Distinct molecular etiologies of male and female hepatocellular carcinoma

Heini M. Natri[1]*, Melissa A. Wilson[1], Kenneth Buetow[1]

[1] Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, Arizona, USA

Corresponding author:

E-mail: hnatri@asu.edu (H.M.N)

Abstract

Sex-differences in cancer occurrence and mortality are evident across tumor types; men exhibit higher rates of incidence and often poorer responses to treatment. Targeted approaches to the treatment of tumors that account for these sex-differences require the characterization and understanding of the fundamental biological mechanisms that differentiate them. Hepatocellular Carcinoma (HCC) is the second leading cause of cancer death worldwide, with the incidence rapidly rising. HCC exhibits a male-bias in occurrence and mortality, but previous studies have failed to explore the sex-specific dysregulation of gene expression in HCC. Here, we characterize the sex-shared and sex-specific regulatory changes in HCC tumors in the TCGA LIHC cohort. By using a sex-specific differential expression analysis of tumor and tumor-adjacent samples, we uncovered etiologically relevant genes and pathways differentiating male

and female HCC. While both sexes exhibited activation of pathways related to apoptosis and cell cycle, males and females differed in the activation of several signaling pathways, with females showing PPAR pathway enrichment while males showed PI3K, 305 PI3K/AKT, FGFR, EGFR, NGF, GF1R, Rap1, DAP12, and IL-2 signaling pathway enrichment. Using eQTL analyses, we discovered germline variants with differential effects on tumor gene expression between the sexes. 24.3% of the discovered eQTLs exhibit differential effects between the sexes, illustrating the substantial role of sex in modifying the effects of eQTLs in HCC. The genes that showed sex-specific dysregulation in tumors and those that harbored a sex-specific eQTL converge in clinically relevant pathways, suggesting that the molecular etiologies of male and female HCC are partially driven by differential genetic effects on gene expression. Overall, our results provide new insight into the role of inherited genetic regulation of transcription in modulating sex-differences in HCC etiology and provide a framework for future studies on sex-biased cancers.

Background

Differences in cancer occurrence and mortality between sexes are evident across tumor types; males exhibit higher rates of cancer incidence and often poorer response to treatment, including some forms of chemotherapy and immunotherapy [1,2]. While differences in risk factors may explain some portion of the sex-bias, the bias remains after appropriate adjustment for these factors [3,4]. A recent study examining the mutational profiles of tumors from males and females across The Cancer Genome Atlas (TCGA) found sex-differences in mutational profiles, calling for the consideration of sex as a biological variable in studies on cancer occurrence, etiology, and treatment [5]. Despite these underlying molecular differences, sex is rarely considered in the development of cancer therapies.

Across tumor types analyzed, the largest sex-differences in autosomal mutational profiles were seen in liver hepatocellular carcinoma (HCC), indicating that male and female HCC are etiologically distinct [5]. Furthermore, HCC exhibits sex-bias in occurrence, with a male-to-female incidence ratio between 1.3:1

45 and 5.5:1 across populations [6,7]. The sexes also differ in the clinical manifestation of HCC, males
46 exhibiting an earlier onset and more/larger nodules [8]. HCC is the second leading cause of cancer
47 mortality worldwide, accounting for 8.2% of all cancer deaths [2], and the incidence in the US has
48 doubled in the last 3 decades, attributable to increased rates of obesity [7], calling for the development of
49 new interventions and targeted therapies.

50 Sex-specific gene regulation may partially underlie differences between the sexes in disease prevalence
51 and severity [9,10]. Previous work observed extensive sex-biased signatures in gene expression in HCC
52 and other sex-biased cancers [11]. However, this study focused solely on comparing male and female
53 tumor samples, without consideration sex-differences in non-diseased and tumor-adjacent tissues. To
54 understand cancer-specific processes, it is necessary to contrast the sex-differences in gene expression
55 identified in HCC with those in non-tumor and tumor tissues. For the targeted treatment of tumors, it is
56 necessary to understand whether sex differences in cancer reflect unique cancer-specific changes, or are
57 reflective of healthy sex-differences that may underlie observed sex-bias in cancer occurrence and disease
58 etiology.

59 In addition to sex differences in overall gene expression due to the wide effects of sex as a biological
60 variable, genetic variants may alter gene expression in a sex-specific manner. A pan-cancer analysis of
61 the TCGA dataset identified 128 germline variants altering gene expression levels (eQTLs) in HCC [12].
62 However, this study purposefully controlled for and removed the effect of sex and, to date, a sex-specific
63 eQTL analysis in HCC has not been performed. Sex-stratified analyses can reveal sex-biased genetic
64 effects on gene expression that may be obscured in a joint analysis of both sexes - e.g. cases where the
65 regulatory variant has a zero or very small effect in one sex, or the eQTL exhibits an opposite effect
66 direction in the two sexes [13]. eQTLs that are discovered in one sex but not in the whole sample analysis
67 are likely to affect gene expression in a sex-dependent manner, and while a combined analysis of both
68 sexes achieves a greater statistical power to detect sex-shared effects, it dilutes the signal of sex-
69 dependent effects [14].

70 Targeted approaches to the treatment of male and female HCC require the characterization and
71 understanding of the fundamental biological mechanisms that differentiate them. Here, we analyzed data
72 from TCGA and The Genotype-Tissue Expression project (GTEx) to examine the sex-specific patterns of
73 gene expression and regulation in HCC. Here, we have contrasted the sex-biased patterns of gene
74 expression in HCC tumors with healthy and tumor-adjacent liver tissues, allowing us to detect sex-
75 differences in gene expression shared between and specific to the different tissues. We show that male
76 and female HCC exhibit differences in the dysregulation of genes and the germline genetic regulation of
77 tumor gene expression. Importantly, these orthogonal approaches identify genes that converge in shared
78 pathways, indicating sex-specific etiology in HCC. The results presented here have implications for the
79 development of targeted therapies for male and female HCC.

80 # Methods

81 ## Data

82 GTEx (release V6p) whole transcriptome (RNAseq) data (dbGaP accession #8834) were downloaded
83 from dbGaP. TCGA LIHC Affymetrix Human Omni 6 array genotype data, whole exome sequencing
84 (WES) and RNAseq data (dbGaP accession #11368) were downloaded from NCI Genomic Data
85 Commons [15]. In total, RNAseq data from 91 male and 45 female GTEx donors, germline genotypes and
86 tumor RNAseq data from 248 male and 119 female TCGA LIHC donors, as well as paired tumor and
87 tumor-adjacent samples from 28 male and 22 female TCGA LIHC donors were utilized in this study.
88 FASTQ read files were extracted from the TCGA LIHC WES BAM files using the *strip_reads*() function
89 of *XYAlign* [16]. We used *FastQC* [17] to assess the WES and RNAseq FASTQ quality. Reads were
90 trimmed using *TRIMMOMATIC IlluminaClip* [18], with the following parameters: seed mismatches 2,
91 palindrome clip threshold 30, simple clip threshold 10, leading quality value 3, trailing quality value 3,
92 sliding window size 4, minimum window quality 30 and minimum read length of 50.

Read mapping and read count quantification

Sequence homology between the X and Y chromosomes may cause the mismapping of short sequencing reads derived from the sex chromosomes and affect downstream analyses [16]. To overcome this, reads were mapped to custom sex-specific reference genomes using *HISAT2* [19]. Female samples were mapped to the human reference genome GRCh38 with the Y-chromosome hard-masked. Male samples were mapped to the human reference genome with Y-chromosomal pseudoautosomal regions hard-masked. Gene-level counts from RNAseq were quantified using *Subread featureCounts* [20]. Reads overlapping multiple features (genes or RNA families with conserved secondary structures) were counted for each feature.

Germline variant calling and filtering

BAM files were processed according to Broad Institute *GATK* (*Genome Analysis Toolkit*) best practices [21–23]: Read groups were added with *Picard Toolkit*'s *AddOrReplaceReadGroups* and optical duplicates marked with *Picard Toolkit*'s *MarkDuplicates* (v.2.18.1, http://broadinstitute.github.io/picard/). Base quality scores were recalibrated with *GATK* (v.4.0.3.0) *BaseRecalibrator*. Germline genotypes were called from whole blood Whole Exome Sequence samples from 248 male and 119 female HCC cases using the scatter-gather method with *GATK HaplotypeCaller* and *GenotypeGVCFs* [21]. Affymetrix 6.0 array genotypes were lifted to GRCh38 using the UCSC *LiftOver* tool [24] and converted to VCF. Filters were applied to retain variants with a minimum quality score > 30, minor allele frequency > 10%, minor allele count > 10, and no call rate < 10% across all samples.

Clinical characteristics and cellular content of tumor samples

Confounding effects, e.g. differences in clinical and pathological characteristics or cell type composition of the sequenced samples, may contribute to the observed effect modification when utilizing stratified

116 analyses. We examined the differences in the clinical characteristics between males and females in the
117 TCGA LIHC cohort. We used a *t*-test to test for the equality of means in patient age and cell type
118 proportions, and Fisher's exact test to test to detect differences in risk factors and pathological
119 classifications (Supplementary Tables S1 and S2).

120 **Filtering of gene expression data**

121 FPKM (Fragments Per Kilobase of transcript per Million mapped reads) expression values for each gene
122 were obtained using *EdgeR* [25]. Each expression dataset was filtered to retain genes with mean
123 FPKM≥0.5 and read count of ≥6 in at least 10 samples across all samples under investigation. In the
124 comparative analysis of differentially expressed genes (DEGs) between the tumor vs. tumor-adjacent
125 samples in males, females, and both sexes, genes that reached the previously described expression
126 thresholds in at least one tissue in at least one sex were retained. This assures that the DEGs detected in
127 the sex-specific and combined analyses are not due to filtering.

128 **Differential expression analysis**

129 For differential expression (DE) analysis, filtered, untransformed read count data were quantile
130 normalized and logCPM transformed with *voom* [26]. From the TCGA LIHC dataset, paired tumor and
131 tumor-adjacent samples were available for 22 females and 28 males. From the GTEx liver dataset, 91
132 male and 45 female samples were used in the DE analysis. A multi-factor design with sex and tissue type
133 as predictor variables were used to fit the linear model. *duplicateCorrelation* function was used to
134 calculate the correlation between measurements made between tumor and tumor-adjacent samples on the
135 same subject, and this inter-subject correlation was accounted for in the linear modeling. As the paired
136 tumor samples differed significantly between the sexes in terms of race, tumor grade, and HBV status,
137 (Supplementary Tables S1 and S2), these parameters were included in the linear models as covariates.
138 Due to missing values in the covariate data, the final numbers of sample pairs used in the analyses were
139 18 females and 26 males.

140 DEGs between comparisons were identified using the *limma*/*voom* pipeline [26] by computing empirical
141 Bayes statistics with *eBayes*. An FDR-adjusted *p*-value threshold of 0.01 and an absolute $\log_2$ fold-change
142 (FC) threshold of 2 were used to select significant DEGs.

143 To reliably detect genes that are expressed in a sex-biased way in HCC but not in non-diseased liver or in
144 tumor-adjacent tissue, we examined genes that were DE in the male vs. female tumor comparison using
145 the previously described significance thresholds, but not in the male vs. female comparisons of normal or
146 tumor-adjacent samples with a relaxed significance threshold of FDR-adjusted *p*-value ≤ 0.1 and absolute
147 $\log_2(FC) \geq 0$.

148 To detect genes that are dysregulated in tumors compared to matched tumor-adjacent samples in each sex,
149 we identified DEGs in the tumor vs. tumor-adjacent comparison of males, females, and in the whole
150 sample. DEGs that were identified in one sex but not in the other or in the combined analysis of both
151 sexes were considered sex-specific. DEGs identified in the combined analysis were considered sex-
152 shared. This approach allows the identification of high-confidence sex-specific events that are a result of
153 the underlying biological differences as opposed to sampling or statistical power. ANOVA and Kruskal-
154 Wallis tests were used to test for equality of fold changes of sex-shared and sex-specific DEGs across
155 male, female, and all samples.

156 **Overrepresentation of biological functions and canonical pathways**

157 We further analyzed the sex-shared and sex-specific tumor vs. tumor-adjacent DEGs as well as the sex-
158 specific eQTL target genes (eGenes) to identify sex-shared and sex-specific pathways driving HCC
159 etiology. We used the *NetworkAnalyst* web tool [27], which utilizes a hypergeometric test to compute *p*-
160 values for the overrepresentation of genes in regards to GO terms and KEGG and Reactome pathways.
161 An FDR-adjusted *p*-value threshold of 0.01 was used to select significantly overrepresented GO terms
162 and canonical pathways.

163 **Accounting for confounding effects and population structure**

164 Gene expression values are affected by genetic, environmental, and technical factors, many of which may
165 be unknown or unmeasured. Technical confounding factors introduce sources of variance that may greatly
166 reduce the statistical power of association studies, and even cause false signals [28]. Thus, it is necessary
167 to account for known and unknown technical confounders. This is often achieved by detecting a set of
168 latent confounding factors with methods such as principal component analysis (PCA) or Probabilistic
169 Estimation of Expression Residuals (PEER) [29]. These surrogate variables are then used as covariates in
170 downstream analyses. We derived 10 PEER factors from the filtered tumor gene expression data and used
171 the weights of these factors as covariates in the eQTL analysis. We used the R package *SNPRelate* [30] to
172 perform PCA on the germline genotype data. We accounted for population structure by applying the first
173 three genotype PCs as covariates in the eQTL analysis.

174 **eQTL analysis**

175 We used eQTL analyses to detect germline genetic effects on tumor gene expression. Similar to the DE
176 analysis, we utilized combined and sex-stratified analyses to detect sex-shared and sex-specific effects.
177 Germline genotypes and tumor gene expression data from 248 male and 119 female donors in the TCGA
178 LIHC cohort were used in the eQTL analysis. Filtered count data was normalized by fitting the FPKM
179 values of each gene and sample to the quantiles of the normal distribution. To account for technical
180 confounders and population structure, 10 *de novo* PEER factors and three genotype principal components
181 were used as covariates. *Cis*-acting (proximal) eQTLs were detected by linear regression as implemented
182 in *QTLtools* v.1.1 [31]. Variants within 1Mb of the gene under investigation were considered for testing.
183 We used the permutation pass with 10,000 permutations to get adjusted *p*-values for associations between
184 the gene expression levels and the top-variants in *cis*: first, permutations are used to derive a nominal *p*-
185 value threshold per gene that reflects the number of independent tests per cis-window. Then, *QTLtools*
186 uses a forward–backward stepwise regression to determine the best candidate variant per signal [31].

187 FDR-adjusted *p*-values were calculated to correct for multiple phenotypes tested, and an adjusted *p*-value
188 threshold of 0.01 was used to select significant associations. To allow the comparison of effect sizes of
189 sex-specific and sex-shared eQTLs across the sexes, effects of each variant located within the 1Mb *cis*-
190 window were obtained using the *QTLtools* nominal pass.

191 Similarly to the tumor vs. tumor-adjacent DEGs, eQTLs that were detected in one sex but not in the other
192 or in the combined analysis were considered sex-specific, while eQTLs detected in the combined analysis
193 were considered sex-shared. ANOVA and Kruskal-Wallis tests were used to test for equality of effect
194 sizes of sex-shared and sex-specific eQTLs across male, female, and all samples.

195 Estimating statistical power in the eQTL analysis

196 We used the R package *powereQTL* [32] to estimate the effect of the sample size to the statistical power
197 to detect eQTLs in the combined analysis of both sexes and in the sex-specific analyses (Fig. S2).

198 Genomic annotations of eQTLs

199 We used the R package *Annotatr* to annotate the genomic locations of eQTLs [33]. Variant sites were
200 annotated for promoters, 5'UTRs, exons, introns, 3'UTRs, CpGs (CpG islands, CpG shores, CpG shelves),
201 and putative regulatory regions based on ChromHMM [34] annotations.

202 Results

203 Sex-specific patterns of gene expression in HCC

204 We identified sex-differences in gene expression in non-diseased liver (GTEx; 21 sex-biased genes with
205 an FDR-adjusted *p*-value ≤ 0.01 and an absolute $\log_2 FC \geq 2$), tumor-adjacent tissue (TCGA LIHC; 21
206 genes), and HCC (TCGA LIHC; 53 genes) to characterize the shared and unique sex-differences that may
207 drive the observed sex-biases in HCC occurrence and etiology (Fig. 1, Supplementary Tables S3-5). X-

208 linked *XIST* and Y-linked genes were expressed in a sex-biased way across all tissues. While sex-biased
209 gene expression in non-diseased and tumor-adjacent tissues may contribute to the sex-differences in
210 cancer occurrence, sex-biased expression in tumors is suggestive of distinct molecular etiologies of male
211 and female HCC. We identified 34 genes that show sex-differences in expression in HCC, but not in
212 tumor-adjacent tissue or non-diseased liver, even with a relaxed significance threshold (Fig. 1A). Notably,
213 Notch-regulating *DTX1* (Fig. 1B) and signal transducer *CD24* were downregulated in male HCC.

214 To further examine the sex-shared and sex-specific mechanisms driving HCC etiology, we detected DEGs
215 between tumor and tumor-adjacent samples in males and females, as well as in the combined analysis of
216 both sexes. Dimensionality reduction of gene expression data shows that variation among the tumor and
217 tumor-adjacent samples is driven by tissue type and sex (Fig. 1C). When inspecting the tumor samples
218 only, the first dimension is largely driven by sex (Supplementary Fig. S1). In the combined analysis of
219 male and female samples, we detected 691 tumor vs. tumor-adjacent DEGs (Supplementary Table S6). In
220 male- and female-specific analyses, we detected 715 and 542 tumor vs. tumor-adjacent DEGs,
221 respectively (Supplementary Tables S7 and S8). Out of the total of 903 unique DEGs, 76.5% were shared
222 between the sexes. We identified 103 female-specific and 108 male-specific tumor vs. tumor-adjacent
223 DEGs. Notably, substantially more DEGs were detected in sex-specific analyses than in the unstratified
224 analysis (Fig. 1D). Specifically, DEGs that showed different magnitudes in fold change between the sexes
225 (based on ANOVA/Kruskal-Wallis tests) were detected in the sex-specific analyses (Fig. 2C, 2D), while
226 DEGs with similar fold changes across all comparisons were detected in the combined analysis as well as
227 the sex-specific analyses (Fig. 2A). Sex-shared DEGs that were only detected in the combined analysis,
228 and not in the sex-specific analyses, showed a large variance in expression and, due to limited power,
229 were not detected as statistically significant DEGs in sex-specific analyses (Fig. 2B). Tumor-infiltrating
230 immune cells may produce spurious signals in DE analyses, which is evident from the detection of
231 various immunoglobulin genes in tumor vs. tumor-adjacent comparisons (Supplementary Tables S6-8).
232 However, male and female samples did not significantly differ in terms of cellular content

233 (Supplementary Table S2), and thus such spurious signals are unlikely to affect male-female comparisons.
234 The observed differences in gene expression are thus likely to reflect actual sex-differences rather than
235 confounding differences in sample characteristics or composition.

236 To put these results in a broader context, we analyzed the male- and female-specific DEGs (tumor vs.
237 tumor-adjacent) for the overrepresentation of functional pathways. We found that the sex-shared and sex-
238 specific DEGs were enriched in pathways relevant to oncogenesis and cancer progression (Supplementary
239 Tables 9-11). We identified pathways that were overrepresented in only one of the sexes but not in the
240 other or in the combined analysis of both sexes, indicating that male and female HCC are partially driven
241 by different mechanisms and processes (Fig. 1E-F).

242 **Differential cis-eQTL effects in male and female HCC**

243 To further investigate the mechanisms of sex-difference in HCC etiology, we used eQTL analyses to
244 detect germline genetic effects on tumor gene expression in both the joint and sex-stratified analyses (Fig.
245 3A). We detected 1,204, 761, and 245 eQTLs in the combined, male-specific, and female-specific
246 analyses, respectively (Supplementary Tables S12-14). As expected, genomic annotations show that most
247 eQTLs are located on non-coding regions (Fig. 3B, Supplementary Tables S15-S17). Consistent with
248 previous reports, most *cis*-eQTLs were located near transcription start sites (TSSs), with 63% of all
249 eQTLs across the combined and sex-specific analyses being located within 20kb of TSSs. On average,
250 384 variants were tested per gene. 31% of the unique shared and sex-specific cis-eQTLs in HCC were
251 also identified as eQTLs in the liver data in the GTEx project analysis release V7, indicating shared tissue
252 origin. Out of the total of 1,595 unique associations, 75.7% were shared between the sexes. We detected
253 295 male-specific and 92 female-specific eQTLs. Since these associations were not detected in the
254 unstratified analysis, they are likely not a result of differential power to detect associations due to
255 different sample sizes, but exhibit effect modification by sex. Sex-specific associations exhibited
256 differences in effect size between the sexes (based on ANOVA/Kruskal-Wallis tests, Fig. 4C, 4D), and

257 the sex-specific effect is diluted in the combined analysis (Fig. 4C, 4D). Sex-shared large effect eQTLs
258 were detected in sex-specific and combined analyses (Fig. 4A), and, due to the larger sample size, sex-
259 shared low-effect eQTLs are detected in the combined analysis only (Fig. 4B).

260 We detected 27 shared eGenes that were associated with independent variants in males and females. This
261 could be due to actual biological differences in gene regulation, or due to technical constraints, in
262 particular, missing genotypes in one sex affecting the permutation scheme to select the top-variant for
263 each target gene. To overcome this and to detect high confidence instances of differential gene regulation
264 between the sexes, we further examined the sex-shared and sex-specific eGenes: we found 24 genes that
265 are under germline regulatory control in only male HCC (Fig. 3C), including *POGLUT1*, which is an
266 essential regulator of Notch signaling (Fig. 3D). No genes were found to be associated with nearby
267 variants in females only, likely due to reduced power to detect associations in females (Supplementary
268 Fig. S2). Male-specific eGenes were overrepresented in pathways related to cell cycle, apoptosis, and
269 cancer (Supplementary Table S18). Concordant with previous studies [14,35], none of the male-specific
270 eGenes were differentially expressed between male and female HCC, indicating that the male-specific
271 eQTLs are not a result of differences in overall gene expression levels between males and females, but are
272 likely to arise from factors such as differential chromatin accessibility or transcription factor activity. The
273 observation that non of the sex-biased autosomal genes in tumors harbor significant *cis*-eQTLs
274 (Supplementary Table S19) also suggests that while sex-specific *cis*-eQTLs may contribute to differences
275 in variance, sex-biased gene expression is likely a result of *trans*-effects, e.g. sex-chromosomal effects on
276 autosomal gene expression, or, more widely, a result of sex as a biological variable, e.g. hormonal effects.

277 Discussion

278 Distinct molecular etiologies of male and female HCC

279 It is well established that patterns of gene expression vary between the sexes across different tissues.
280 Previous studies have confounded these differences with those which may be driving etiological
281 differences between male and female tumors. For example, Yuan et al. previously reported extensive sex-
282 biased signatures in gene expression in HCC and other strongly sex-biased cancers [11]. While they
283 identified immunity and cancer-associated enriched pathways based on sex-biased genes detected in HCC
284 tumors, their approach was limited as it did not include the examination of non-diseased liver nor tumor-
285 adjacent tissues. From the results presented here, we are able to distinguish the differences detected in
286 comparisons of male and female HCC from those reflecting sex-differences in the healthy liver or in
287 tumor-adjacent tissue, as well as to detect genes that are dysregulated in HCC in a sex-shared or sex-
288 specific manner.

289 We characterized differences in gene expression between male and female HCC cases. Notably, sex-
290 differences in gene expression were the largest in the tumor tissue, with 53 genes (including 32 autosomal
291 genes) being expressed in a sex-biased way. These sex-differences point to distinct mechanisms
292 underlying HCC oncogenesis between the sexes, and may partially underlie the observed sex-biases in
293 HCC occurrence and onset. We detected 34 genes that were expressed in a sex-biased way in HCC
294 tumors, but not in healthy or tumor-adjacent liver tissues. Some of these genes are of particular interest in
295 the context of HCC: Notch-regulating *DTX1*, found here to be underexpressed in males compared to
296 females, has been identified as a putative tumor suppressor gene in head and neck squamous cell
297 carcinoma [36]. Another female-biased gene detected here, *CD24*, has a crucial role in T cell homeostasis
298 and autoimmunity [37]. The opposing roles of *CD24* expression in cancer and autoimmune diseases raise
299 interesting questions on the role of sex-differences in immunity underlying sex-differences in cancer.
300 Future studies will focus on better understanding the differential regulation of immune functions between
301 the sexes, and how these differences contribute to the observed biases in disease occurrence and etiology.

302 By sex-specific analyses of matched tumor and tumor-adjacent samples, we detected genes that are
303 uniquely dysregulated in male and female HCC. Further examination of these genes revealed sex-

304 differences in the pathway activation, indicating that the molecular etiologies of male and female HCC
305 are partly driven by distinct functional pathways. Males and females differed in the activation of several
306 signaling pathways, with females showing PPAR pathway enrichment while males showed PI3K,
307 PI3K/AKT, FGFR, EGFR, NGF, GF1R, Rap1, DAP12, and IL-2 signaling pathway enrichment (Fig. 1E,
308 Supplementary Tables 9-10). As these signaling pathways are notable targets for anti-cancer and anti-
309 metastasis therapies [38–44], the results presented here have implications for the targeted treatment of
310 male and female HCC.

311 Sex-specific germline genetic effects on tumor gene expression may drive the molecular 312 etiologies of male and female HCC

313 Sex-specific regulatory functions may underlie sex-differences in cancer etiology, progression, and
314 outcome. We detected sex-differences in the germline genetic regulation of tumor gene expression in
315 HCC, including 24 genes that were under germline regulatory control only in male HCC (Fig. 3).
316 Functional annotations of these male-specific eGenes provide insight into possible regulatory mechanisms
317 contributing to the observed male-bias in HCC and sex-differences in HCC etiology. Protein O-
318 glucosyltransferase 1 (*POGLUT1*) was found to be under germline regulation in male HCC, but not in
319 female HCC or in the joint analysis of both sexes (Fig. 3D). The eQTL associated with *POGLUT1* is
320 located on a promoter region of its target (Supplementary Table S15). *POGLUT1* is an enzyme that is
321 responsible for O-linked glycosylation of proteins. Altered glycosylation of proteins has been observed in
322 many cancers [45,46], including liver cancer [47,48]. *POGLUT1* is an essential regulator of Notch
323 signaling and is likely involved in cell fate and tissue formation during development. Genes involved in
324 Notch and PI3K/AKT signaling were also found to be expressed in a sex-biased way in HCC tumors and
325 overrepresented among the male-specific DEGs detected in the tumor vs. tumor-adjacent comparison,
326 showing that sex-specific eQTLs and sex-specific dysregulated genes converge in canonical pathways.
327 Notch signaling pathway was also detected as overrepresented (FDR-adj. p-value ≤ 0.01) among the 24

328 male-specific eGenes. PI3K-AKT is known to co-operate with Notch by triggering inflammation and
329 immunosuppression [49]. These results point to a major role of the Notch/PI3K/AKT axis in the
330 development of HCC in males. PI3K/AKT/mTOR signaling is of particular interest in the context of
331 HCC, as it has been implicated in HCC carcinogenesis [50], is involved in hepatic gluconeogenesis [51],
332 and is activated in a sex-biased way in the liver and other tissues [52]. The role of Notch and PI3K/AKT
333 signaling in HCC may differ between early and late-stage tumors and among molecular subtypes, and
334 further studies are necessary to understand the possible oncogenic properties of these pathways among
335 HCC subtypes and between the sexes. In the future, analyses of data collected as a part of the
336 International Cancer Genomics Consortium project may elucidate the sex-specific processes of HCC
337 oncogenesis among the Japanese, as well as the interactions between sex and hepatitis infections in
338 shaping HCC etiology. However, each dataset has a unique ancestry composition and are not directly
339 comparable for validation purposes.

340 Conclusions

341 In summary, we discovered differential regulatory functions in HCC tumors between the sexes. This work
342 provides a framework for future studies on sex-biased cancers. Further studies are required to identify and
343 validate sex-specific genetic effects on tumor gene expression and its consequences in HCC and other
344 sex-biased cancers across diverse populations.

345 Abbreviations

346 HCC: Hepatocellular Carcinoma. TCGA: The Cancer Genome Atlas. GTEx: Genotype x Tissue
347 Expression Project. HBV: Hepatitis B virus. HCV: Hepatitis C virus. DEG: Differentially expressed gene.
348 eQTL: Expression quantitative trait loci. TSS: Transcription start site.

349 Declarations

Acknowledgments

We thank Dr. Nicholas Banovich and anonymous reviewers for their feedback on previous versions of this manuscript.

Funding

This study was supported by ASU Center for Evolution and Medicine postdoctoral fellowship and the Marcia and Frank Carlucci Charitable Foundation postdoctoral award from the Prevent Cancer Foundation for HMN, ASU School of Life Sciences and the Biodesign Institute startup funds for MAW, and ASU Center for Evolution and Medicine Venture funds.

Availability of data and materials

Data used in this study are available at dbGaP NCI Genomic Data Commons.

Authors' contributions

Conception and design: HMN, MAW, KB. Development of methodology: HMN, MAW. Acquisition of data: MAW, KB. Analysis and interpretation of data: HMN. Writing, review, and/or revision of the manuscript: HMN, MAW, KB. Study supervision: MAW, KB. All authors approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

369 The authors declare no competing interests.

370 Additional files

371 **Additional files contain supplementary information with two figures, and 19 supplementary tables.**

372 **Fig S1.** A multi-dimensional scaling plot of the TCGA LIHC tumor samples of each sex (N male = 248,
373 N female = 119). Euclidean distances between samples were calculated based on 100 genes with the
374 largest standard deviations between samples.

375 **Fig S2.** Estimation of statistical power in the combined (grey), male-specific (blue), and female-specific
376 (red) eQTL analyses with a *p*-value level 0.01 and 384 variants. Increased power in the combined analysis
377 allows the detection of sex-shared low-effect eQTLs.

378 **Table S1.** Categorical clinical characteristics of male and female patients in the TCGA LIHC cohort.
379 Table S1. Categorical clinical characteristics of samples derived from male and female patients in the
380 TCGA LIHC cohort. The following symbols are used to indicate statistical significance: * = $p \leq 0.10$, **
381 = $p \leq 0.05$, *** = $p \leq 0.01$.

382 **Table S2.** Continuous clinical characteristics of male and female patients in the TCGA LIHC cohort.

383 **Table S3.** Sex-biased gene expression in the GTEx liver tissue samples with an FDR-adjusted *p*-value ≤
384 0.1 and absolute $\log_2$ fold-change ≥ 0. N males = 91, N females = 45.

385 **Table S4.** Sex-biased gene expression in the TCGA LIHC tumor-adjacent samples with an FDR-adjusted
386 *p*-value ≤ 0.1 and an absolute $\log_2$ fold-change ≥ 0. N males = 22, N females = 18.

387 **Table S5.** Sex-biased gene expression in TCGA LIHC tumor samples with an FDR-adjusted *p*-value ≤
388 0.01 and an absolute $\log_2$ fold-change ≥ 2. N males = 22, N females = 18.

389 **Table S6.** Differentially expressed genes between matched tumor and tumor-adjacent samples in the
390 combined analysis of male and female samples with an FDR-adjusted *p*-value ≤ 0.01 and an absolute $\log_2$
391 fold-change ≥ 2. N male sample pairs = 22, N female sample pairs = 18.

392 **Table S7.** Differentially expressed genes between matched male tumor and tumor-adjacent samples with
393 an FDR-adjusted *p*-value ≤ 0.01 and an absolute $\log_2$ fold-change ≥ 2. N of sample pairs = 22.

394 **Table S8.** Differentially expressed genes between matched female tumor and tumor-adjacent samples
395 with an FDR-adjusted *p*-value ≤ 0.01 and an absolute $\log_2$ fold-change ≥ 2. N of sample pairs = 18.

396 **Table S9.** Overrepresented GO terms and canonical pathways in the sex-shared tumor vs. tumor-adjacent
397 DEGs. Significant terms and pathways were selected based on an FDR-adjusted *p*-value threshold of
398 0.01.

399 **Table S10.** Overrepresented GO terms and canonical pathways in the male-specific tumor vs. tumor-
400 adjacent DEGs. Significant terms and pathways were selected based on an FDR-adjusted *p*-value
401 threshold of 0.01.

402 **Table S11.** Overrepresented GO terms and canonical pathways in the female-specific tumor vs. tumor-
403 adjacent DEGs. Significant terms and pathways were selected based on an FDR-adjusted *p*-value
404 threshold of 0.01.

405 **Table S12.** *cis*-eQTLs detected in the combined analysis of both sexes. N=367 (N males = 248, N
406 females = 119). Significant eQTLs were selected based on an FDR-adjusted *p*-value threshold 0.01.

407 **Table S13.** *cis*-eQTLs detected in the male-specific analysis. N=248. Significant eQTLs were selected
408 based on an FDR-adjusted *p*-value threshold 0.01.

409 **Table S14.** *cis*-eQTLs detected in the female-specific analysis. N=119. Significant eQTLs were selected
410 based on an FDR-adjusted *p*-value threshold 0.01.

411 **Table S15.** Genomic annotations of eQTLs detected in the combined analysis of both sexes.

412 **Table S16.** Genomic annotations of eQTLs detected in the male-specific analysis.

413 **Table S17.** Genomic annotations of eQTLs detected in the female-specific analysis.

414 **Table S18.** Overrepresented canonical pathways in the male-specific eQTL target genes. Significant
415 terms and pathways were selected based on an FDR-adjusted *p*-value threshold of 0.01.

416 **Table S19.** Top-variants associated with autosomal genes that were expressed in a sex-biased way in
417 HCC tumor samples.

418 Figure Captions

419 **Fig. 1. Patterns of gene expression and molecular etiologies of male and female HCC. A:** Sex-biased
420 gene expression in HCC. A volcano plot of DEGs between male (N=26) and female (N=18) HCC tumor
421 samples. X-linked genes are indicated in pink, Y-linked in green, and autosomal in black. Significant
422 genes were selected based on an FDR-adjusted *p*-value threshold of 0.01 and absolute $\log_2$(FC) threshold
423 of 2. Multiple transcripts of the long non-coding RNA *XIST* are independently expressed. Genes that were
424 not expressed in a sex-biased way in healthy liver (GTEx) or in the tumor-adjacent tissues are indicated
425 with an asterisk. **B:** An example of a gene exhibiting a sex-bias in HCC but not in healthy liver or tumor-
426 adjacent tissues. *DTX1* expression in log(CPM) is shown for male and female samples in each tissue. **C:**
427 A multi-dimensional scaling plot of the paired TCGA LIHC tumor and tumor-adjacent samples of each
428 sex. Euclidean distances between samples were calculated based on 100 genes with the largest standard
429 deviations between samples. Tissue type (dimension 1) and sex (dimension 2) drive the overall patterns of
430 gene expression in HCC. **D:** Venn-diagram of the overlap of DEGs in the sex-specific and combined
431 analysis of matched tumor and tumor-adjacent samples. Substantially more DEGs were identified in the
432 sex-specific analyses. **E:** Sex-specific and sex-shared DEGs were analyzed for the overrepresentation of

433 functional pathways. Sex-specific patterns of pathway enrichment point to differential processes driving
434 the etiology of male and female HCC. **F:** Examples of sex-specific and sex-shared pathways.

435 **Fig 2.** Absolute $\log_2$-fold changes of DEGs detected from tumor vs. tumor-adjacent comparisons in the
436 sex-specific analyses and combined analysis of both sexes. Absolute $\log_2$-fold changes are given for
437 female samples, male samples, and across all samples. Global *p*-values for ANOVA are shown for each
438 DEG type. Adjusted *p*-values based on Kruskal-Wallis tests are shown for each pairwise comparison.

439 **Fig. 3. Sex-specific genetic effects on tumor gene expression in HCC. A:** QQ-plot of eQTL
440 associations in the combined analysis of both sexes (grey), male-specific analysis (blue), and female-
441 specific analysis (red). **B:** Genomic annotations of eQTLs in the combined analysis of both sexes, male-
442 specific analysis, and female-specific analysis. **C:** Overlap of eGenes detected in combined and sex-
443 specific analyses. **D:** An example of a male-specific eQTL. *POGLUT1* expression in tumors is modulated
444 by a germline variant in *cis* in male HCC, but not in female HCC nor in the combined analysis of both
445 sexes, indicating effect modification by sex. Numbers of individuals with each genotype, adjusted
446 significance, and effect size (β) are given for each model.

447 **Fig 4.** Absolute effect sizes of sex-shared and sex-specific eQTLs in males, females, and the whole study
448 sample. Due to the larger sample size, sex-shared low-effect eQTLs are only detected as significant in the
449 combined analysis (A). Sex-shared large effect eQTLs are detected in the combined analysis as well as
450 the sex-specific analyses (B). Sex-specific eQTLs exhibit a larger effect in one sex than the other, and the
451 effect is diluted in the combined analysis (C, D). Sex-shared large effect eQTLs can be detected in sex-
452 specific and combined analyses. Global *p*-values for ANOVA are shown for each eQTL type. Adjusted *p*-
453 values based on Kruskal-Wallis tests are shown for each pairwise comparison.

454 # References

455 1. Clocchiatti A, Cora E, Zhang Y, Dotto GP. Sexual dimorphism in cancer. Nat Rev Cancer.
456 2016;16: 330.

457 2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018:
458 GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer
459 J Clin. 2018; doi:10.3322/caac.21492

460 3. Wisnivesky JP, Halm EA. Sex differences in lung cancer survival: do tumors behave differently
461 in elderly women? J Clin Oncol. 2007;25: 1705–1712.

462 4. OuYang P-Y, Zhang L-N, Lan X-W, Xie C, Zhang W-W, Wang Q-X, et al. The significant
463 survival advantage of female sex in nasopharyngeal carcinoma: a propensity-matched analysis. Br J
464 Cancer. 2015;112: 1554–1561.

465 5. Li CH, Haider S, Shiah Y-J, Thai K, Boutros PC. Sex Differences in Cancer Driver Genes and
466 Biomarkers. Cancer Res. 2018;78: 5527–5537.

467 6. Wands J. Hepatocellular carcinoma and sex. N Engl J Med. 2007;357: 1974–1976.

468 7. Petrick JL, Braunlin M, Laversanne M, Valery PC, Bray F, McGlynn KA. International trends in
469 liver cancer incidence, overall and by histologic subtype, 1978-2007. Int J Cancer. 2016;139: 1534–1545.

470 8. Ladenheim MR, Kim NG, Nguyen P, Le A, Stefanick ML, Garcia G, et al. Sex differences in
471 disease presentation, treatment and clinical outcomes of patients with hepatocellular carcinoma: a single-
472 centre cohort study. BMJ Open Gastroenterol. 2016;3: e000107.

473 9. Gilks WP, Abbott JK, Morrow EH. Sex differences in disease genetics: evidence, evolution, and
474 detection. Trends Genet. 2014;30: 453–463.

475 10. Morrow EH. The evolution of sex differences in disease. Biol Sex Differ. 2015;6: 5.

476 11. Yuan Y, Liu L, Chen H, Wang Y, Xu Y, Mao H, et al. Comprehensive Characterization of
477 Molecular Differences in Cancer between Male and Female Patients. Cancer Cell. 2016;29: 711–722.

478 12. Gong J, Mei S, Liu C, Xiang Y, Ye Y, Zhang Z, et al. PancanQTL: systematic identification of
479 cis-eQTLs and trans-eQTLs in 33 cancer types. Nucleic Acids Res. 2018;46: D971–D976.

480 13. Behrens G, Winkler TW, Gorski M, Leitzmann MF, Heid IM. To stratify or not to stratify: power
481 considerations for population-based genome-wide association studies of quantitative traits. Genet
482 Epidemiol. 2011;35: 867–879.

483 14. Dimas AS, Nica AC, Montgomery SB, Stranger BE, Raj T, Buil A, et al. Sex-biased genetic
484 effects on gene regulation in humans. Genome Res. 2012;22: 2368–2375.

485 15. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a Shared
486 Vision for Cancer Genomic Data. N Engl J Med. 2016;375: 1109–1112.

487 16. Webster TH, Couse M, Grande BM, Karlins E, Phung TN, Richmond PA, et al. Identifying,
488 understanding, and correcting technical biases on the sex chromosomes in next-generation sequencing
489 data [Internet]. bioRxiv. 2018. p. 346940. doi:10.1101/346940

490 17. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data. In:
491 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Internet]. 2010. Available:
492 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

493 18. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
494 Bioinformatics. 2014;30: 2114–2120.

495 19. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements.
496 Nat Methods. 2015;12: 357–360.

497 20. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning
498 sequence reads to genomic features. Bioinformatics. 2014;30: 923–930.

499 21. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome
500 Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.
501 Genome Res. 2010;20: 1297–1303.

502 22. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for
503 variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:
504 491–498.

505 23. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al.
506 From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.
507 Curr Protoc Bioinformatics. 2013;43: 11.10.1–33.

508 24. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC
509 Genome Browser Database: update 2006. Nucleic Acids Res. 2006;34: D590–8.

510 25. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
511 expression analysis of digital gene expression data. Bioinformatics. 2010;26: 139–140.

512 26. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis
513 tools for RNA-seq read counts. Genome Biol. 2014;15: R29.

514 27. Xia J, Gill EE, Hancock REW. NetworkAnalyst for statistical, visual and network-based meta-
515 analysis of gene expression data. Nat Protoc. 2015;10: 823–844.

516 28. Leek JT, Storey JD. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable
517 Analysis. PLoS Genet. 2007;3: e161.

518 29. Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression
519 residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc.
520 2012;7: 500–507.

521 30. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing
522 toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28: 3326–
523 3328.

524 31. Delaneau O, Ongen H, Brown AA, Fort A, Panousis N, Dermitzakis E. A complete tool set for
525 molecular QTL discovery and analysis. bioRxiv. 2016; doi:10.1101/068635

526 32. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45: 580–
527 585.

528 33. Cavalcante RG, Sartor MA. annotatr: genomic regions in context. Bioinformatics. 2017;33:
529 2381–2383.

530 34. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat
531 Methods. 2012;9: 215–216.

532 35. Kukurba KR, Parsana P, Balliu B, Smith KS, Zappala Z, Knowles DA, et al. Impact of the X
533 Chromosome and sex on regulatory variation. Genome Res. 2016;26: 768–777.

534 36. Gaykalova DA, Zizkova V, Guo T, Tiscareno I, Wei Y, Vatapalli R, et al. Integrative
535 computational analysis of transcriptional and epigenetic alterations implicates DTX1 as a putative tumor
536 suppressor gene in HNSCC. Oncotarget. 2017;8: 15349–15363.

537 37. Liu Y, Zheng P. CD24: a genetic checkpoint in T cell homeostasis and autoimmune diseases.
538 Trends Immunol. 2007;28: 315–320.

539 38. Gou Q, Gong X, Jin J, Shi J, Hou Y. Peroxisome proliferator-activated receptors (PPARs) are
540 potential drug targets for cancer therapy. Oncotarget. 2017;8: 60704–60709.

541 39. Mayer IA, Arteaga CL. The PI3K/AKT Pathway as a Target for Cancer Treatment. Annu Rev
542 Med. 2016;67: 11–28.

543 40. Porta R, Borea R, Coelho A, Khan S, Araújo A, Reclusa P, et al. FGFR a promising druggable
544 target in cancer: Molecular biology and new drugs. Crit Rev Oncol Hematol. 2017;113: 256–267.

545 41. Seshacharyulu P, Ponnusamy MP, Haridas D, Jain M, Ganti AK, Batra SK. Targeting the EGFR
546 signaling pathway in cancer therapy. Expert Opin Ther Targets. 2012;16: 15–31.

547 42. Demir IE, Tieftrunk E, Schorn S, Friess H, Ceyhan GO. Nerve growth factor & TrkA as novel
548 therapeutic targets in cancer. Biochim Biophys Acta. 2016;1866: 37–50.

549 43. Zhang Y-L, Wang R-C, Cheng K, Ring BZ, Su L. Roles of Rap1 signaling in tumor cell
550 migration and invasion. Cancer Biol Med. 2017;14: 90–99.

551 44. Choudhry H, Helmi N, Abdulaal WH, Zeyadi M, Zamzami MA, Wu W, et al. Prospects of IL-2
552 in Cancer Immunotherapy. Biomed Res Int. 2018;2018: 9056173.

553 45. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. Nat Rev
554 Cancer. 2015;15: 540–555.

555 46. Stowell SR, Ju T, Cummings RD. Protein glycosylation in cancer. Annu Rev Pathol. 2015;10:
556 473–510.

557 47. Mehta A, Herrera H, Block T. Chapter Seven - Glycosylation and Liver Cancer. In: Drake RR,
558 Ball LE, editors. Advances in Cancer Research. Academic Press; 2015. pp. 257–279.

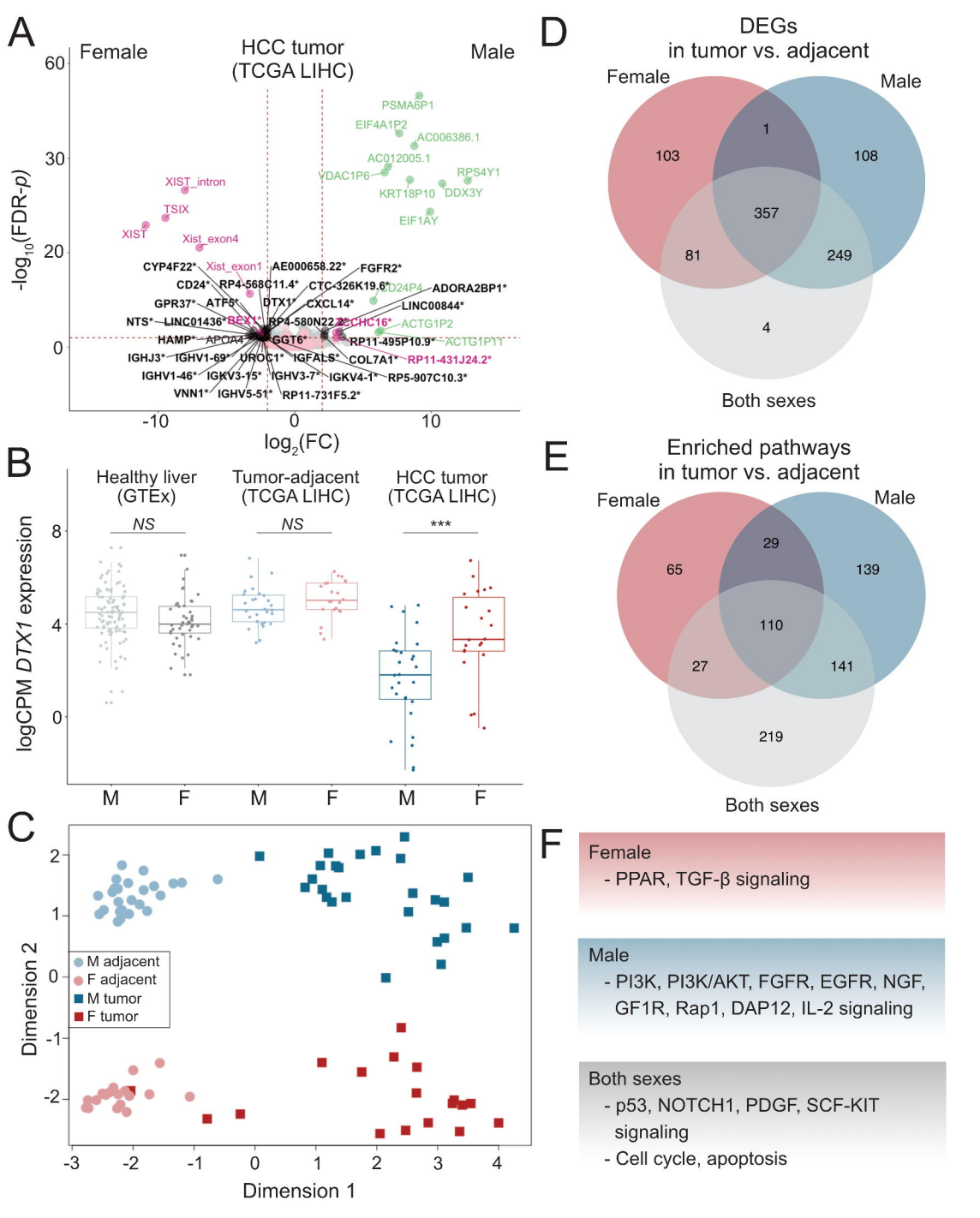
559 48. Liang K-H, Yeh C-T. O-glycosylation in liver cancer: Clinical associations and potential
560 mechanisms. Liver Research. 2017;1: 193–196.

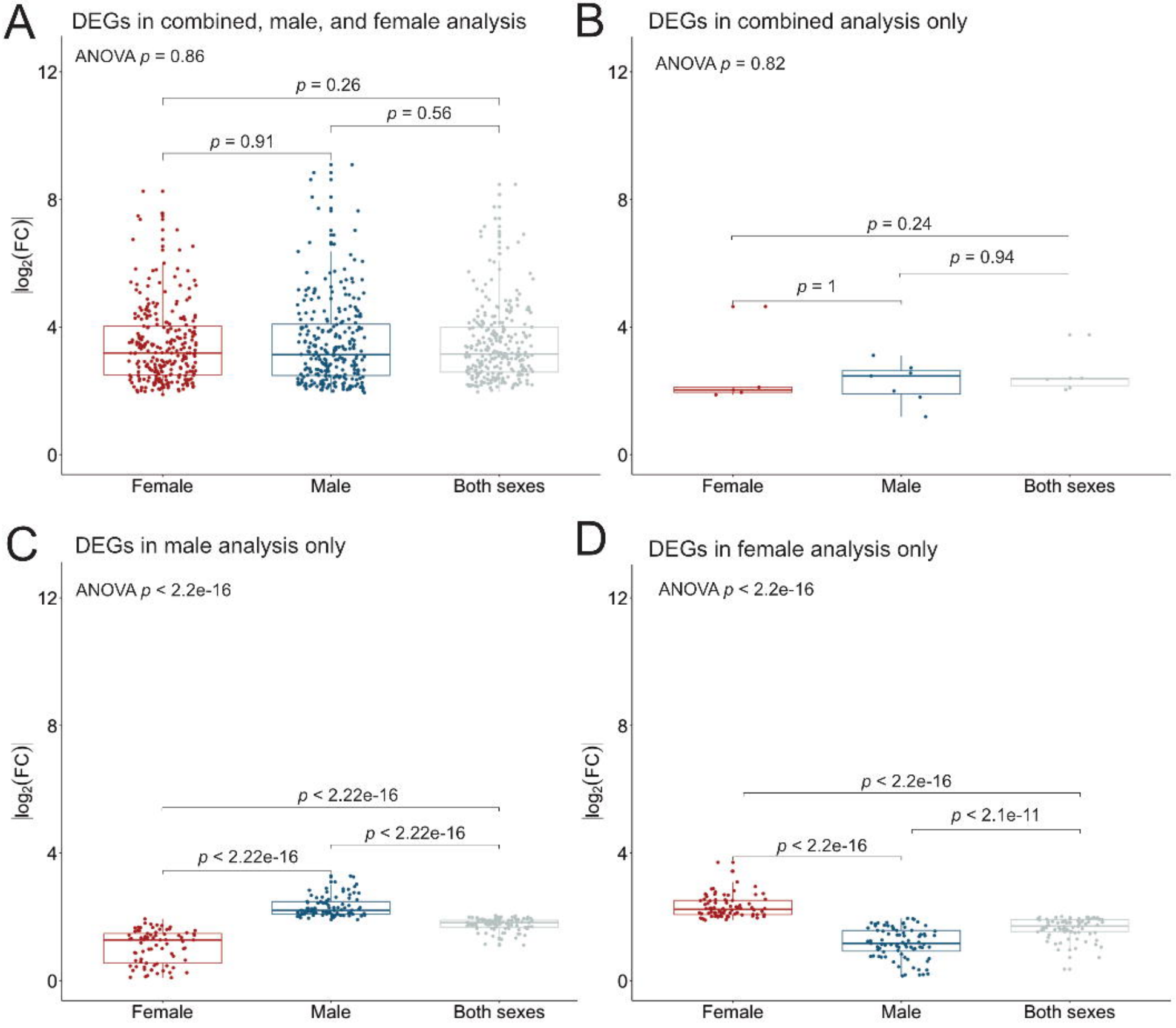
561 49. Villegas SN, Gombos R, García-López L, Gutiérrez-Pérez I, García-Castillo J, Vallejo DM, et al.
562 PI3K/Akt Cooperates with Oncogenic Notch by Inducing Nitric Oxide-Dependent Inflammation. Cell
563 Rep. 2018;22: 2541–2549.

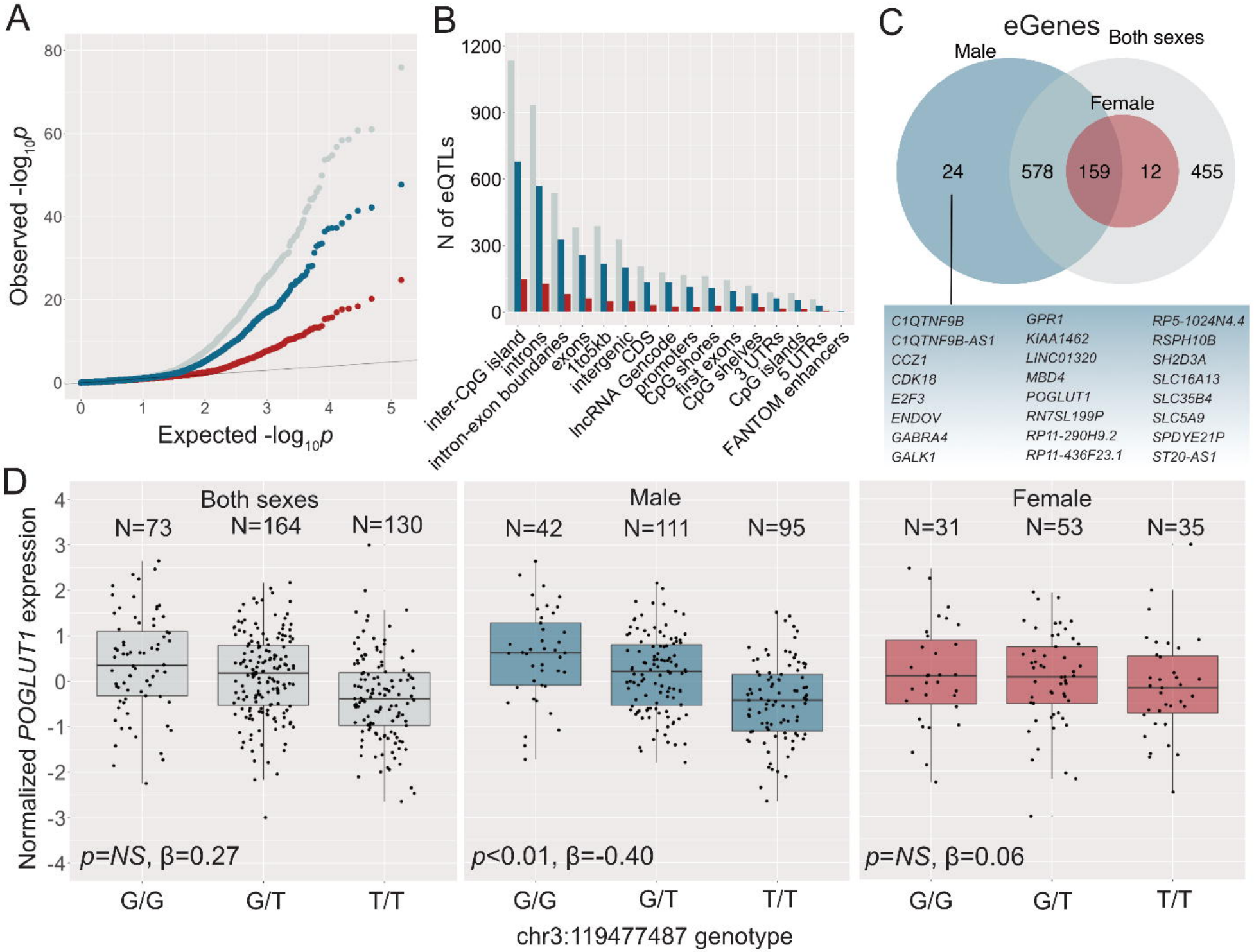
564 50. Wang G-L, Iakova P, Wilde M, Awad S, Timchenko NA. Liver tumors escape negative control of
565 proliferation via PI3K/Akt-mediated block of C/EBP alpha growth inhibitory activity. Genes Dev.
566 2004;18: 912–925.

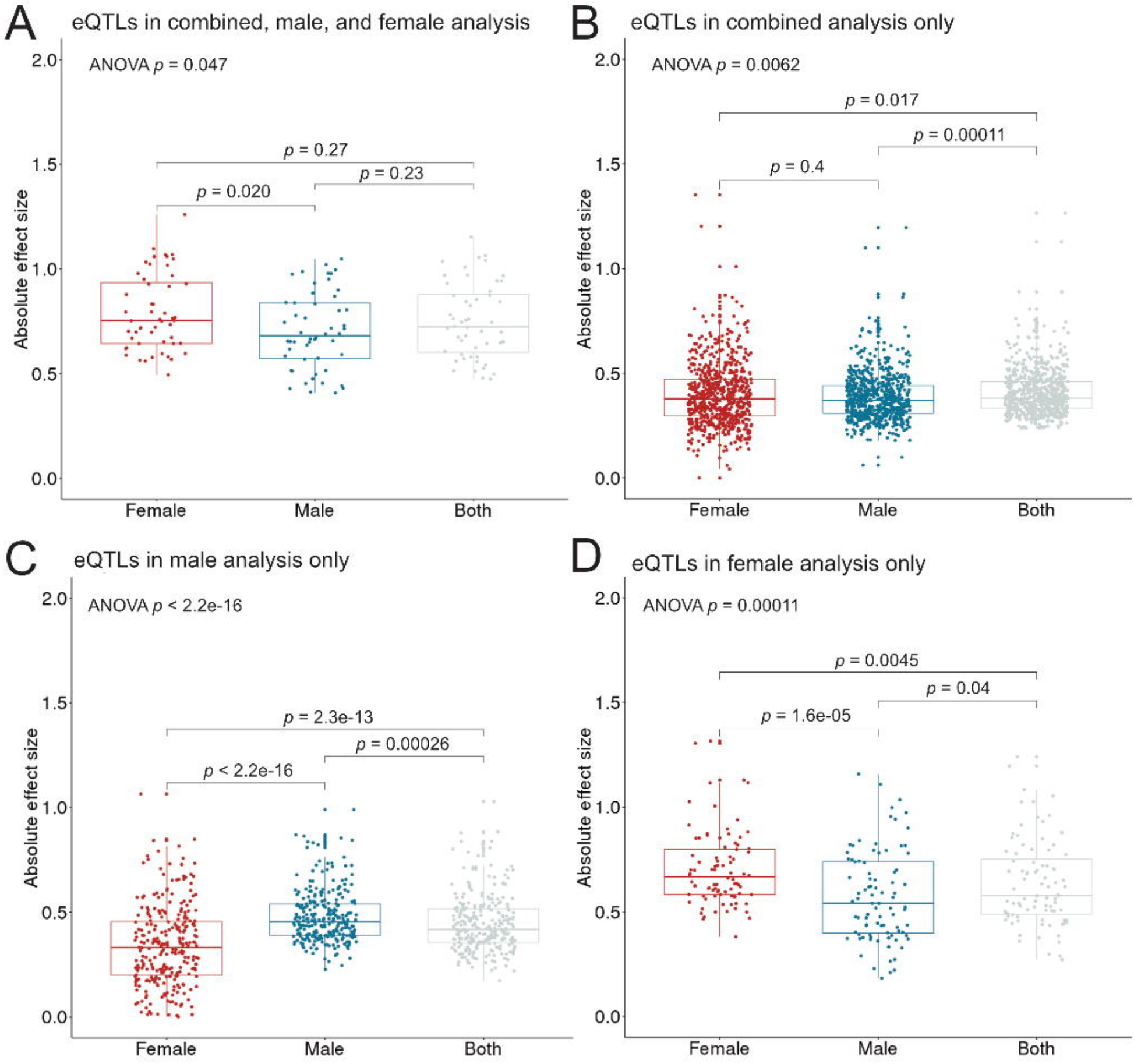
567 51. Huang X, Liu G, Guo J, Su Z. The PI3K/AKT pathway in obesity and type 2 diabetes. Int J Biol
568 Sci. 2018;14: 1483–1496.

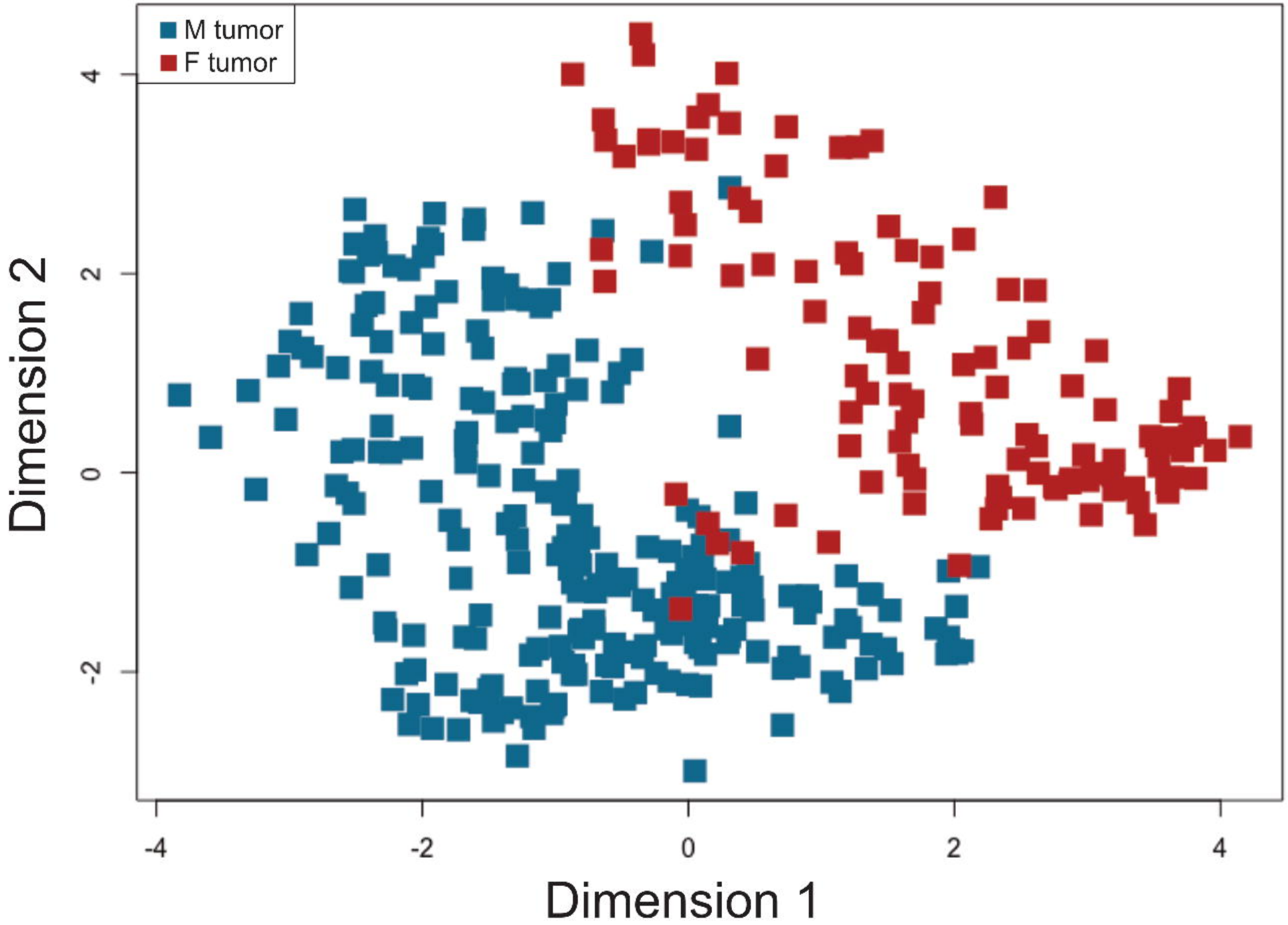
569 52. Baar EL, Carbajal KA, Ong IM, Lamming DW. Sex- and tissue-specific changes in mTOR
570 signaling with age in C57BL/6J mice. Aging Cell. 2016;15: 155–166.

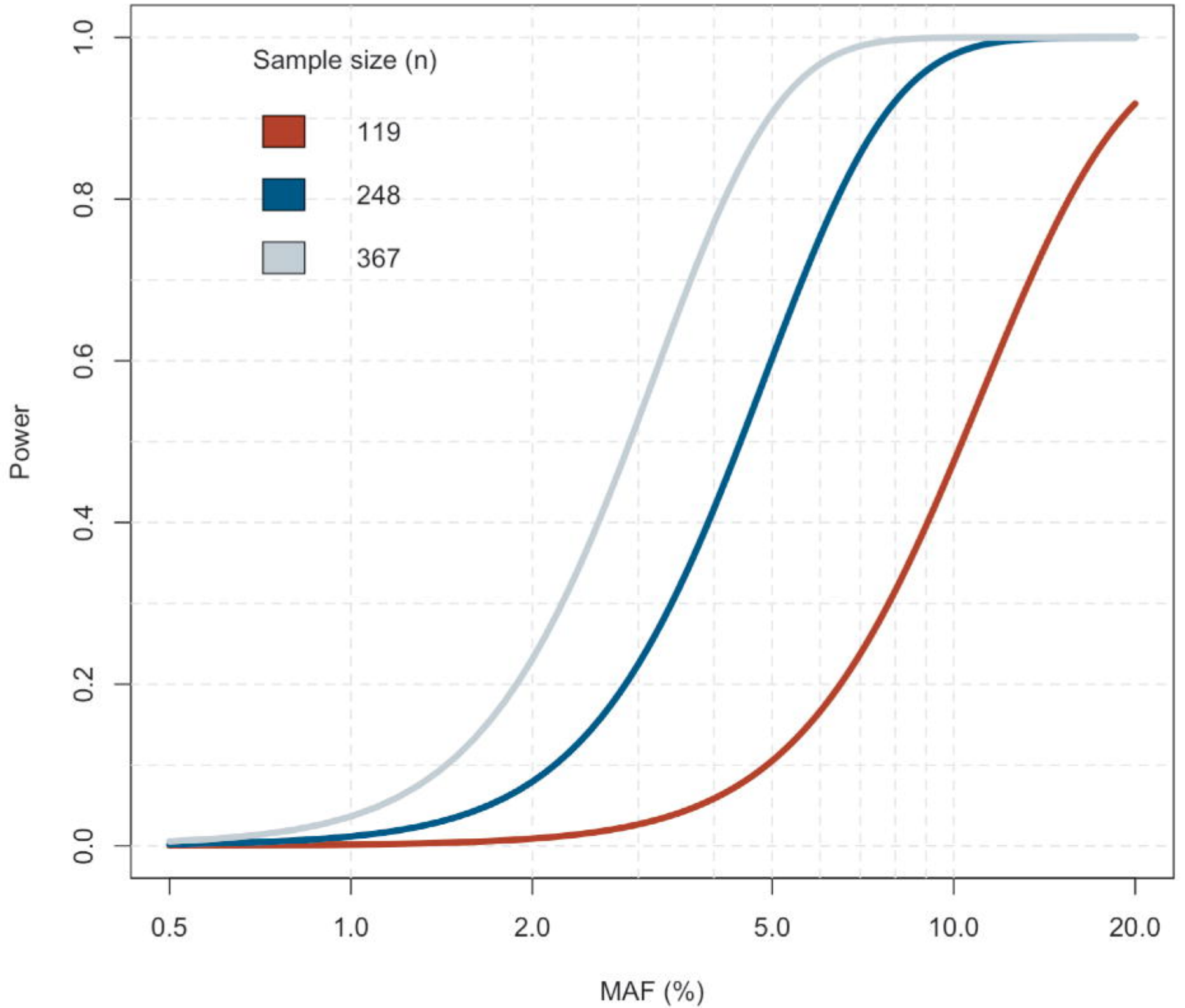

A
Female
HCC tumor
(TCGA LIHC)
Male
-log10(FDR-p)
log2(FC)
PSMA6P1
EIF4A1P2
AC006386.1
AC012005.1
VDAC1P6
RPS4Y1
KRT18P10
DDX3Y
EIF1AY
XIST_intron
TSIX
XIST
Xist_exon4
Xist_exon1
CYP4F22*
AE000658.22*
FGFR2*
CD24*
RP4-568C11.4*
CTC-326K19.6*
ADORA2BP1*
GPR37*
ATF5*
DTX1*
CXCL14*
LINC00844*
NTS*
LINC01436*
BEX1*
RP4-580N22.2*
CHC16*
ACTG1P2
HAMP*
APOA4*
GGT6*
RP11-495P10.9*
ACTG1P11
IGHJ3*
IGHV1-69*
UROC1*
IGFALS*
COL7A1*
RP11-431J24.2*
IGHV1-46*
IGKV3-15*
IGHV3-7*
IGKV4-1*
RP5-907C10.3*
VNN1*
IGHV5-51*
RP11-731F5.2*
B
Healthy liver
(GTEx)
Tumor-adjacent
(TCGA LIHC)
HCC tumor
(TCGA LIHC)
NS
NS
***
logCPM DTX1 expression
M
F
M
F
M
F
C
M adjacent
F adjacent
M tumor
F tumor
Dimension 1
Dimension 2
D
DEGs
in tumor vs. adjacent
Female
Male
103
1
108
357
81
249
4
Both sexes
E
Enriched pathways
in tumor vs. adjacent
Female
Male
65
29
139
110
27
141
219
Both sexes
F
Female
- PPAR, TGF-β signaling
Male
- PI3K, PI3K/AKT, FGFR, EGFR, NGF, GF1R, Rap1, DAP12, IL-2 signaling
Both sexes
- p53, NOTCH1, PDGF, SCF-KIT signaling
- Cell cycle, apoptosis


A
DEGs in combined, male, and female analysis
ANOVA p = 0.86
p = 0.26
p = 0.56
p = 0.91
|log2(FC)|
12
8
4
0
Female
Male
Both sexes
B
DEGs in combined analysis only
ANOVA p = 0.82
p = 0.24
p = 0.94
p = 1
C
DEGs in male analysis only
ANOVA p < 2.2e-16
p < 2.22e-16
p < 2.22e-16
p < 2.22e-16
D
DEGs in female analysis only
ANOVA p < 2.2e-16
p < 2.2e-16
p < 2.1e-11
p < 2.2e-16


A
Observed -log10p
Expected -log10p
B
N of eQTLs
inter-CpG island
introns
intron-exon boundaries
exons
1to5kb
intergenic
CDS
lncRNA Gencode
promoters
CpG shores
first exons
CpG shelves
3 UTRs
CpG islands
5 UTRs
FANTOM enhancers
C
eGenes
Male
Both sexes
Female
24
578
159
12
455
C1QTNF9B
C1QTNF9B-AS1
CCZ1
CDK18
E2F3
ENDOV
GABRA4
GALK1
GPR1
KIAA1462
LINC01320
MBD4
POGLUT1
RN7SL199P
RP11-290H9.2
RP11-436F23.1
RP5-1024N4.4
RSPH10B
SH2D3A
SLC16A13
SLC35B4
SLC5A9
SPDYE21P
ST20-AS1
D
Normalized POGLUT1 expression
Both sexes
N=73
N=164
N=130
p=NS, β=0.27
Male
N=42
N=111
N=95
p<0.01, β=-0.40
Female
N=31
N=53
N=35
p=NS, β=0.06
G/G
G/T
T/T
chr3:119477487 genotype


A
eQTLs in combined, male, and female analysis
ANOVA p = 0.047
p = 0.27
p = 0.23
p = 0.020
Absolute effect size
2.0
1.5
1.0
0.5
0.0
Female
Male
Both
B
eQTLs in combined analysis only
ANOVA p = 0.0062
p = 0.017
p = 0.00011
p = 0.4
Absolute effect size
2.0
1.5
1.0
0.5
0.0
Female
Male
Both
C
eQTLs in male analysis only
ANOVA p < 2.2e-16
p = 2.3e-13
p = 0.00026
p < 2.2e-16
Absolute effect size
2.0
1.5
1.0
0.5
0.0
Female
Male
Both
D
eQTLs in female analysis only
ANOVA p = 0.00011
p = 0.0045
p = 0.04
p = 1.6e-05
Absolute effect size
2.0
1.5
1.0
0.5
0.0
Female
Male
Both


M tumor
F tumor
Dimension 2
Dimension 1
4
2
0
-2
-4
-2
0
2
4


Sample size (n)
119
248
367
Power
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.0
2.0
5.0
10.0
20.0
MAF (%)