an accurate
452 landscape of ASE in proteinuric kidney disease. Our primary finding is that, unlike in control
453 tissue, proteinuric kidneys exhibit a significantly higher proportion of genome-wide ASE in
454 GLOM compared to TUBE. We quantified this discrepancy using a new metric, the GLOM-to-
455 TUBE ASE ratio (GTAR), and demonstrated that a high GTAR was strongly associated with
456 improved kidney survival and faster disease remission.

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

468 Our study also provides insights into the molecular mechanisms driving ASE. We
469 demonstrated that most ASE events are individual-specific. However, ASE genes shared across
470 many individuals were predominantly driven by known molecular mechanisms (i.e., eQTLs,
471 sQTLs, and genomic imprinting). More importantly, we provide an explanation for individual-
472 specific ASE, previously speculated as stochastic or random effects. By integrating deep learning
473 models[41], we revealed that genes with an "unknown" mechanism of ASE were significantly
474 enriched for rare, heterozygous predicted regulatory variants in their promoter regions. This
475 finding strongly suggests that a substantial portion of rare, individual-specific ASE may not be
476 stochastic, but rather a deterministic outcome of an individual's unique gene regulatory
477 architecture, which eQTL/sQTL studies are underpowered to detect.

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

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

495 This study has several limitations. First, the use of short-read RNA-seq can fail to resolve
496 haplotypes across long distances, particularly for genes with few heterozygous variants.
497 However, we mitigated this by integrating population-level phasing from SHAPEIT4[32] and RNA
498 read-backed phasing from phASER, which leverages spliced reads to connect distant variants.
499 The high quality of our phasing is supported by our low estimation of gene-level haplotypic flips
500 (2.5-3%) in known imprinted genes. Second, our primary analysis was performed on bulk RNA-
501 seq from microdissected glomerular and tubulointerstitial compartments. While this provides
502 crucial compartmental resolution, it lacks the granularity of single-cell analysis. We began to
503 address this by integrating KPMP single-cell data[24], which revealed that frequently shared ASE
504 genes tend to be cell-type specific. Nonetheless, performing ASE analysis at a true single-cell
505 resolution is needed to fully delineate the cell-specific drivers of ASE in kidney disease. Third,
506 our comparison to the control cohort required using an RNA-only pipeline due to data
507 availability, although the stark difference in GTAR was robust. Lastly, our identification of rare
508 variants driving "unknown" ASE relies on computational predictions, and functional validation
509 of these specific promoter variants is necessary.

510 In conclusion, we provide the first comprehensive landscape of allele-specific expression
511 in kidney tissue in proteinuric disease, revealing widespread transcriptional changes in the
512 glomerular compartments. We establish a new metric, GTAR, which links a higher proportion of
513 glomerular ASE to a protective transcriptional state and improved kidney survival. This study
514 demonstrates that rigorous ASE analysis is a powerful tool for uncovering disease biology and
515 identifying clinically relevant biomarkers in human kidney disease.
516

517 **References**

518 1. Cleary, S. & Seoighe, C. Perspectives on Allele-Specific Expression. *Annu. Rev. Biomed.*
519 *Data Sci.* **4**, 101–122 (2021).
520 2. Pastinen, T. Genome-wide allele-specific analysis: insights into regulatory variation. *Nat.*
521 *Rev. Genet.* **11**, 533–538 (2010).
522 3. Mohammadi, P. *et al.* Genetic regulatory variation in populations informs transcriptome
523 analysis in rare disease. *Science* **366**, 351–356 (2019).
524 4. Knowles, D. A. *et al.* Allele-specific expression reveals interactions between genetic
525 variation and environment. *Nat. Methods* **14**, 699–702 (2017).
526 5. Robles-Espinoza, C. D., Mohammadi, P., Bonilla, X. & Gutierrez-Arcelus, M. Allele-specific
527 expression: applications in cancer and technical considerations. *Curr. Opin. Genet. Dev.* **66**,
528 10–19 (2021).
529 6. Castel, S. E., Mohammadi, P., Chung, W. K., Shen, Y. & Lappalainen, T. Rare variant
530 phasing and haplotypic expression from RNA sequencing with phASER. *Nat. Commun.* **7**,
531 12817 (2016).
532 7. Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools and
533 best practices for data processing in allelic expression analysis. *Genome Biol.* **16**, 195
534 (2015).
535 8. Choi, Y., Chan, A. P., Kirkness, E., Telenti, A. & Schork, N. J. Comparison of phasing
536 strategies for whole human genomes. *PLOS Genet.* **14**, e1007308 (2018).
537 9. van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J. K. WASP: allele-specific software
538 for robust molecular quantitative trait locus discovery. *Nat. Methods* **12**, 1061–1063 (2015).
539 10. Panousis, N. I., Gutierrez-Arcelus, M., Dermitzakis, E. T. & Lappalainen, T. Allelic mapping
540 bias in RNA-sequencing is not a major confounder in eQTL studies. *Genome Biol.* **15**, 467
541 (2014).
542 11. Mattevi, S., Mazzarotto, F. & Martini, P. Allele-specific expression analysis: pipelines,
543 applications, challenges, and unmet needs. *Comput. Biol. Med.* **196**, 110890 (2025).
544 12. Fan, J. *et al.* ASEP: Gene-based detection of allele-specific expression across individuals in a
545 population by RNA sequencing. *PLOS Genet.* **16**, e1008786 (2020).
546 13. Muto, Y. *et al.* Single cell transcriptional and chromatin accessibility profiling redefine
547 cellular heterogeneity in the adult human kidney. *Nat. Commun.* **12**, 2190 (2021).

548 14. Bielski, C. M. *et al.* Widespread Selection for Oncogenic Mutant Allele Imbalance in Cancer.
549 *Cancer Cell* **34**, 852-862.e4 (2018).
550 15. Lee, C., Kang, E. Y., Gandal, M. J., Eskin, E. & Geschwind, D. H. Profiling allele-specific
551 gene expression in brains from individuals with autism spectrum disorder reveals preferential
552 minor allele usage. *Nat. Neurosci.* **22**, 1521–1532 (2019).
553 16. Langmyhr, M. *et al.* Allele-specific expression of Parkinson's disease susceptibility genes in
554 human brain. *Sci. Rep.* **11**, 504 (2021).
555 17. McKean, D. M. *et al.* Loss of RNA expression and allele-specific expression associated with
556 congenital heart disease. *Nat. Commun.* **7**, 12824 (2016).
557 18. Sigurdsson, M. I. *et al.* Allele-specific expression in the human heart and its application to
558 postoperative atrial fibrillation and myocardial ischemia. *Genome Med.* **8**, 127 (2016).
559 19. van Beek, D. *et al.* Allele-specific expression analysis for complex genetic phenotypes
560 applied to a unique dilated cardiomyopathy cohort. *Sci. Rep.* **13**, 564 (2023).
561 20. Liu, H. *et al.* Kidney multiome-based genetic scorecard reveals convergent coding and
562 regulatory variants. *Science* **387**, eadp4753 (2025).
563 21. Castel, S. E. *et al.* A vast resource of allelic expression data spanning human tissues. *Genome*
564 *Biol.* **21**, 234 (2020).
565 22. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human
566 tissues. *Science* **369**, 1318–1330 (2020).
567 23. Garza, F. A. & Leslie, S. W. Anatomy, Abdomen and Pelvis: Kidneys. in *StatPearls*
568 (StatPearls Publishing, Treasure Island (FL), 2025).
569 24. Lake, B. B. *et al.* An atlas of healthy and injured cell states and niches in the human kidney.
570 *Nature* **619**, 585–594 (2023).
571 25. Ogobuiro, I. & Tuma, F. Physiology, Renal. in *StatPearls* (StatPearls Publishing, Treasure
572 Island (FL), 2025).
573 26. Meliambro, K., He, J. C. & Campbell, K. N. Podocyte-targeted therapies - progress and
574 future directions. *Nat. Rev. Nephrol.* **20**, 643–658 (2024).
575 27. Mann, N., Sun, H. & Majmundar, A. J. Mechanisms of podocyte injury in genetic kidney
576 disease. *Pediatr. Nephrol. Berl. Ger.* **40**, 1523–1538 (2025).
577 28. Groopman, E. E. *et al.* Diagnostic Utility of Exome Sequencing for Kidney Disease. *N. Engl.*
578 *J. Med.* **380**, 142–151 (2019).

579 29. Connaughton, D. M. *et al.* Monogenic causes of chronic kidney disease in adults. *Kidney Int.*
580 **95**, 914–928 (2019).
581 30. Gadegbeku, C. A. *et al.* Design of the Nephrotic Syndrome Study Network (NEPTUNE) to
582 evaluate primary glomerular nephropathy by a multidisciplinary approach. *Kidney Int.* **83**,
583 749–756 (2013).
584 31. Martin, M. *et al. WhatsHap: Fast and Accurate Read-Based Phasing*.
585 http://biorxiv.org/lookup/doi/10.1101/085050 (2016) doi:10.1101/085050.
586 32. Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L. & Dermitzakis, E. T. Accurate,
587 scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).
588 33. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.
589 *Nature* **562**, 203–209 (2018).
590 34. Jun, G. *et al.* Detecting and estimating contamination of human DNA samples in sequencing
591 and array-based genotype data. *Am. J. Hum. Genet.* **91**, 839–848 (2012).
592 35. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–
593 21 (2013).
594 36. Asiimwe, R. & Alexander, D. STAR+WASP reduces reference bias in the allele-specific
595 mapping of RNA-seq reads. *BioRxiv Prepr. Serv. Biol.* 2024.01.21.576391 (2024)
596 doi:10.1101/2024.01.21.576391.
597 37. Baran, Y. *et al.* The landscape of genomic imprinting across diverse adult human tissues.
598 *Genome Res.* **25**, 927–936 (2015).
599 38. Han, S. K. *et al.* Mapping genomic regulation of kidney disease and traits through high-
600 resolution and interpretable eQTLs. *Nat. Commun.* **14**, 2229 (2023).
601 39. Tucci, V., Isles, A. R., Kelsey, G., Ferguson-Smith, A. C., & Erice Imprinting Group.
602 Genomic Imprinting and Physiological Processes in Mammals. *Cell* **176**, 952–965 (2019).
603 40. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human
604 tissues. *Science* **369**, 1318–1330 (2020).
605 41. Loeb, G. B. *et al.* Variants in tubule epithelial regulatory elements mediate most heritable
606 differences in human kidney function. *Nat. Genet.* **56**, 2078–2092 (2024).
607 42. Nembaware, V. *et al.* Genome-wide survey of allele-specific splicing in humans. *BMC*
608 *Genomics* **9**, 265 (2008).

609 43. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
610 RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
611 44. Rossi, L. *et al.* Urinary Excretion of Kidney Aquaporins as Possible Diagnostic Biomarker of
612 Diabetic Nephropathy. *J. Diabetes Res.* **2017**, 4360357 (2017).
613 45. Gao, C. & Zhang, W. Urinary AQP5 is independently associated with eGFR decline in
614 patients with type 2 diabetes and nephropathy. *Diabetes Res. Clin. Pract.* **155**, 107805
615 (2019).
616 46. Ren, Y., Yu, M., Zheng, D., He, W. & Jin, J. Lysozyme promotes renal fibrosis through the
617 JAK/STAT3 signal pathway in diabetic nephropathy. *Arch. Med. Sci. AMS* **20**, 233–247
618 (2024).
619 47. Lee, Y. H. *et al.* Urinary mRNA Signatures as Predictors of Renal Function Decline in
620 Patients With Biopsy-Proven Diabetic Kidney Disease. *Front. Endocrinol.* **12**, 774436
621 (2021).
622 48. Paueksakon, P., Najafian, B., Alpers, C. E. & Fogo, A. B. AJKD Atlas of Renal Pathology:
623 Lysozyme-Associated Nephropathy. *Am. J. Kidney Dis. Off. J. Natl. Kidney Found.* **86**, e1–
624 e3 (2025).
625 49. Menzies, R. I., Tam, F. W., Unwin, R. J. & Bailey, M. A. Purinergic signaling in kidney
626 disease. *Kidney Int.* **91**, 315–323 (2017).
627 50. Taylor, S. R. J. *et al.* P2X7 deficiency attenuates renal injury in experimental
628 glomerulonephritis. *J. Am. Soc. Nephrol. JASN* **20**, 1275–1281 (2009).
629 51. Turner, C. M. *et al.* Increased expression of the pro-apoptotic ATP-sensitive P2X7 receptor in
630 experimental and human glomerulonephritis. *Nephrol. Dial. Transplant. Off. Publ. Eur. Dial.*
631 *Transpl. Assoc. - Eur. Ren. Assoc.* **22**, 386–395 (2007).
632 52. Zhu, Z. *et al.* CerS6 links ceramide metabolism to innate immune responses in diabetic
633 kidney disease. *Nat. Commun.* **16**, 1528 (2025).
634 53. Wang, X. *et al.* CERS6-derived ceramides aggravate kidney fibrosis by inhibiting PINK1-
635 mediated mitophagy in diabetic kidney disease. *Am. J. Physiol. Cell Physiol.* **325**, C538–
636 C549 (2023).
637 54. Guo, Y. *et al.* Modulation of podocyte extracellular matrix remodeling in membranous
638 nephropathy by the NFATc3/LRRC55/BK channel pathway. *J. Cell Commun. Signal.* **19**,
639 e70022 (2025).

640 55. Ooi, J. D. *et al.* FcγRIIB regulates T-cell autoreactivity, ANCA production, and neutrophil
641 activation to suppress anti-myeloperoxidase glomerulonephritis. *Kidney Int.* **86**, 1140–1149
642 (2014).
643 56. Sharp, P. E. H. *et al.* FcγRIIb on myeloid cells and intrinsic renal cells rather than B cells
644 protects from nephrotoxic nephritis. *J. Immunol. Baltim. Md 1950* **190**, 340–348 (2013).
645 57. Niederer, H. A., Clatworthy, M. R., Willcocks, L. C. & Smith, K. G. C. FcgammaRIIB,
646 FcgammaRIIIB, and systemic lupus erythematosus. *Ann. N. Y. Acad. Sci.* **1183**, 69–88
647 (2010).
648 58. Sharp, P. E. H. *et al.* Increased incidence of anti-GBM disease in Fcgamma receptor 2b
649 deficient mice, but not mice with conditional deletion of Fcgr2b on either B cells or myeloid
650 cells alone. *Mol. Immunol.* **50**, 49–56 (2012).
651 59. Caroleo, M. C. *et al.* Human kidney podocyte cell population as a novel biological target of
652 nerve growth factor. *Growth Factors Chur Switz.* **33**, 14–22 (2015).
653 60. Alpers, C. E. *et al.* Nerve growth factor receptor expression in fetal, mature, and diseased
654 human kidneys. *Lab. Investig. J. Tech. Methods Pathol.* **69**, 703–713 (1993).
655 61. Aloe, L., Rossi, S. & Manni, L. Altered expression of nerve growth factor and its receptors in
656 the kidneys of diabetic rats. *J. Nephrol.* **24**, 798–805 (2011).
657 62. Liu, Y. Cellular and molecular mechanisms of renal fibrosis. *Nat. Rev. Nephrol.* **7**, 684–696
658 (2011).
659 63. Bowers, S. L. K. *et al.* Inhibition of fibronectin polymerization alleviates kidney injury due
660 to ischemia-reperfusion. *Am. J. Physiol. Renal Physiol.* **316**, F1293–F1298 (2019).
661 64. Chen, H. *et al.* Identification of CCL19 as a Novel Immune-Related Biomarker in Diabetic
662 Nephropathy. *Front. Genet.* **13**, 830437 (2022).
663 65. Sato, Y. & Yanagita, M. Resident fibroblasts in the kidney: a major driver of fibrosis and
664 inflammation. *Inflamm. Regen.* **37**, 17 (2017).
665 66. Chen, D. *et al.* Single-cell RNA-seq with spatial transcriptomics to create an atlas of human
666 diabetic kidney disease. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **37**, e22938 (2023).
667 67. Wada, T., Sakai, N., Matsushima, K. & Kaneko, S. Fibrocytes: a new insight into kidney
668 fibrosis. *Kidney Int.* **72**, 269–273 (2007).

68. Sakai, N. *et al.* Secondary lymphoid tissue chemokine (SLC/CCL21)/CCR7 signaling regulates fibrocytes in renal fibrosis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 14098–14103 (2006).
69. Zhou, H.-L., Wang, Y.-T., Gao, T., Wang, W.-G. & Wang, Y.-S. Distribution and expression of fibroblast-specific protein chemokine CCL21 and chemokine receptor CCR7 in renal allografts. *Transplant. Proc.* **45**, 538–545 (2013).
70. Vojtusek, I. K. *et al.* First Characterization of ADAMTS-4 in Kidney Tissue and Plasma of Patients with Chronic Kidney Disease-A Potential Novel Diagnostic Indicator. *Diagn. Basel Switz.* **12**, 648 (2022).
71. Zhao, J. *et al.* A novel role of ADAMTS16 in renal fibrosis through activating TGF-β/Smad signaling. *Cell. Signal.* **122**, 111347 (2024).
72. Chen, C.-H., Ke, G.-M., Lin, P.-C. & Lin, K.-D. 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).
73. Bansal, R. *et al.* Integrin alpha 11 in the regulation of the myofibroblast phenotype: implications for fibrotic diseases. *Exp. Mol. Med.* **49**, e396 (2017).
74. Xiao, P.-T. *et al.* Targeting Neuraminidase 4 Attenuates Kidney Fibrosis in Mice. *Adv. Sci. Weinh. Baden-Wurtt. Ger.* **11**, e2406936 (2024).
75. Zhang, Y. *et al.* Activation of HIF-1α C-terminal transactivation domain promotes tubulointerstitial fibrosis through hexokinase 2-mediated metabolic reprogramming. *Cell. Signal.* **127**, 111531 (2025).
76. Medzikovic, L. *et al.* Myocardial fibrosis and calcification are attenuated by microRNA-129-5p targeting Asporin and Sox9 in cardiac fibroblasts. *JCI Insight* **8**, e168655 (2023).
77. Huang, S. *et al.* 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).
78. Lopez-Silva, C. *et al.* Circulating Protein and Metabolite Correlates of Histologically Confirmed Diabetic Kidney Disease. *Kidney Med.* **6**, 100920 (2024).
79. Sas, K. M. *et al.* 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).

80. Parikh, S. V. *et al.* Molecular imaging of the kidney in lupus nephritis to characterize response to treatment. *Transl. Res. J. Lab. Clin. Med.* **182**, 1–13 (2017).
81. Lu, J.-D. *et al.* 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).
82. van der Plas, E. *et al.* Associations between neurofilament light-chain protein, brain structure, and chronic kidney disease. *Pediatr. Res.* **91**, 1735–1740 (2022).
83. Xu, S. *et al.* Using clusterProfiler to characterize multiomics data. *Nat. Protoc.* **19**, 3292–3320 (2024).

Acknowledgments:

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

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

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

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

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

NIMH, and NINDS.

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

Funding:

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

Author contributions:

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

768 **Competing interests:**

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

771

772 **Data and materials availability:**

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

783

784 **Supplementary Materials**

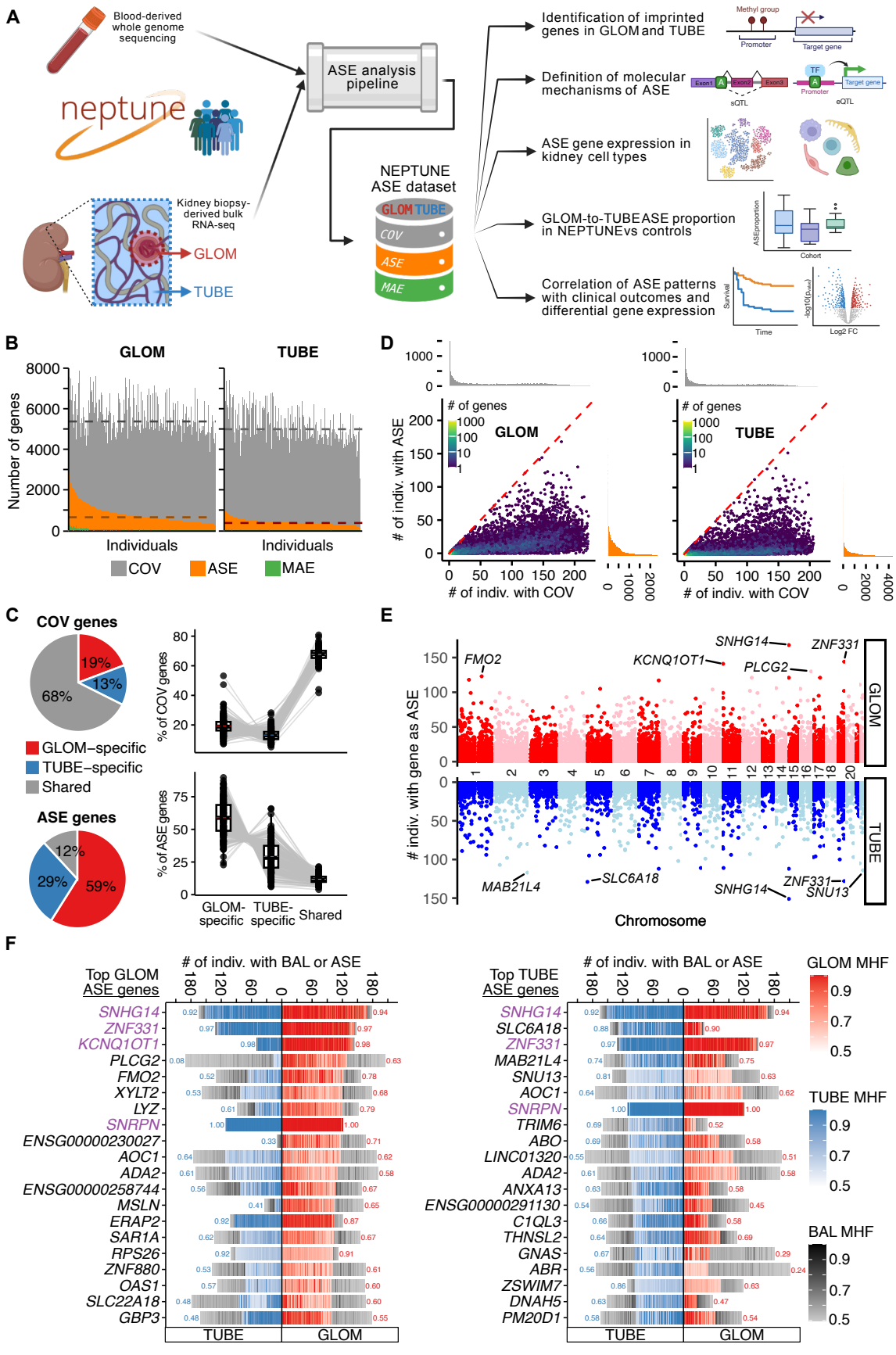
785 Methods

786 Figs. S1 to S25

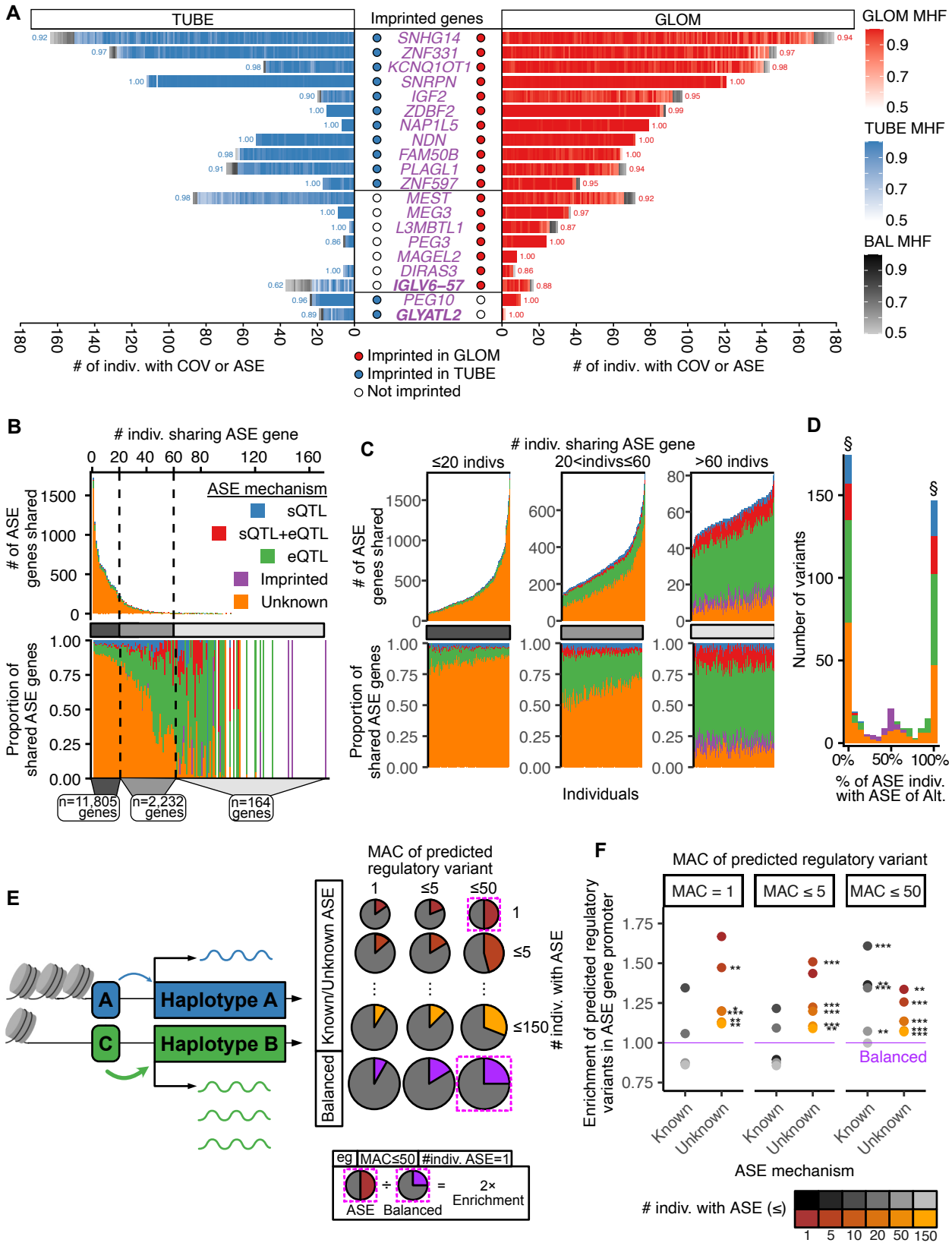
787 Tables S1 to S11

788 Members of the Nephrotic Syndrome Study Network

789

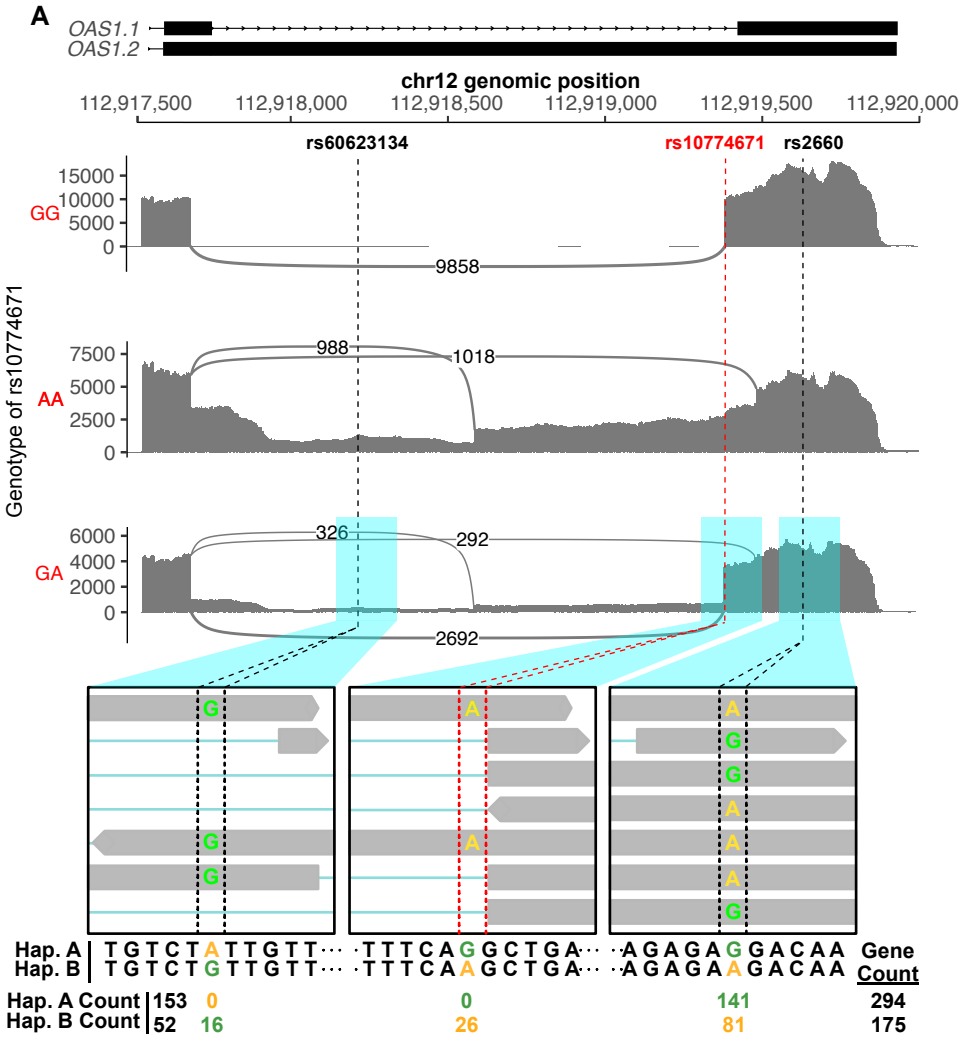

A
Blood-derived whole genome sequencing
neptune
ASE analysis pipeline
Kidney biopsy-derived bulk RNA-seq
GLOM
TUBE
NEPTUNE ASE dataset
GLOM TUBE
COV
ASE
MAE
Identification of imprinted genes in GLOM and TUBE
Methyl group
Promoter
Target gene
Definition of molecular mechanisms of ASE
Exon1
Exon2
Exon3
sQTL
TF
eQTL
ASE gene expression in kidney cell types
GLOM-to-TUBE ASE proportion in NEPTUNE vs controls
ASEproportion
Cohort
Correlation of ASE patterns with clinical outcomes and differential gene expression
Survival
Time
Log2 FC
B
GLOM
TUBE
Number of genes
Individuals
COV
ASE
MAE
C
COV genes
19%
13%
68%
GLOM–specific
TUBE–specific
Shared
ASE genes
12%
29%
59%
% of COV genes
% of ASE genes
GLOM-specific
TUBE-specific
Shared
D
# of genes
GLOM
TUBE
# of indiv. with ASE
# of indiv. with COV
E
FMO2
KCNQ1OT1
SNHG14
PLCG2
ZNF331
MAB21L4
SLC6A18
SNU13
GLOM
TUBE
indiv. with gene as ASE
Chromosome
F
Top GLOM ASE genes
# of indiv. with BAL or ASE
SNHG14
ZNF331
KCNQ1OT1
PLCG2
FMO2
XYLT2
LYZ
SNRPN
ENSG00000230027
AOC1
ADA2
ENSG00000258744
MSLN
ERAP2
SAR1A
RPS26
ZNF880
OAS1
SLC22A18
GBP3
TUBE
GLOM
Top TUBE ASE genes
SNHG14
SLC6A18
ZNF331
MAB21L4
SNU13
AOC1
SNRPN
TRIM6
ABO
LINC01320
ADA2
ANXA13
ENSG00000291130
C1QL3
THNSL2
GNAS
ABR
ZSWIM7
DNAH5
PM20D1
GLOM MHF
TUBE MHF
BAL MHF

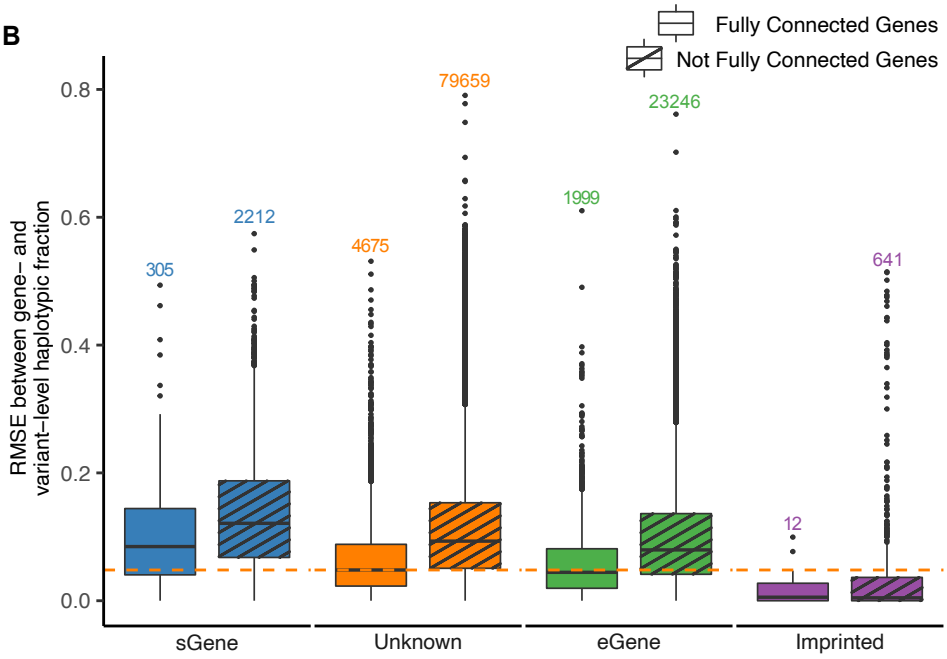
791 **Fig. 1. Study design and overview of ASE in kidney glomerular and tubulointerstitial**
792 **compartments in NEPTUNE. (A)** Schematic overview of our study. (**B**) Individual-level data:
793 Each bar represents one individual, ordered by number of ASE genes; stacked bars are
794 cumulative. Dashed lines indicate the median of COV (grey) and ASE (orange) genes per
795 sample. (**C**) Overlap of COV (top) and ASE (bottom) genes between GLOM and TUBE in
796 paired samples (n=206). Pie charts show the mean proportions of shared and compartment-
797 specific genes across all individuals. Boxplots show the distribution of these proportions. (**D**)
798 Gene-level data: each dot represents a gene, with the number of individuals of this gene as COV
799 (x-axis), compared to that of ASE (y-axis). The heatmap represents point density to indicate
800 overlapping data points. Marginal histograms are shown on both axes. (**E**) The number of
801 individuals sharing an ASE gene (y-axis) is plotted (GLOM: top, TUBE: bottom) against its
802 genomic coordinate (x-axis). The five most common ASE genes are labeled. (**F**) The top 20 ASE
803 genes in GLOM (left) and TUBE (right), ordered by number of ASE individuals of this gene.
804 The horizontal axis shows the number of ASE and BAL individuals in GLOM or TUBE.
805 Individual stacked bars have color grading determined by major haplotypic fraction (MHF) and
806 sorted by the binomial test *P* value. Imprinted genes (from our imprinting analysis, Fig. 2A) are
807 indicated in purple. The proportion of individuals with ASE among those with COV is shown as
808 the number adjacent to each bar. Indiv, individual.


A
TUBE
Imprinted genes
GLOM
GLOM MHF
TUBE MHF
BAL MHF
SNHG14
ZNF331
KCNQ1OT1
SNRPN
IGF2
ZDBF2
NAP1L5
NDN
FAM50B
PLAGL1
ZNF597
MEST
MEG3
L3MBTL1
PEG3
MAGEL2
DIRAS3
IGLV6-57
PEG10
GLYATL2
# of indiv. with COV or ASE
Imprinted in GLOM
Imprinted in TUBE
Not imprinted
B
indiv. sharing ASE gene
ASE mechanism
sQTL
sQTL+eQTL
eQTL
Imprinted
Unknown
# of ASE genes shared
Proportion of shared ASE genes
n=11,805 genes
n=2,232 genes
n=164 genes
C
indiv. sharing ASE gene
≤20 indivs
20<indivs≤60
>60 indivs
Individuals
D
Number of variants
% of ASE indiv. with ASE of Alt.
E
Haplotype A
Haplotype B
MAC of predicted regulatory variant
Known/Unknown ASE
Balanced
indiv. with ASE
eg MAC≤50 #indiv. ASE=1
ASE Balanced
2× Enrichment
F
MAC of predicted regulatory variant
MAC = 1
MAC ≤ 5
MAC ≤ 50
Enrichment of predicted regulatory variants in ASE gene promoter
Balanced
Known
Unknown
ASE mechanism
indiv. with ASE (≤)

809

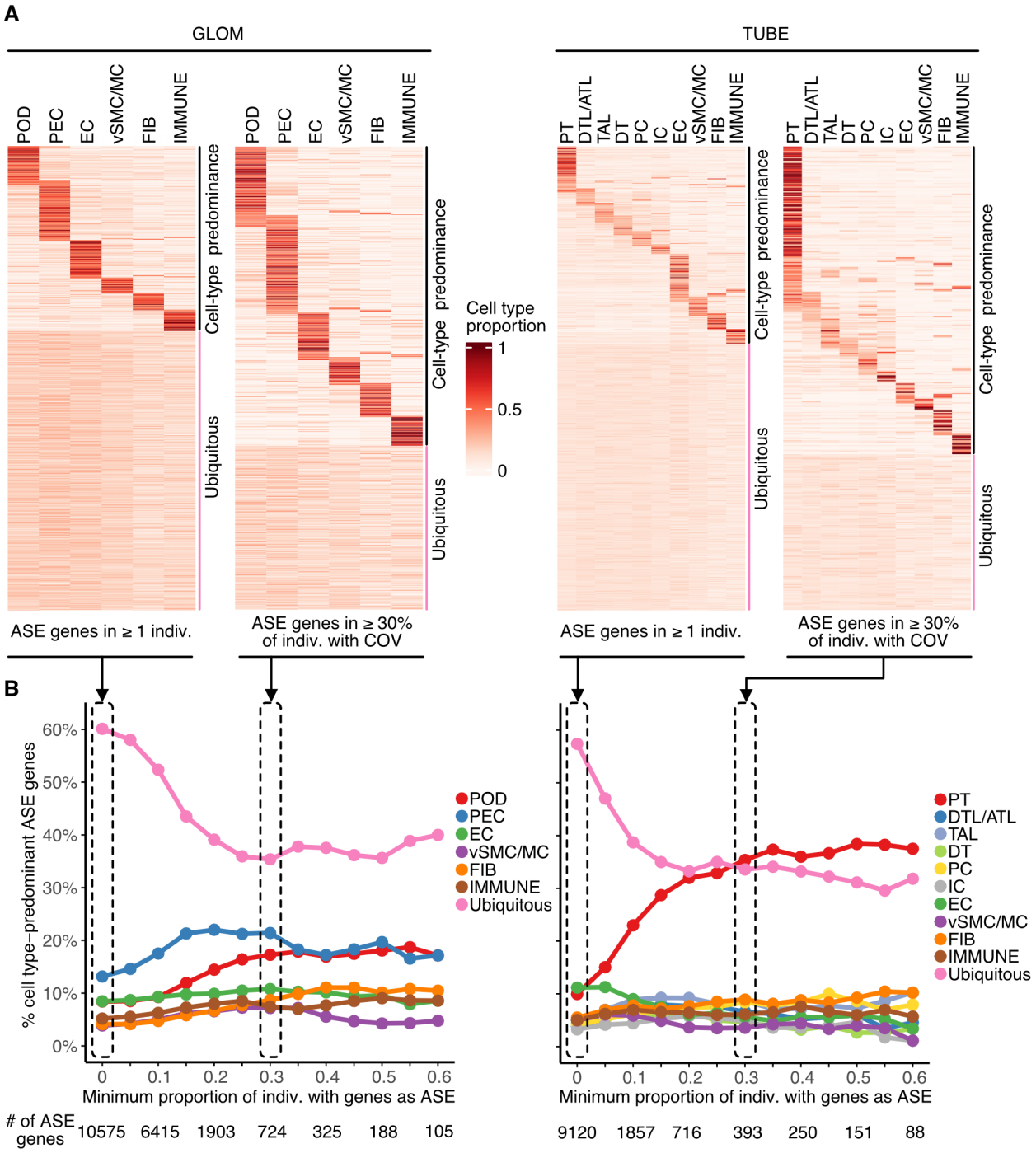
810 **Fig. 2. Molecular mechanisms of ASE in NEPTUNE. (A)** The 20 imprinted genes in GLOM
811 and/or TUBE, identified by our imprinting analysis. Similar to **Fig. 1F**, bars show the number of
812 individuals with BAL or ASE in GLOM or TUBE, with individual stacked bars color graded by
813 MHF. Circles indicate predicted imprinted status of each gene in the respective compartment,
814 with novel genes shown in bold. **(B)** The top panel is a distribution of the ASE gene frequency
815 across the GLOM samples, stratified and colored by their molecular mechanisms, while the
816 bottom panel shows their proportions. Dashed lines split ASE genes into three groups by number
817 of individuals sharing the ASE gene. The total gene count per group is shown below. **(C)** Each
818 bar represents one individual, stratified by the three ASE gene groups defined in (B). The top
819 panels show numbers of ASE genes, colored by each molecular mechanism, while the bottom
820 panels show their proportions. **(D)** A distribution of the alternative allele preference of ASE
821 variants in GLOM, stratified and color coded by their molecular mechanisms. § indicates perfect
822 allelic preference (PAP) of Ref. (0%) or Alt. (100%). **(E)** Schematic of the promoter variant
823 analysis. **(F)** Enrichment of predicted promoter regulatory variants of ASE gene-individual pairs,
824 comparing known and unknown ASE mechanisms to balanced gene-individual pairs in GLOM.
825 Each dot represents a group of ASE gene-individual pairs, defined by their mechanism
826 (known/unknown), ASE frequency, and the minor allele count (MAC) of the predicted
827 regulatory variants. The y-axis shows enrichment relative to balanced genes (purple line = 1.0).
828 The significance was determined by a one-sided binomial test of the observed count against the
829 expected probability, defined by the proportion of balanced gene-individual pairs with predicted
830 regulatory variants at the corresponding MAC. (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ ***).


A
OAS1.1
OAS1.2
chr12 genomic position
112,917,500
112,918,000
112,918,500
112,919,000
112,919,500
112,920,000
rs60623134
rs10774671
rs2660
Genotype of rs10774671
GG
AA
GA
9858
988
1018
326
292
2692
Hap. A
Hap. B
Hap. A Count
Hap. B Count
Gene Count
294
175


B
Fully Connected Genes
Not Fully Connected Genes
RMSE between gene– and variant–level haplotypic fraction
305
2212
4675
79659
1999
23246
12
641
sGene
Unknown
eGene
Imprinted

831

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



850

851 **Fig. 4. Cell-type specificity of ASE genes in NEPTUNE kidney compartments.** (**A**) Heatmaps
852 of cell-type proportions for all ASE genes (ASE observed in ≥1 individuals) and common ASE
853 genes (ASE observed in ≥30% individuals with COV) in GLOM (left) and TUBE (right). Rows
854 represent individual ASE genes; columns represent cell types; color scale indicates cell-type
855 proportion. The right panel of each heatmap indicates whether ASE genes are "cell-type

856 predominance" or "ubiquitous". All analyses were restricted to genes with COV in ≥10%
857 individuals. **(B)** Trends in cell-type predominance of ASE genes in GLOM (left) and TUBE
858 (right) as a function of ASE frequency. The x-axis is the minimum proportion of individuals with
859 ASE among individuals with COV, and the y-axis is the percentage of ASE genes dominantly
860 expressed in a given cell type. Colors represent cell types in each compartment, and numbers
861 below the x-axis denote the total number of ASE genes meeting the corresponding threshold.
862 The dotted black box corresponds to the heatmap in (A). Unique GLOM cell types are podocytes
863 (POD) and parietal epithelial cells (PEC). Unique TUBE cell types are proximal tubule cells
864 (PT), intercalated cells (IC), distal convoluted tubule cells (DT), thick ascending limb cells
865 (TAL), descending/ascending thin limb cells (DTL/ATL), and principal cells (PC). Common cell
866 types include endothelial cells (EC), vascular smooth muscle/mesangial cells (vSMC/MC),
867 fibroblasts (FIB), and immune cells (IMMUNE).

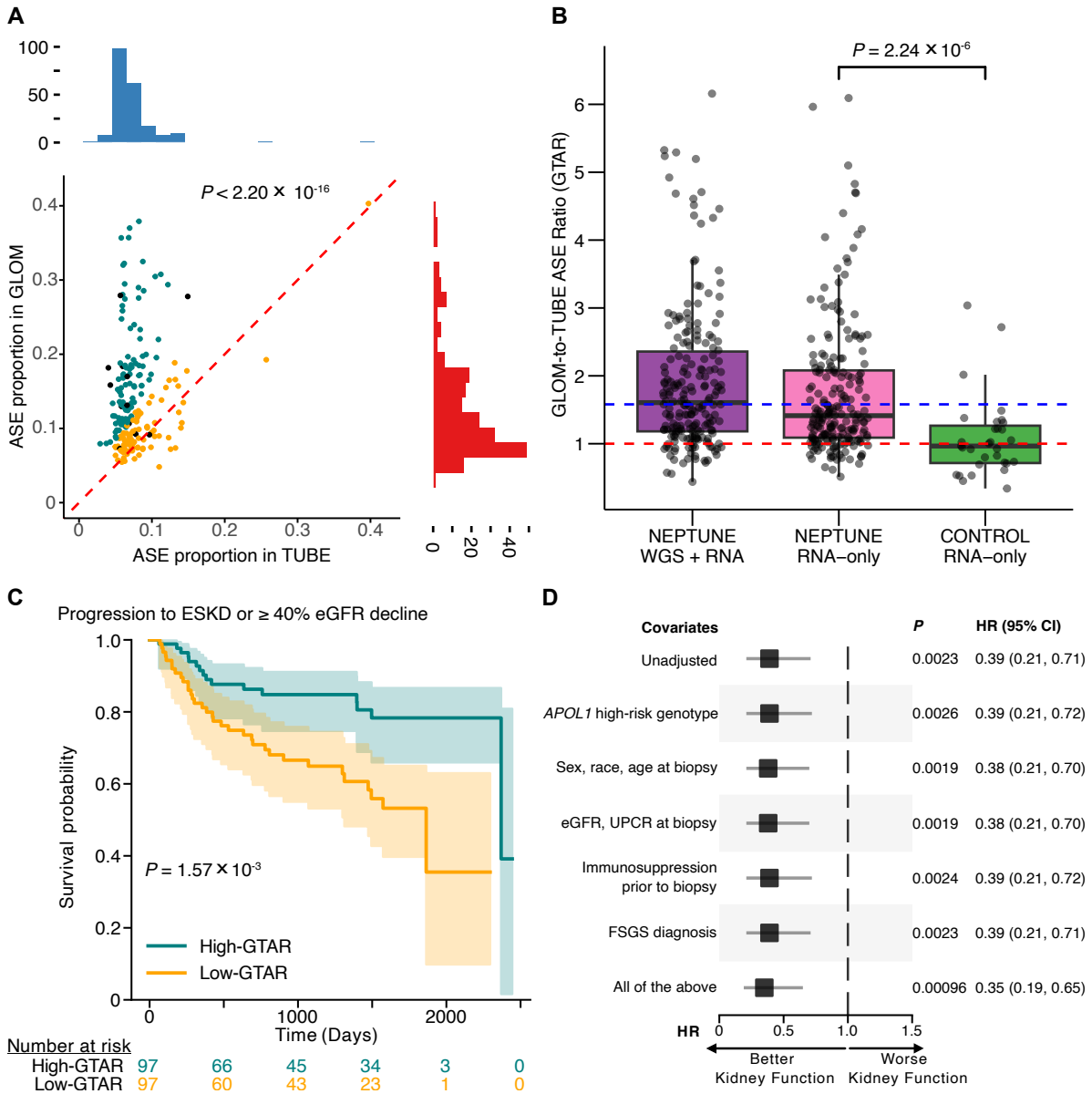
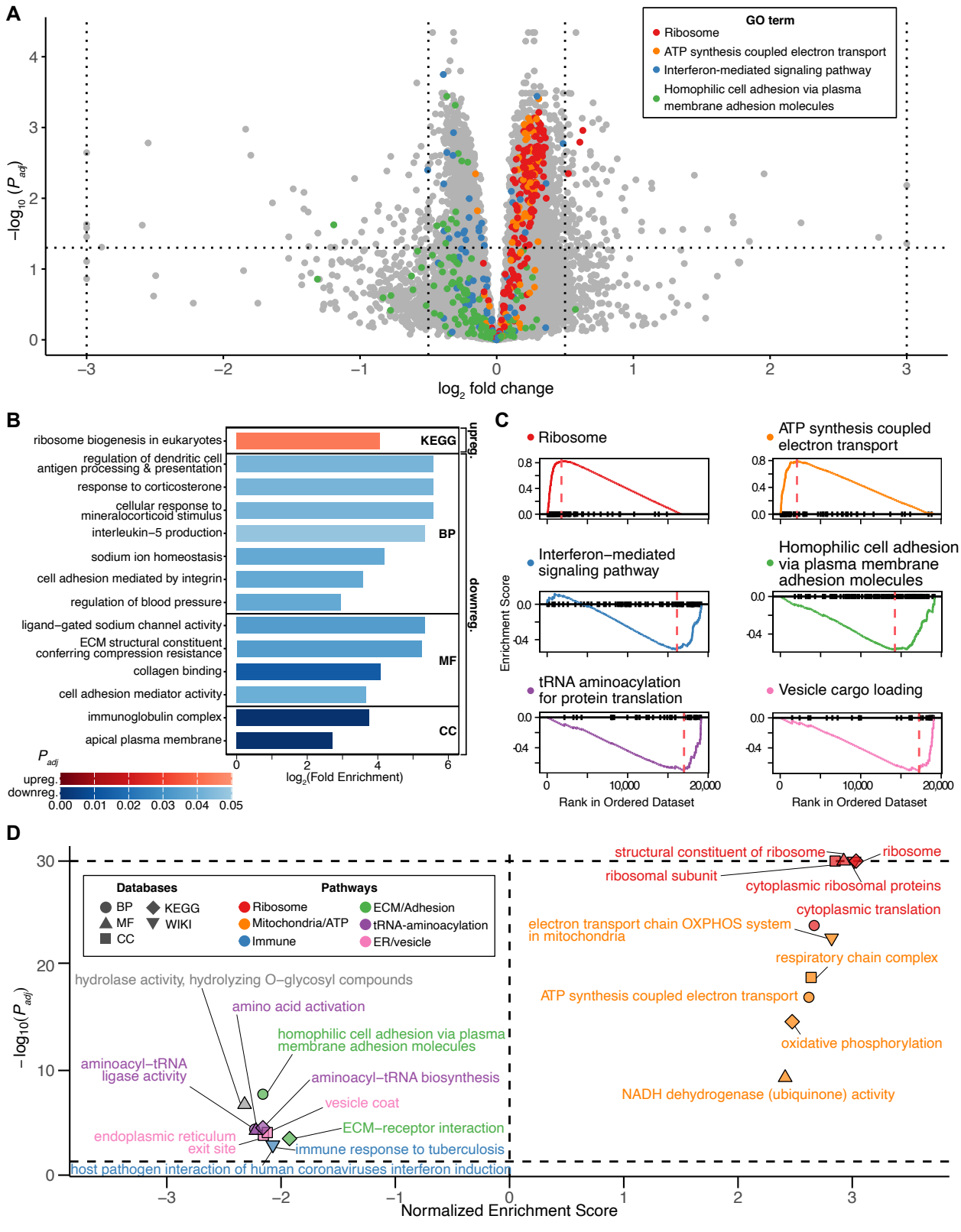


Fig. 5. Genome-wide ASE proportion in GLOM/TUBE in NEPTUNE versus controls, and correlation of GTAR with clinical outcomes in NEPTUNE participants. (**A**) Scatterplot of genome-wide ASE proportion (ASE/COV per sample) in paired GLOM and TUBE samples. Each dot is a NEPTUNE individual with both GLOM and 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 yellow. Samples without endpoint data are in black (n = 12). Marginal histograms illustrate the distribution of genome-wide ASE proportion in each compartment. The Wilcoxon

877 signed-rank test *P* is shown to compare GLOM and TUBE proportions. The red dashed line
878 indicates equal GLOM and TUBE ASE proportions. (**B**) GLOM-to-TUBE ASE ratio (GTAR)
879 per individual, stratified by cohort and ASE analysis pipeline. Each dot is an individual with
880 paired GLOM and TUBE data (NEPTUNE n=206, control n=31). The blue dashed line
881 represents the median GTAR for NEPTUNE (WGS + RNA pipeline), and the red dashed line
882 indicates GTAR = 1. **(C)** Kaplan–Meier survival analysis of NEPTUNE individuals (n = 194),
883 comparing the high- and low-GTAR groups. Endpoint is composite of ESKD or ≥ 40% eGFR
884 decline. **(D)** Forest plot of Cox proportional hazards model, with hazard ratio (HR) in reference
885 to high-GTAR group, and adjusted for multiple kidney disease-related covariates. ESKD, end-
886 stage kidney disease; eGFR, estimated glomerular filtration rate; UPCR, urine protein-to-
887 creatinine ratio; FSGS, focal segmental glomerulosclerosis.


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

889 **Fig. 6. Differential glomerular gene expression between high- and low-GTAR groups and**
890 **gene set enrichment analyses (GSEA).** (**A**) Volcano plot of differentially expressed genes
891 (DEGs) with the low GTAR group as reference. The horizontal dashed line indicates P_{adj}
892 threshold (0.05), and the inner and outer vertical dashed lines indicate $\log_2$(fold change) cutoffs
893 of ±0.5 and ±3, respectively. Genes exceeding ±3 are aligned with the outer vertical lines.
894 Significantly upregulated GO terms related to ribosome and mitochondrial/ATP synthesis (red
895 and orange) and downregulated immune-related and membrane-associated pathways (blue and
896 green) are highlighted. The corresponding GO terms are shown in *(C)* with matching colors. (**B**)
897 ORA of significantly upregulated ($\log_2$FC > 0.5, P_{adj} < 0.05) and downregulated ($\log_2$FC < –0.5,
898 P_{adj} < 0.05) genes across multiple databases. Bars color scales reflect P_{adj} from ORA
899 (upregulated: red, downregulated: blue). (**C**) GSEA enrichment plots of significant GO terms,
900 grouped into six functional categories: ribosome (red), mitochondria/ATP (orange), immune
901 (blue), ECM/adhesion (green), tRNA-aminoacylation (purple), and ER/vesicle (pink). (**D**) GSEA
902 summary of the top upregulated and downregulated GO terms ranked by normalized enrichment
903 score across multiple databases. Terms with $-\log_{10}(P_{adj})$ > 30 are truncated and aligned at the top
904 horizontal dashed line. The bottom dashed line indicates the threshold for statistical significance
905 (P_{adj} = 0.05). The vertical dashed line shows where there is no enrichment. Colors correspond to
906 the six categories in *(C)* and grey color denotes GO terms not included in these categories.
907 Shapes indicate the database source.