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)

The authors declare no potential conflicts of interest.

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 of the 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. By sex-specific analysis of tumor and tumor-adjacent samples, we uncovered etiologically relevant genes, pathways, and networks differentiating male and female HCC. While both sexes exhibited activation of pathways related to apoptosis and p53 signaling, pathways involved in innate and adaptive immunity showed differential activation between the sexes. Using eQTL analyses, we discovered germline regulatory 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 role of sex in modifying the effects of regulatory variants in HCC. The genes that showed sex-specific dysregulation and harbored a sex-specific eQTL converge in canonical 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 insights into the role of genetic regulation of transcription in modulating sex-differences in HCC etiology and provide a framework for future studies on sex-biased cancers.

Introduction

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

Sex-specific gene regulation may partially underlie differences between the sexes in disease prevalence and severity (9,10). Previous work observed extensive sex-biased signatures in gene expression in HCC and other sex-biased cancers (11). However, this study focused solely on comparing male and female tumor samples, while ignoring sex-differences in healthy and tumor-adjacent tissues. To understand cancer-specific processes, it is necessary to contrast the sex-differences in gene expression identified in HCC with those in healthy and tumor tissues. For the targeted treatment of tumors, it is necessary to understand whether sex differences in cancer reflect unique cancer-specific changes, or are reflective of healthy sex-differences that may underlie observed sex-bias in cancer occurrence and disease etiology.

In addition to sex differences in overall gene expression due to the wide effects of sex as a biological variable, genetic variants may alter gene expression in a sex-specific manner. A pan-cancer analysis of the TCGA dataset identified 128 germline variants altering gene expression levels (eQTLs) in HCC (12). However, this study purposefully controlled for and removed the effect of sex and, to date, a sex-specific eQTL analysis in HCC has not been performed. Sex-stratified analyses can reveal sex-biased genetic effects on gene expression that may be obscured in a joint analysis of both sexes - e.g. cases where the regulatory variant has a zero or very small effect in one sex, or the eQTL exhibits an opposite effect direction in the two sexes (13). eQTLs that are discovered in one sex but not in the whole sample analysis are likely to affect gene

expression in a sex-dependent manner, and while a combined analysis of both sexes achieves a greater statistical power to detect sex-shared effects, it dilutes the signal of sex-dependent effects (14).

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

Materials and Methods

Data

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

97 **Read mapping and read count quantification**

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

106 **Germline variant calling and filtering**

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

117 **Clinical characteristics and cellular content of tumor samples**

118 Confounding effects, e.g. differences in clinical and pathological characteristics or cell type
119 composition of the sequenced samples, may contribute to the observed effect modification when
120 utilizing stratified analyses. We examined the differences in the clinical characteristics between
121 males and females in the TCGA LIHC cohort. We used a *t*-test to test for the equality of means
122 in patient age, and Fisher's exact test to test to detect differences in risk factors and pathological
123 classifications (Supplementary Table S1). To examine the cellular composition of tumor

124 samples, we performed cell type enrichment analysis for 64 immune and stromal cell types using
125 *xCell* (25).

126 **Filtering of gene expression data**

127 FPKM (Fragments Per Kilobase of transcript per Million mapped reads) expression values for
128 each gene were obtained using *EdgeR* (26). Each expression dataset was filtered to retain genes
129 with mean FPKM>0.5 and read count of >6 in at least 10 samples across all samples under
130 investigation. In the comparative analysis of differentially expressed genes (DEGs) between the
131 tumor vs. tumor-adjacent samples in males, females, and both sexes, genes that reached the
132 previously described expression thresholds in at least one tissue in at least one sex were retained.
133 This assures that the DEGs detected in the sex-specific and combined analyses are not due to
134 filtering.

135 **Differential expression analysis**

136 For differential expression (DE) analysis, filtered, untransformed read count data were quantile
137 normalized and logCPM transformed with *voom* (27). From the TCGA LIHC dataset, paired
138 tumor and tumor-adjacent samples were available for 22 females and 28 males. From the GTEx
139 liver dataset, 91 male and 45 female samples were used in the DE analysis. A multi-factor design
140 with sex and tissue type as predictor variables were used to fit the linear model.
141 *duplicateCorrelation* function was used to calculate the correlation between measurements made
142 on the same subject, and this inter-subject correlation was accounted for in the linear modeling.

143 DEGs between comparisons were identified using the *limma*/*voom* pipeline (27) by computing
144 empirical Bayes statistics with *eBayes*. An FDR-adjusted *p*-value threshold of 0.01 and an
145 absolute $\log_2$ fold-change (FC) threshold of 2 were used to select significant DEGs.

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

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

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

160 We further analyzed the sex-shared and sex-specific tumor vs. tumor-adjacent DEGs as well as
161 the sex-specific eQTL target genes (eGenes) to identify sex-shared and sex-specific pathways
162 driving HCC etiology. We used the Gene Ontology Consortium's web tool (28) which utilizes
163 the PANTHER classification system (29) and a binomial test (30) to determine whether a list of
164 genes is overrepresented in pathways in regards to molecular functions, biological processes, and
165 cellular functions. Fisher's exact test was used to calculate overrepresentation scores. We used
166 the *NetworkAnalyst* webtool (31), which utilizes a hypergeometric test to compute *p*-values for
167 the overrepresentation of genes in regards to KEGG and Reactome pathways. An FDR-adjusted
168 *p*-value threshold of 0.01 was used to select significantly overrepresented GO terms and
169 canonical pathways.

170 **eQTL analysis**

171 We used eQTL analyses to detect germline genetic effects on tumor gene expression. Similar to
172 the DE analysis, we utilized combined and sex-stratified analyses to detect sex-shared and sex-
173 specific effects. Germline genotypes and tumor gene expression data from 248 male and 119
174 female donors in the TCGA LIHC cohort were used in the eQTL analysis. Filtered count data
175 was normalized by fitting the FPKM values of each gene and sample to the quantiles of the
176 normal distribution. To account for technical confounders and population structure, 10 *de novo*
177 PEER factors and three genotype principal components were used as covariates (Supplementary
178 Methods). *Cis*-acting (proximal) eQTLs were detected by linear regression as implemented in
179 *QTLtools* v.1.1 (32). Variants within 1Mb of the gene under investigation were considered for

180 testing. We used the permutation pass with 10,000 permutations to get adjusted *p*-values for
181 associations between the phenotypes and the top-variants in *cis*. FDR-adjusted *p*-values were
182 calculated to correct for multiple phenotypes tested, and an adjusted *p*-value threshold of 0.01
183 was used to select significant associations. To allow the comparison of effect sizes of sex-
184 specific and sex-shared eQTLs across the sexes, effects of each variant located within the 1Mb
185 *cis*-window were obtained using the *QTLtools* nominal pass.

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

191 Results

192 Sex-specific patterns of gene expression in HCC

193 We identified sex-differences in gene expression in healthy liver (GTEx; 21 sex-biased genes),
194 tumor-adjacent tissue (TCGA LIHC; 72), and HCC (TCGA LIHC; 24) to characterize the shared
195 and unique sex-differences that may drive the observed sex-biases in HCC occurrence and
196 etiology (Fig. 1, Supplementary Tables S2-4). X-linked *XIST* and Y-linked genes were expressed
197 in a sex-biased way across all tissues. While sex-biased gene expression in healthy and tumor-
198 adjacent tissues may contribute to the sex-differences in cancer occurrence, sex-biased
199 expression in tumors is suggestive of distinct molecular etiologies of male and female HCC. We
200 identified 13 genes, including 7 protein-coding genes (*CD24*, *CYP4F22*, *DTX1*, *HRCT1*,
201 *HSD11B1*, *PIGR*, *ZCCHC16)* and 6 non-coding RNAs or pseudogenes (*ADORA2BP1*, *CTD-*
202 *2325A15.3*, *RP11-431J24.2*, *RP11-495P10.5*, *RP11-495P10.7*, *RP11-495P10.9*) that show sex-
203 differences in expression in HCC, but not in tumor-adjacent tissue or healthy liver, even with a
204 relaxed significance threshold (Fig. 1A). Notably, Notch-regulating *DTX1* (Fig. 1B) and signal
205 transducer *CD24* were downregulated in male HCC.

206 To further examine the sex-shared and sex-specific mechanisms driving HCC etiology, we
207 detected DEGs between tumor and tumor-adjacent samples in males and females, as well as in

208 the combined analysis of both sexes. Dimensionality reduction of gene expression data shows
209 that variation among the tumor and tumor-adjacent samples is driven by tissue type and sex (Fig.
210 1C). When inspecting the tumor samples only, the first dimension is driven by sex
211 (Supplementary Fig. S1). In the combined analysis of male and female samples, we detected 581
212 tumor vs. tumor-adjacent DEGs (Supplementary Table S5). In male- and female-specific
213 analyses, we detected 606 and 416 tumor vs. tumor-adjacent DEGs, respectively (Supplementary
214 Tables S6-7). Out of the total of 756 unique DEGs, 77% were shared between the sexes. We
215 identified 83 female-specific and 92 male-specific tumor vs. tumor-adjacent DEGs. Notably,
216 substantially more DEGs were detected in sex-specific analyses than in the unstratified analysis
217 (Fig. 1D). Specifically, DEGs that showed different magnitudes in fold change between the sexes
218 (based on ANOVA/Kruskal-Wallis tests) were detected in the sex-specific analyses, while DEGs
219 with similar fold changes across all comparisons were detected in the combined analysis as well
220 as the sex-specific analyses (Supplementary Fig. S2). Sex-shared DEGs that were only detected
221 in the combined analysis, and not in the sex-specific analyses, showed a large variance in
222 expression and, due to limited power, were not detected as statistically significant DEGs in sex-
223 specific analyses (Supplementary Fig. S2). Tumor-infiltrating immune cells may produce
224 spurious signals in DE analyses, which is evident from the detection of various immunoglobulin
225 genes in tumor vs. tumor-adjacent comparisons (Supplementary Tables S5-7). However, cell
226 content analyses based on gene expression data did not exhibit sex-differences in the prevalence
227 of tumor-infiltrating immune cells in this data (Supplementary Fig. S3), and thus such spurious
228 signals are unlikely to affect male-female comparisons. Furthermore, male and female cases did
229 not differ in terms of tumor pathology (Supplementary Table S1). The observed differences in
230 gene expression are thus likely to reflect actual sex-differences rather than confounding
231 differences in sample characteristics or composition.

232 To put these results in a broader context, we analyzed the male- and female-specific DEGs
233 (tumor vs. tumor-adjacent) for the overrepresentation of functional pathways and canonical
234 networks. We found that the sex-shared and sex-specific DEGs were enriched in pathways
235 relevant to oncogenesis and cancer progression (Supplementary Tables 8-10). Most of the sex-
236 specific top pathways were non-overlapping, indicating that male and female HCC are driven by
237 different mechanisms and processes (Fig. 1E-F).

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

239 To further investigate the mechanisms of sex-difference in HCC etiology, we used eQTL
240 analyses to detect germline genetic effects on tumor gene expression in both the joint and sex-
241 stratified analyses (Fig. 2A). We detected 1,204, 761, and 245 eQTLs in the combined, male-
242 specific, and female-specific analyses, respectively (Supplementary Tables S11-13). As
243 expected, genomic annotations show that most eQTLs are located on non-coding regions
244 (Supplementary Methods, Fig. 2B). Consistent with previous reports, most *cis*-eQTLs were
245 located near transcription start sites (TSSs), with 63% of all eQTLs across the combined and sex-
246 specific analyses being located within 20kb of TSSs. On average, 384 variants were tested per
247 gene. 31% of the unique shared and sex-specific cis-eQTLs in HCC were also identified as
248 eQTLs in the liver data in the GTEx project, indicating shared tissue origin. Out of the total of
249 1,595 unique associations, 75.7% were shared between the sexes. We detected 295 male-specific
250 and 92 female-specific eQTLs. Since these associations were not detected in the unstratified
251 analysis, they are likely not a result of differential power to detect associations due to different
252 sample sizes, but exhibit effect modification by sex. Sex-specific associations exhibited
253 differences in effect size between the sexes (based on ANOVA/Kruskal-Wallis tests), and the
254 sex-specific effect is diluted in the combined analysis (Supplementary Fig. S6). Sex-shared large
255 effect eQTLs were detected in sex-specific and combined analyses, and, due to the larger sample
256 size, sex-shared low-effect eQTLs are detected in the combined analysis only (Supplementary
257 Fig. S6).

258 We detected 27 shared eGenes that were regulated by independent regulatory variants in males
259 and females. This could be due to actual biological differences in gene regulation, or due to
260 technical constraints, in particular, missing genotypes in one sex affecting the permutation
261 scheme to select the top-variant for each target gene. To overcome this and to detect high
262 confidence instances of differential gene regulation between the sexes, we further examined the
263 sex-shared and sex-specific eGenes: we found 24 genes that are under germline regulatory
264 control in only male HCC (Fig. 2C), including *POGLUT1*, which is an essential regulator of
265 Notch signaling (Fig. 2D). No genes were found to be under regulatory control in females only,
266 likely due to reduced power to detect associations in females (Supplementary Fig. S5). Male-
267 specific eGenes were overrepresented in pathways related to cell cycle, apoptosis, and cancer

268 (Supplementary Table S17). Concordant with previous studies (14,33), none of the male-specific
269 eGenes were differentially expressed between male and female HCC, indicating that the male-
270 specific eQTLs are not a result of differences in overall gene expression levels between males
271 and females, but are likely to arise from factors such as differential chromatin accessibility or
272 transcription factor activity. The observation that the sex-biased autosomal genes in tumors did
273 not harbor significant *cis*-eQTLs (Supplementary Table S18) also suggests that the sex-biased
274 gene expression is a result of trans-effects, e.g. sex-chromosomal effects on autosomal gene
275 expression, or, more widely, a result of sex as a biological variable, e.g. hormonal effects.

276 Discussion

277 Distinct molecular etiologies of male and female HCC

278 It is well established that patterns of gene expression vary between the sexes across different
279 tissues. Previous studies have confounded these differences with those which may be driving
280 etiological differences between male and female tumors. For example, Yuan et al. previously
281 reported extensive sex-biased signatures in gene expression in HCC and other strongly sex-
282 biased cancers (11). While they identified immunity and cancer-associated enriched pathways
283 based on sex-biased genes detected in HCC tumors, their approach was limited as it did not
284 include the examination of tumor-adjacent tissues. From the results presented here, we are able
285 to distinguish the differences detected in comparisons of male and female HCC from those
286 reflecting sex-differences in the healthy liver or in tumor-adjacent tissue, as well as to detect
287 genes that are dysregulated in HCC in a sex-shared or sex-specific manner.

288 We characterized differences in gene expression between male and female HCC cases. Notably,
289 sex-differences in gene expression were the largest in the tumor-adjacent tissue, with 72 genes
290 (including 20 autosomal genes) being expressed in a sex-biased way. These sex-differences point
291 to distinct mechanisms underlying HCC oncogenesis between the sexes, and may partially
292 underlie the observed sex-biases in HCC occurrence and onset. We detected 13 genes that were
293 expressed in a sex-biased way in HCC tumors, but not in healthy or tumor-adjacent liver tissues.
294 Some of these genes are of particular interest in the context of HCC: Notch-regulating *DTX1*,
295 found here to be underexpressed in males compared to females, has been identified as a putative
296 tumor suppressor gene in head and neck squamous cell carcinoma (34). Another female-biased

297 gene detected here, *CD24*, has a crucial role in T cell homeostasis and autoimmunity (35). The
298 opposing roles of *CD24* expression in cancer and autoimmune diseases raise interesting
299 questions on the role of sex-differences in immunity underlying sex-differences in cancer. Future
300 studies will focus on better understanding the differential regulation of immune functions
301 between the sexes, and how these differences contribute to the observed biases in disease
302 occurrence and etiology.

303 By sex-specific analyses of matched tumor and tumor-adjacent samples, we detected genes that
304 are uniquely dysregulated in male and female HCC. Further examination of these genes revealed
305 sex-differences in the pathway and network activation, indicating that the molecular etiologies of
306 male and female HCC are partly driven by distinct functional pathways. Males and females
307 differed in oncogenic processes with females showing RAS pathway enrichment while males
308 showed PI3K/AKT pathway enrichment (Fig. 1E, Supplementary Tables 9-10). While both sexes
309 exhibited an enrichment of genes involved in cell cycle, apoptosis and p53 signaling, only males
310 showed activation of pathways regulating to natural killer cell cytotoxicity, and females showed
311 an activation of pathways involved in NF-kB (nuclear factor kappa-light-chain-enhancer of
312 activated B cell) activation. Furthermore, female-specific DEGs were enriched in genes involved
313 in innate immunity Toll-like receptor (TLR) signaling, while male-specific DEGs exhibited an
314 enrichment of genes involved in adaptive immunity and Notch signaling. TLR signaling has a
315 crucial role in the regulation of the immune system by evoking an inflammatory response (36).
316 As TLR and Notch signaling pathways are notable targets for anti-cancer and anti-metastasis
317 therapies (37,38), the results presented here have implications to the targeted treatment of male
318 and female HCC.

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

321 Sex-specific regulatory functions may underlie sex-differences in cancer etiology, progression,
322 and outcome. We detected sex-differences in the germline genetic regulation of tumor gene
323 expression in HCC, including 24 genes that were under germline regulatory control only in male
324 HCC (Fig. 2). Functional annotations of these male-specific eGenes provide insight into possible
325 regulatory mechanisms contributing to the observed male-bias in HCC and sex-differences in

326 HCC etiology. Protein O-glucosyltransferase 1 (*POGLUT1*) was found to be under germline
327 regulation in male HCC, but not in female HCC or in the joint analysis of both sexes (Fig. 3D).
328 The eQTL associated with *POGLUT1* is located on a promoter region of its target
329 (Supplementary Table S15). *POGLUT1* is an enzyme that is responsible for O-linked
330 glycosylation of proteins. Altered glycosylation of proteins has been observed in many cancers
331 (39,40), including liver cancer (41,42). *POGLUT1* is an essential regulator of Notch signaling
332 and is likely involved in cell fate and tissue formation during development. Genes involved in
333 Notch signaling were also found to be expressed in a sex-biased way in HCC tumors and highly
334 overrepresented among the male-specific DEGs detected in the tumor vs. tumor-adjacent
335 comparison, showing that sex-specific eQTLs and sex-specific dysregulated genes converge in
336 canonical pathways. While the Notch signaling pathway was not detected as statistically
337 significantly overrepresented (FDR-adj. p-value < 0.01) among the 24 male-specific eGenes,
338 these genes exhibited an enrichment of genes belonging to the KEGG Notch signaling pathway
339 (4 out of 47 genes detected, 0.542 expected, FDR-adj. p=0.06) and Reactome term "NOTCH1
340 Intracellular Domain Regulates Transcription" (4 out of 50 genes detected, 0.412 expected,
341 FDR-adj. p=0.05). These results point to a possible role of Notch in the development of HCC in
342 males. Notch signaling is of particular interest in the context of HCC, as it is involved in liver
343 development (43,44), the development of sexually dimorphic traits (45), and tumorigenesis (46).
344 The role of Notch signaling in HCC may differ between early and late-stage tumors and among
345 molecular subtypes, and further studies are necessary to understand the possible oncogenic
346 properties of Notch among HCC subtypes and between the sexes.

347 In summary, we discovered differential genetic and regulatory functions in HCC tumors between
348 the sexes. This work provides a framework for future studies on sex-biased cancers. We are
349 calling for the generation of suitable large datasets and further studies to identify and validate
350 sex-specific genetic effects on tumor gene expression and its consequences in HCC and other
351 sex-biased cancers.

352 Acknowledgments

353 This study was supported by ASU Center for Evolution and Medicine postdoctoral fellowship
354 and the Marcia and Frank Carlucci Charitable Foundation postdoctoral award from the Prevent

Cancer Foundation for HMN, ASU School of Life Sciences startup funds for MAW, and ASU Center for Evolution and Medicine Venture funds. We thank Dr. Nicholas Banovich and three anonymous reviewers for their feedback on previous versions of this manuscript.

Citations

1. Clocchiatti A, Cora E, Zhang Y, Dotto GP. Sexual dimorphism in cancer. Nat Rev Cancer. 2016;16:330.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin [Internet]. 2018; Available from: http://dx.doi.org/10.3322/caac.21492
3. Wisnivesky JP, Halm EA. Sex differences in lung cancer survival: do tumors behave differently in elderly women? J Clin Oncol. 2007;25:1705–12.
4. OuYang P-Y, Zhang L-N, Lan X-W, Xie C, Zhang W-W, Wang Q-X, et al. The significant survival advantage of female sex in nasopharyngeal carcinoma: a propensity-matched analysis. Br J Cancer. 2015;112:1554–61.
5. Li CH, Haider S, Shiah Y-J, Thai K, Boutros PC. Sex Differences in Cancer Driver Genes and Biomarkers. Cancer Res. 2018;78:5527–37.
6. Wands J. Hepatocellular carcinoma and sex. N Engl J Med. 2007;357:1974–6.
7. Petrick JL, Braunlin M, Laversanne M, Valery PC, Bray F, McGlynn KA. International trends in liver cancer incidence, overall and by histologic subtype, 1978-2007. Int J Cancer. 2016;139:1534–45.
8. Ladenheim MR, Kim NG, Nguyen P, Le A, Stefanick ML, Garcia G, et al. Sex differences in disease presentation, treatment and clinical outcomes of patients with hepatocellular carcinoma: a single-centre cohort study. BMJ Open Gastroenterol. 2016;3:e000107.
9. Gilks WP, Abbott JK, Morrow EH. Sex differences in disease genetics: evidence, evolution, and detection. Trends Genet. 2014;30:453–63.
10. Morrow EH. The evolution of sex differences in disease. Biol Sex Differ. 2015;6:5.
11. Yuan Y, Liu L, Chen H, Wang Y, Xu Y, Mao H, et al. Comprehensive Characterization of Molecular Differences in Cancer between Male and Female Patients. Cancer Cell. 2016;29:711–22.

385 12. Gong J, Mei S, Liu C, Xiang Y, Ye Y, Zhang Z, et al. PancanQTL: systematic identification
386 of cis-eQTLs and trans-eQTLs in 33 cancer types. Nucleic Acids Res. 2018;46:D971–6.
387 13. Behrens G, Winkler TW, Gorski M, Leitzmann MF, Heid IM. To stratify or not to stratify:
388 power considerations for population-based genome-wide association studies of quantitative
389 traits. Genet Epidemiol. 2011;35:867–79.
390 14. Dimas AS, Nica AC, Montgomery SB, Stranger BE, Raj T, Buil A, et al. Sex-biased genetic
391 effects on gene regulation in humans. Genome Res. 2012;22:2368–75.
392 15. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a
393 Shared Vision for Cancer Genomic Data. N Engl J Med. 2016;375:1109–12.
394 16. Webster TH, Couse M, Grande BM, Karlins E, Phung TN, Richmond PA, et al. Identifying,
395 understanding, and correcting technical biases on the sex chromosomes in next-generation
396 sequencing data [Internet]. bioRxiv. 2018 [cited 2018 Dec 16]. page 346940. Available from:
397 https://www.biorxiv.org/content/early/2018/07/18/346940
398 17. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data [Internet].
399 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. 2010. Available from:
400 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
401 18. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
402 Bioinformatics. 2014;30:2114–20.
403 19. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory
404 requirements. Nat Methods. 2015;12:357–60.
405 20. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning
406 sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
407 21. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
408 Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA
409 sequencing data. Genome Res. 2010;20:1297–303.
410 22. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for
411 variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet.
412 2011;43:491–8.
413 23. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et
414 al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best
415 practices pipeline. Curr Protoc Bioinformatics. 2013;43:11.10.1–33.

416 24. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC
417 Genome Browser Database: update 2006. Nucleic Acids Res. 2006;34:D590–8.
418 25. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity
419 landscape. Genome Biol. 2017;18:220.
420 26. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
421 expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
422 27. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis
423 tools for RNA-seq read counts. Genome Biol. 2014;15:R29.
424 28. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11:
425 expanded annotation data from Gene Ontology and Reactome pathways, and data analysis
426 tool enhancements. Nucleic Acids Res. 2017;45:D183–9.
427 29. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with
428 the PANTHER classification system. Nat Protoc. 2013;8:1551–66.
429 30. Cho RJ, Campbell MJ. Transcription, genomes, function. Trends Genet. 2000;16:409–15.
430 31. Xia J, Gill EE, Hancock REW. NetworkAnalyst for statistical, visual and network-based
431 meta-analysis of gene expression data. Nat Protoc. 2015;10:823–44.
432 32. Delaneau O, Ongen H, Brown AA, Fort A, Panousis N, Dermitzakis E. A complete tool set
433 for molecular QTL discovery and analysis. bioRxiv [Internet]. 2016; Available from:
434 http://biorxiv.org/content/early/2016/08/10/068635.abstract
435 33. Kukurba KR, Parsana P, Balliu B, Smith KS, Zappala Z, Knowles DA, et al. Impact of the X
436 Chromosome and sex on regulatory variation. Genome Res. 2016;26:768–77.
437 34. Gaykalova DA, Zizkova V, Guo T, Tiscareno I, Wei Y, Vatapalli R, et al. Integrative
438 computational analysis of transcriptional and epigenetic alterations implicates DTX1 as a
439 putative tumor suppressor gene in HNSCC. Oncotarget. 2017;8:15349–63.
440 35. Liu Y, Zheng P. CD24: a genetic checkpoint in T cell homeostasis and autoimmune diseases.
441 Trends Immunol. 2007;28:315–20.
442 36. Gu J, Liu Y, Xie B, Ye P, Huang J, Lu Z. Roles of toll-like receptors: From inflammation to
443 lung cancer progression. Biomed Rep. 2018;8:126–32.
444 37. Li L, Tang P, Li S, Qin X, Yang H, Wu C, et al. Notch signaling pathway networks in cancer
445 metastasis: a new target for cancer therapy. Med Oncol. 2017;34:180.

446 38. Braunstein MJ, Kucharczyk J, Adams S. Targeting Toll-Like Receptors for Cancer Therapy.
447 Target Oncol. 2018;13:583–98.
448 39. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. Nat Rev
449 Cancer. 2015;15:540–55.
450 40. Stowell SR, Ju T, Cummings RD. Protein glycosylation in cancer. Annu Rev Pathol.
451 2015;10:473–510.
452 41. Mehta A, Herrera H, Block T. Chapter Seven - Glycosylation and Liver Cancer. In: Drake
453 RR, Ball LE, editors. Advances in Cancer Research. Academic Press; 2015. page 257–79.
454 42. Liang K-H, Yeh C-T. O-glycosylation in liver cancer: Clinical associations and potential
455 mechanisms. Liver Research. 2017;1:193–6.
456 43. McDaniell R, Warthen DM, Sanchez-Lara PA, Pai A, Krantz ID, Piccoli DA, et al. NOTCH2
457 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway.
458 Am J Hum Genet. 2006;79:169–73.
459 44. Oda T, Elkahloun AG, Pike BL, Okajima K, Krantz ID, Genin A, et al. Mutations in the
460 human Jagged1 gene are responsible for Alagille syndrome. Nat Genet. 1997;16:235–42.
461 45. Deng H, Jasper H. Sexual Dimorphism: How Female Cells Win the Race. Curr Biol.
462 2016;26:R212–5.
463 46. Ntziachristos P, Lim JS, Sage J, Aifantis I. From fly wings to targeted cancer therapies: a
464 centennial for notch signaling. Cancer Cell. 2014;25:318–34.

465 ## Supporting Information

466 **Supporting Information contains Supplementary Methods, six figures, and 18 tables.**

467 **Fig S1.** A multi-dimensional scaling plot of the TCGA LIHC tumor samples of each sex.
468 Euclidean distances between samples were calculated based on 100 genes with the largest
469 standard deviations between samples.

470 **Fig S2.** Absolute $\log_2$-fold changes of DEGs detected from tumor vs. tumor-adjacent
471 comparisons in the sex-specific analyses and combined analysis of both sexes. Absolute $\log_2$-fold
472 changes are given for female samples, male samples, and across all samples. Global *p*-values for
473 ANOVA are shown for each DEG type. Adjusted *p*-values based on Kruskal-Wallis tests are
474 shown for each pairwise comparison.

475 **Fig S3.** Cellular content of TCGA LIHC tumor samples based on cell-type enrichment analysis
476 from bulk-RNAseq data. Heatmap is based on scaled and mean centered estimates of cell type
477 proportions.

478 **Fig S4.** PEER factors associated with the continuous and categorical sample and patient
479 covariates. **A-B:** R^2 and FDR-adjusted *p*-values between PEER factors and continuous technical
480 and biological covariates. **C-D:** ANOVA test statistic and FDR-adjusted *p*-values between PEER
481 factors and categorical technical and biological covariates.

482 **Fig S5.** Estimation of statistical power in the combined (grey), male-specific (blue), and female-
483 specific (red) eQTL analyses with a *p*-value level 0.01 and 384 variants. Increased power in the
484 combined analysis allows the detection of sex-shared low-effect eQTLs.

485 **Fig S6.** Absolute effect sizes of sex-shared and sex-specific eQTLs in males, females, and the
486 whole study sample. Due to the larger sample size, sex-shared low-effect eQTLs are only
487 detected as significant in the combined analysis. "Sex-shared" facet shows eQTLs that were
488 identified as significant in both sex-specific analyses and in the combined analysis of both sexes.
489 Sex-specific eQTLs exhibit a larger effect in one sex than the other, and the effect is diluted in
490 the combined analysis. Sex-shared large effect eQTLs can be detected in sex-specific and
491 combined analyses. Global *p*-values for ANOVA are shown for each DEG type. Adjusted *p*-
492 values based on Kruskal-Wallis tests are shown for each pairwise comparison.

493 **Table S1.** Clinical characteristics of male and female patients in the TCGA LIHC cohort.
494 Bonferroni adjusted *p*-value threshold for Fisher's exact test = 0.0006.

495 **Table S2.** Sex-biased gene expression in the GTEx liver tissue samples with an FDR-adjusted *p*-
496 value < 0.1 and absolute $\log_2$ fold-change > 0. N males = 91, N females = 45.

497 **Table S3.** Sex-biased gene expression in the TCGA LIHC tumor-adjacent samples with an FDR-
498 adjusted *p*-value < 0.1 and an absolute $\log_2$ fold-change > 0. N males = 28, N females = 22.

499 **Table S4.** Sex-biased gene expression in TCGA LIHC tumor samples with an FDR-adjusted *p*-
500 value < 0.01 and an absolute $\log_2$ fold-change > 2. N males = 28, N females = 22. Empirical *p*-
501 value based on sex-label shuffling with 1000 permutations.

502 **Table S5.** Differentially expressed genes between matched tumor and tumor-adjacent samples in
503 the combined analysis of male and female samples with an FDR-adjusted *p*-value < 0.01 and an
504 absolute $\log_2$ fold-change > 2. Each tissue N=50, N female = 22, N male = 28.

505 **Table S6.** Differentially expressed genes between matched male tumor and tumor-adjacent
506 samples with an FDR-adjusted *p*-value < 0.01 and an absolute $\log_2$ fold-change > 2. N of sample
507 pairs = 28.

508 **Table S7.** Differentially expressed genes between matched female tumor and tumor-adjacent
509 samples with an FDR-adjusted *p*-value < 0.01 and an absolute $\log_2$ fold-change > 2. N of sample
510 pairs = 22.

511 **Table S8.** Overrepresented GO terms and canonical pathways in the sex-shared tumor vs. tumor-
512 adjacent DEGs. Significant terms and pathways were selected based on an FDR-adjusted *p*-value
513 threshold 0.01.

514 **Table S9.** Overrepresented GO terms and canonical pathways in the male-specific tumor vs.
515 tumor-adjacent DEGs. Significant terms and pathways were selected based on an FDR-adjusted
516 *p*-value threshold 0.01.

517 **Table S10.** Overrepresented GO terms and canonical pathways in the female-specific tumor vs.
518 tumor-adjacent DEGs. Significant terms and pathways were selected based on an FDR-adjusted
519 *p*-value threshold 0.01.

520 **Table S11.** *cis*-eQTLs detected in the combined analysis of both sexes. N=367 (N males = 248,
521 N females = 119). Significant eQTLs were selected based on an FDR-adjusted *p*-value threshold
522 0.01.

523 **Table S12.** *cis*-eQTLs detected in the male-specific analysis. N=248. Significant eQTLs were
524 selected based on an FDR-adjusted *p*-value threshold 0.01.

525 **Table S13.** *cis*-eQTLs detected in the female-specific analysis. N=119. Significant eQTLs were
526 selected based on an FDR-adjusted *p*-value threshold 0.01.

527 **Table S14.** Genomic annotations of eQTLs detected in the combined analysis of both sexes.

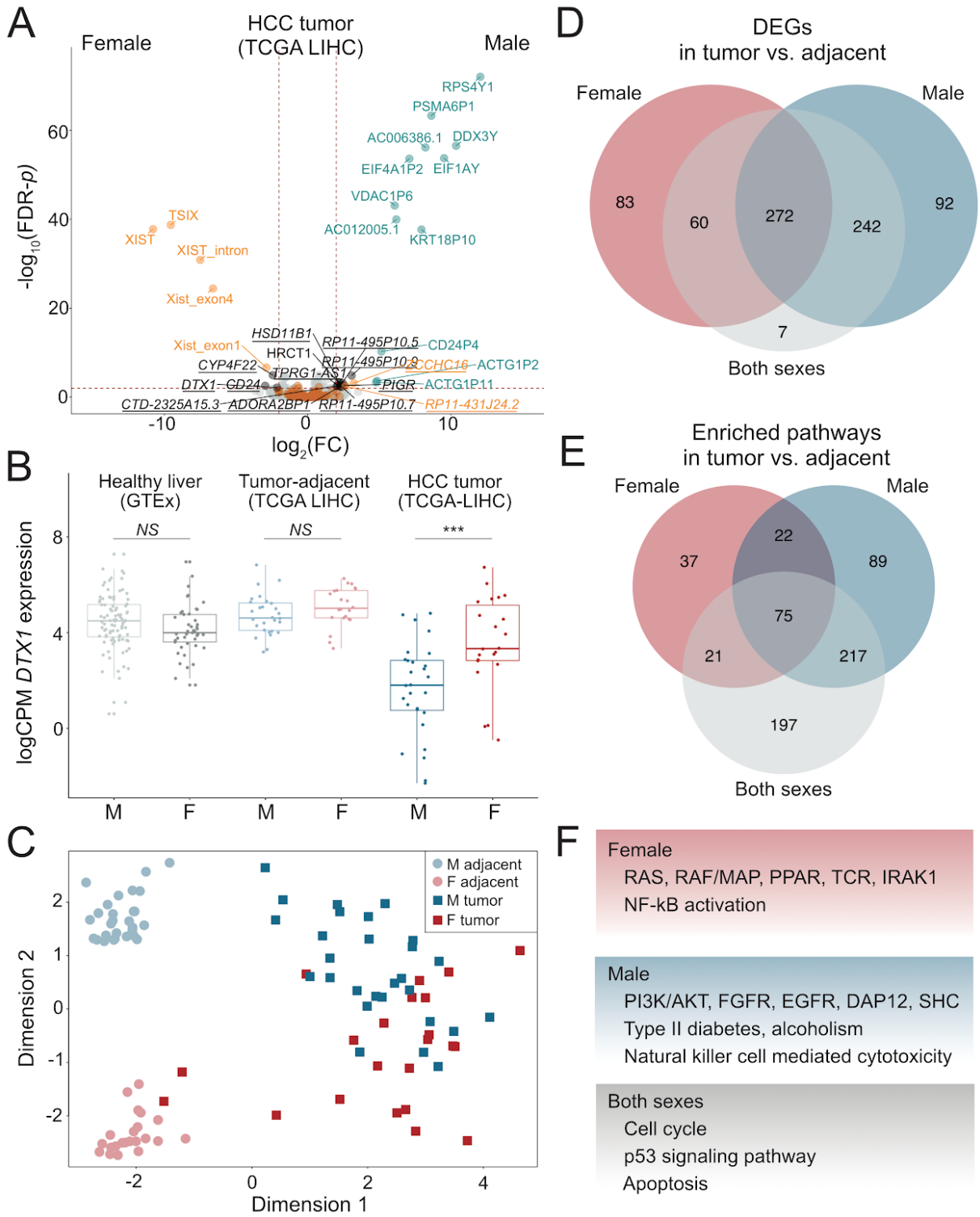
528 **Table S15.** Genomic annotations of eQTLs detected in the male-specific analysis.

529 **Table S16.** Genomic annotations of eQTLs detected in the female-specific analysis.

530 **Table S17.** Overrepresented canonical pathways in the male-specific eQTL target genes.
531 Significant terms and pathways were selected based on an FDR-adjusted *p*-value threshold 0.01.

532 **Table S18.** Top-variants associated with autosomal genes that were expressed in a sex-biased
533 way in HCC tumor samples.

534 Figure Captions



535

536 **Fig. 1. Patterns of gene expression and molecular etiologies of male and female HCC. A:**
537 Sex-biased gene expression in HCC. A volcano plot of DEGs between male (N=28) and female
538 (N=22) HCC tumor samples. X-linked genes are indicated in orange, Y-linked in teal, and

539 autosomal in black. Significant genes were selected based on an FDR-adjusted *p*-value threshold
540 of 0.01 and absolute $\log_2$(FC) threshold of 2. Multiple transcripts of the long non-coding RNA
541 *XIST* are independently expressed. Genes that were not expressed in a sex-biased way in healthy
542 liver (GTEx) or in the tumor-adjacent tissues are indicated in italics and underlined. **B:** An
543 example of a gene exhibiting a sex-bias in HCC but not in healthy liver or tumor-adjacent
544 tissues. *DTX1* expression in log(CPM) is shown for male and female samples in each tissue. **C:**
545 A multi-dimensional scaling plot of the paired TCGA LIHC tumor and tumor-adjacent samples
546 of each sex. Euclidean distances between samples were calculated based on 100 genes with the
547 largest standard deviations between samples. Tissue type (dimension 1) and sex (dimension 2)
548 drive the overall patterns of gene expression in HCC. **D:** Venn-diagram of the overlap of DEGs
549 in the sex-specific and combined analysis of matched tumor and tumor-adjacent samples.
550 Substantially more DEGs were identified in the sex-specific analyses. **E:** Sex-specific and sex-
551 shared DEGs were analyzed for the overrepresentation of pathways and canonical networks.
552 Sex-specific patterns of pathway enrichment point to differential processes driving the etiology
553 of male and female HCC. **F:** Examples of sex-specific and sex-shared pathways.

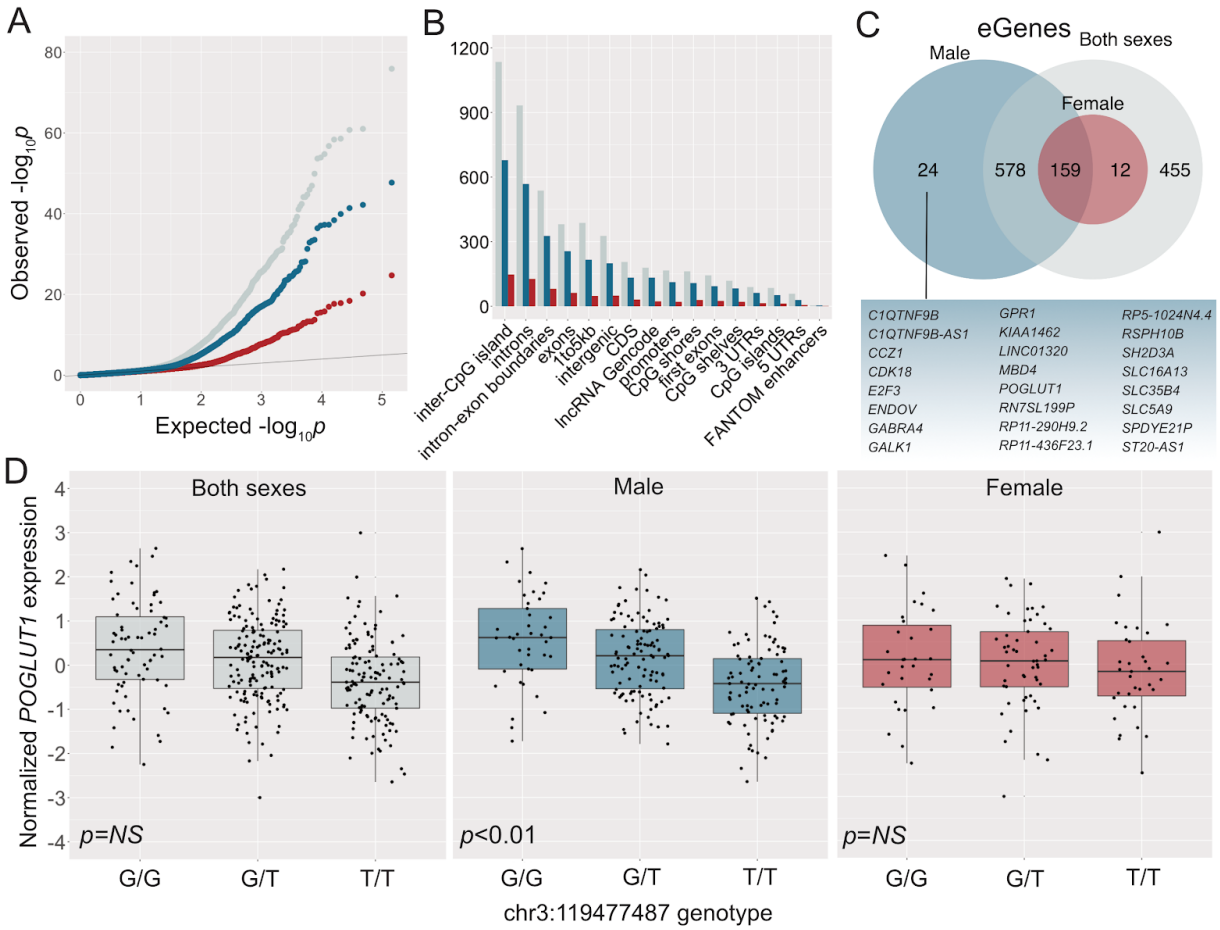


Fig. 2. Sex-specific genetic effects on tumor gene expression in HCC. A: QQ-plot of eQTL associations in the combined analysis of both sexes (grey), male-specific analysis (blue), and female-specific analysis (red). **B:** Genomic annotations of eQTLs in the combined analysis of both sexes, male-specific analysis, and female-specific analysis. **C:** Overlap of eGenes detected in combined and sex-specific analyses. **D:** An example of a male-specific eQTL. *POGLUT1* expression in tumors is modulated by a germline variant in *cis* in male HCC, but not in female HCC nor in the combined analysis of both sexes, indicating effect modification by sex.