# Sex-specific regulatory mechanisms underlying Hepatocellular Carcinoma

Heini M Natri[1,*], Melissa Wilson Sayres[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. Most HCC studies, to date, have failed to explore the sex-specific effects in gene regulatory functions. Here we have characterized the regulatory functions underlying HCC tumors in sex-stratified and combined analyses. By sex-specific analyses of differential expression of tumor and tumor adjacent samples, we uncovered etiologically relevant genes, pathways and canonical 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. We discovered eQTLs overlapping HCC GWAS loci, providing regulatory mechanisms connecting these loci to HCC risk. Furthermore, we discovered genes under germline regulatory control that alter survival in patients with HCC, including some that exhibited differential effects on survival between the sexes. Overall, our results provide new insight into the role of genetic regulation of transcription in modulating sex differences in HCC occurrence and outcome and provide a framework for future studies on sex-biased cancers.


## Author Summary

Sex differences in cancer occurrence and mortality are evident across tumor types, and targeted approaches to the treatment male and female tumors 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. Here, we have characterized the regulatory functions underlying male and female HCC. We show that while HCC shares commonalities across the sexes, it shows demonstrable regulatory differences that could be critical in terms of prevention and treatment: we detected differential activation of pathways involved in innate and adaptive immunity, as well as differential genetic effects on gene expression, between the sexes. Furthermore, we have detected regulatory variants overlapping known HCC GWAS risk loci, providing regulatory mechanisms connecting these risk variants to HCC. Finally, we discovered genes under germline regulatory control altering the overall survival of patients with HCC, including some that showed differential effects between the sexes. Overall, our findings create a paradigm 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 behaviors and environmental exposure may explain some portion of the sex-bias, cellular and molecular differences are also likely to be important. The sexes differ in their endocrinological profile, immunological functions and genetic makeup [3–5]. However, sex differences are rarely considered in the development of cancer therapies, and the contribution from sex chromosome variation to tumor etiology remains poorly understood. Sex-biased gene expression and regulatory functions may underlie differences between the sexes in disease prevalence and severity [6,7]. Analyses of sex-specific regulation of gene expression are essential for

understanding sources of sex difference in cancer incidence, as well as sex-specific mechanisms affecting tumor etiology, disease progression and outcome.

Hepatocellular carcinoma (HCC) exhibits sex-bias in occurrence, with a male-to-female incidence ratio between 1.3:1 and 5.5:1 across populations [8,9]. HCC is the second leading cause of cancer mortality worldwide, accounting for 8.2% of all cancer deaths [10]. HCC risk is influenced by genetic susceptibility, environmental factors such as oncogenic viral infections, metabolic syndrome, and alcohol use, and shaped by numerous biological processes, resulting in a high degree of genetic and transcriptional heterogeneity [11]. HCC incidence in the US has doubled in the last 3 decades, attributable to increased rates of obesity [9].

While sex-bias in HCC is partly attributed to sex-specific differences in risk behaviors and environmental exposure, the relationship of these factors and sex as a biological variable has not been systematically investigated. Previously, sex differences in HCC were explored in conjunction with 12 additional Cancer Genome Atlas (TCGA) tumor types by Yuan et al [12]. They discovered extensive sex-biased signatures in gene expression in HCC and other strongly sex-biased cancers.

HCC shows evidence of molecular subtypes, though their role in sex-bias has yet to be explored. More specifically, clustering analyses of gene expression data have produced a characterization of four distinct HCC subtypes that differ in terms of gene expression, pathway enrichment, and median survival [13]. A study utilizing a weighted co-expression network analysis method discovered hub genes associated with HCC outcome [14]. On the other hand, genetic variants affecting the expression of *TNFRSF10 [15]*, *PAX8 [16],* and *PVT1 [17]* have been found to alter HCC disease progression and outcome, highlighting the role of regulatory variants in HCC. How these progression altering subtypes or variants are related to HCC sex-bias has not been examined.

Sex-specific analyses can reveal genetic and regulatory mechanisms of sexual dimorphism in HCC susceptibility, progression, and mortality. Targeted approaches to the treatment of male and female HCC will require the characterization and understanding of the fundamental biological mechanisms that differentiate them. Here, we analyzed data from The Genotype-Tissue Expression (GTEx) project and The Cancer Genome Atlas (TCGA) to examine the sex-specific patterns of gene expression and regulation in HCC and healthy liver tissue. We find etiologically meaningful differences in regulatory functions in HCC between males and females, providing insight into the mechanisms underlying sex-bias in cancer. Additionally, we observed genes under germline regulatory control that alter survival in sex-biased way in patients with HCC.

# Results

## Sex-specific patterns of gene expression HCC

We found sex-differences in gene expression in normal liver, tumor adjacent, and tumor tissues (Fig 1). We found 13 genes showing a consistent sex-bias in expression in liver, HCC adjacent tissue and HCC. Notably, we find that X-inactivation and X-linked dosage compensation appear to be functioning in the healthy liver, HCC adjacent and HCC tissues. This is concluded from the bulk RNAseq measurements where we observe *XIST* expression only in female samples. As *XIST* is primarily involved in X-inactivation [18], this is further supported by the observation of very little or no sex differences in expression of other X-linked genes. We find Y-linked genes expressed in male samples in all three tissues. Many of these Y-linked genes have functional X-linked homologs. When examining the combined expression levels of these homologous gene pairs, we did not detect notable differences between the sexes (Fig S1).

Interestingly, we identify 9 genes, including 6 protein-coding genes (*CYP4F22*, *HRCT1*, *ZCCHC16*, *DTX1*, *ATF5*, *CD24)*, 2 non-coding RNAs (*CTD-2325A15.3*, *RP11-495P10.9*) and

105 one pseudogene (*ASB9P1*) that show sex-differences in expression in HCC, but not in HCC
106 adjacent or healthy liver. Notably, Notch-regulating *DTX1*, activating transcription factor 5
107 (*ATF5*) and signal transducer *CD24* were downregulated in male HCC.

108 To further investigate the regulatory mechanisms underlying HCC of males and females, we
109 detected differentially expressed genes (DEGs) between tumor and tumor adjacent normal
110 samples in males and females, as well as in a joint analysis of both sexes (Fig 2A). Substantially
111 more DEGs were detected in sex-specific analyses than in the unstratified analysis (Fig 2B).
112 Specifically, DEGs that showed different magnitudes in fold change between the sexes were
113 detected in the sex-specific analyses, while DEGs with similar fold changes across all
114 comparisons were detected in the unstratified analysis as well as the sex-specific analyses (Fig
115 S2). DEGs that were only detected in the unstratified analysis, and not in sex-specific analyses,
116 showed a large variance in expression and were thus not detected as statistically significant DEGs
117 in sex-specific analyses (Fig S2). Tumor-infiltrating immune cells may produce spurious signals
118 in DEG analyses, which is evident from the detection of various immunoglobulin genes in tumor
119 vs. tumor adjacent comparisons (Table S4-6). However, cell content analyses based on gene
120 expression data did not exhibit sex differences in the prevalence of tumor-infiltrating immune
121 cells (Fig S3), and thus such spurious signals are unlikely to affect male-female comparisons.

122 In the joint analysis of male and female samples, we detected 581 DEGs, 23 of which were X-
123 linked (Table S4). In male and female specific analyses, we detected 606 and 416 DEGs,
124 respectively (Supplementary Tables 5 and 6). In both sex-specific analyses, 25 X-linked genes
125 were detected. Some of these X-linked genes had similar expression patterns in male and female
126 comparisons: including the MAGE (Melanoma-associated antigen) gene family members
127 *MAGEA1*, *MAGEA12*, *MAGEA6*, *MAGEC2*, *MAGEA3*, XAGE (X Antigen) family members
128 *XAGEA1* and *XAGEB1*, PAGE (Prostate-associated) family member *PAGE2*, as well as SXX
129 (synovial sarcoma X) family member *SSX1* were overexpressed in both male and female HCC.

130 Interestingly, prostate-associated antigen family member *PAGE4* was only overexpressed in male
131 HCC.

132 To further investigate the sex-specific functions of HCC tumors, we analyzed the lists of male-
133 and female-specific DEGs for the enrichment of functional pathways and canonical networks (Fig
134 2C and 2D). We found that sex-specific DEGs were enriched in pathways relevant to oncogenesis
135 and cancer progression (Supplementary Tables 7 and 8). Most of the top pathways were non-
136 overlapping, strongly indicating that male and female HCC are driven by different mechanisms
137 and processes.

### 138 Differential *cis*-eQTL effects in male and female HCC

139 We used eQTL analyses to detect germline genetic effects on tumor gene expression in males,
140 females, and in both sexes (Fig 3A). Due to the small number of samples from females, we were
141 unable to reliably detect female-specific eQTL associations. However, we detected 24 male-
142 specific eGenes (Fig 3B), which were not detected in the female-specific or in the unstratified
143 analysis. Since these associations were not detected in the unstratified analysis, they are likely not
144 a result of differential power to detect associations due to different sample sizes or allele
145 frequencies between the sexes, but exhibit differential effects between the sexes. None of the
146 male-specific eGenes were differentially expressed between male and female HCC, indicating
147 that the male-specific associations are not a driven of differences in overall gene expression levels
148 between males and females, but are likely to arise from factors such as chromatin accessibility,
149 transcription factor activity, and hormone receptors [19,20]. Genomic annotations show that most
150 of the detected regulatory variants were located on the non-coding regions (Fig 3C).

## 151 Overlap of eQTLs and HCC risk GWAS loci

152 eQTL mapping can be used to find a regulatory mechanism explaining GWAS risk loci by
153 identifying associations between genotypes and intermediate molecular phenotypes (gene
154 expression levels). A number of variants in the HLA region in chr6, as well as variants near
155 Signal Transducer and Activator of Transcription 4 (*STAT4*) in chr4 and Aspartylglucosaminidase
156 (*AGA*) in chr2, have been identified as HCC risk loci [21–25], but the regulatory mechanisms
157 connecting these risk variants to the HCC phenotype are not clear. We discovered sex-shared
158 regulatory variants tightly linked with known HCC risk loci altering gene expression levels in
159 HCC tumors (Table 1).

160 **Table 1. Tightly linked ($r^2$>0.8) HCC risk loci and sex-shared regulatory variants**.

| GWAS variant and study | Position | Adjacent gene(s) | Context | eVariant Position | eVariant distance to GWAS variant | eVariant Annotation | Target eGene |
|---|---|---|---|---|---|---|---|
| rs7574865[22] | chr2:191099907 | *STAT4* | Intron | 190603780 | 496127 | Intergenic, inter-CpG-island | *ORMDL1* |
| rs117413665[24] | chr4:177496993 | *LOC285500* | Intron | 177438565 | 58428 | CpG-shelf | *AGA* |
| rs2523961[25] | chr6:29971803 | *MICD* | Non-coding transcript exon | 29985849, 29659172*, 29626949* | 14046, 312631, 344854 | Intergenic, inter-CpG-island; Intron, intron-exon-boundary, inter-CpG-island; 1 to 5kb, intron, CpG-shore | *MCCD1P2*, *MICD*, *MICD* |
| rs1110446[25] | chr6:30103160 | *TRIM31* | 3'UTR | 30035457, 30010508 | 67703, 92652 | Promotor, exon, intron; Intron, CpG-shelf, lncRNA | *ZNRD1-AS1_2*, *ZNRD1-AS1_3* |
| rs3094137[25] | chr6:30233096 | *TRIM26* - *HCG17* | Intergenic | 29828972* | 404124 | Intron, CpG-shore | *HCG17* |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| rs2596542[21] | chr6:31398818 | *LOC101929072 - MICA* | Intergenic | 31411200, 31411843 | 12382, 13025 | Intron, CpG-shelf, lncRNA; Intron, CpG-shelf, lncRNA | *XXbac-BPG181B23.7*, *MICA* |
| rs9272105[23] | chr6:32632222 | *HLA-DRB1 - LOC107986589* | Intron | 32429758*, 32473499* | 202464, 158723 | Intergenic, inter-CpG-island; Intergenic, inter-CpG-island | *HLA-DRB1* |
| rs9275319[22] | chr6:32698518 | *HLA-DQB1 - MTCO3P1* | Intergenic | 32604372*, 32662070* | 94146, 36448 | Intergenic, inter-CpG-island; 1 to 5kb, CDS, exon, intron, intron-exon boundary, CpG-shelf | *HLA-DQB1* |
| rs2856723[99] | chr6:32699985 | *HLA-DQB1 - MTCO3P1* | Intergenic | 32604372*, 32662070* | 95613, 37915 | Intergenic, inter-CpG-island; 1 to 5kb, CDS, exon, intron, intron-exon boundary, CpG-shelf | *HLA-DQB1* |
| rs9275572[21] | chr6:32711222 | *MTCO3P1 - LOC102725019* | Intergenic | 32637497, 32662070, 32637430, 32662025 | 73725, 49152, 73792, 49197 | CDS, exon, first exon, intron-exon boundary, inter-CpG-island; 1 to 5kb, CDS, exon, intron, intron-exon boundary, CpG-shelf;Promoter, 5'UTR, exon, first exon, intron-exon boundary, inter-CpG-island; 1 to 5kb, exon, intron, intron-exon boundary, 3'UTR, CpG-shelf | *HLA-DQA1*, *HLA-DQB*, *HLA-DQB1-AS1*, *HLA-DQA2*, *HLA-DQB2* |

GWAS study citation is denoted in the risk variant column. Adjacent genes: If the GWAS variant is located within a gene, that gene is listed. If the variant is intergenic, the upstream and downstream genes are listed. Asterisk denotes instances where eVariants outside the LD block associated with reported or adjacent genes were found. eVariant sites have been annotated for genes and putatively regulatory regions (Methods, Table S12). Variant positions and distance are denoted as base pairs.

**Germline genetic effects on HCC tumor gene expression and survival**

168 We used Cox's proportional hazard model to test for effects of eGene expression on patient
169 survival in the TCGA Liver Hepatocellular Carcinoma (LIHC) cohort. Out of the 1204 sex-shared
170 eGenes, 50 were associated with joint survival (FDR-adjusted *p*-value <0.05, Table 2). No
171 female-specific survival associations were detected, likely due to the low number of female
172 samples affecting statistical power to detect survival effects. However, we detected 65 sex-shared
173 eGenes associated with male survival (Table 3). 15 of these were not associated with joint
174 survival, clearly indicating differential effects on overall survival between the sexes.

175 **Table 2. Sex-shared eGenes associated with overall survival in joint analysis of both sexes.**

| Gene name | Likelihood ratio $\chi^2$ | FDR-adjusted *p*-value |
|---|---|---|
| *PSRC1* | 27.30 | 2.09E-04 |
| *MMP1* | 22.24 | 1.45E-03 |
| *CFHR3* | 19.15 | 3.63E-03 |
| *PON1* | 19.45 | 3.63E-03 |
| *RP11-274H2.3* | 17.26 | 7.83E-03 |
| *FAM110D* | 16.49 | 9.83E-03 |
| *CYP3A5* | 15.94 | 1.12E-02 |
| *TAF1B* | 15.43 | 1.15E-02 |
| *BAATP1* | 15.49 | 1.15E-02 |
| *RP4-608O15.3* | 14.64 | 1.46E-02 |
| *AC018641.7* | 14.43 | 1.46E-02 |
| *RP11-459O16.8* | 14.54 | 1.46E-02 |
| *CCDC163P* | 14.27 | 1.47E-02 |
| *RP11-108O10.2* | 14.00 | 1.57E-02 |
| *TMPRSS11D* | 12.78 | 2.08E-02 |
| *RP11-650L12.2* | 12.88 | 2.08E-02 |
| *TMEM147-AS1* | 12.76 | 2.08E-02 |
| *LPA* | 12.90 | 2.08E-02 |

| | | |
|---|---|---|
| *RRM1-AS1* | 13.12 | 2.08E-02 |
| *RP11-105C19.2* | 12.72 | 2.08E-02 |
| *RP11-669E14.4* | 13.13 | 2.08E-02 |
| *CYP4F60P* | 12.02 | 2.88E-02 |
| *RP11-711K1.8* | 11.79 | 3.10E-02 |
| *RGN* | 11.72 | 3.10E-02 |
| *RP11-738E22.3* | 11.41 | 3.15E-02 |
| *LINC01252* | 11.43 | 3.15E-02 |
| *TMEM147* | 11.52 | 3.15E-02 |
| *CCDC146* | 11.51 | 3.15E-02 |
| *RP1-239B22.5* | 11.29 | 3.20E-02 |
| *ANKRD20A7P* | 11.25 | 3.20E-02 |
| *GLMP* | 11.16 | 3.25E-02 |
| *MASTL* | 10.92 | 3.58E-02 |
| *CFHR1* | 10.75 | 3.80E-02 |
| *CTD-2270P14.1* | 10.66 | 3.88E-02 |
| *TRG-AS1* | 10.53 | 3.93E-02 |
| *CHRNA5* | 10.58 | 3.93E-02 |
| *FCRL3* | 10.33 | 4.00E-02 |
| *DDT* | 10.30 | 4.00E-02 |
| *RP4-736L20.3* | 10.32 | 4.00E-02 |
| *IKBIP* | 10.39 | 4.00E-02 |
| *TMEM79* | 10.05 | 4.38E-02 |
| *GABPB1-AS1* | 10.07 | 4.38E-02 |
| *LDAH* | 9.63 | 4.77E-02 |
| *FAM99B* | 9.71 | 4.77E-02 |
| *RP11-109L13.1* | 9.72 | 4.77E-02 |
| *SPSB2* | 9.59 | 4.77E-02 |

| | | |
|---|---|---|
| *DDX11* | 9.62 | 4.77E-02 |
| *DDTL* | 9.69 | 4.77E-02 |
| *CR848007.2* | 9.65 | 4.77E-02 |
| *TECTB* | 9.57 | 4.77E-02 |

176 Cox's proportional hazard ratio: likelihood ratio $\chi^2$, FDR-adjusted *p*-value threshold 0.05.

177 **Table 3. Sex-shared eGenes associated with overall survival in males.**

| **Gene name** | **Likelihood ratio $\chi^2$** | **FDR-adjusted *p*-value** |
|---|---|---|
| *PSRC1* | 30.89944 | 3.27E-05 |
| *MMP1* | 21.751568 | 1.25E-03 |
| *RGN* | 21.758012 | 1.25E-03 |
| *RP11-650L12.2* | 20.488249 | 1.81E-03 |
| *PON1* | 18.294246 | 2.74E-03 |
| *FAM99A** | 18.308255 | 2.74E-03 |
| *FAM99B* | 17.940412 | 2.74E-03 |
| *RP1-239B22.5* | 18.05079 | 2.74E-03 |
| *CHRNA5* | 19.089358 | 2.74E-03 |
| *TMEM147* | 18.899008 | 2.74E-03 |
| *RP11-108O10.2* | 17.018204 | 4.05E-03 |
| *LPA* | 16.509603 | 4.48E-03 |
| *RRM1-AS1* | 16.518558 | 4.48E-03 |
| *CYP3A5* | 16.170768 | 4.98E-03 |
| *AADACP1** | 15.893269 | 5.04E-03 |
| *TMEM147-AS1* | 15.899306 | 5.04E-03 |
| *RP11-274H2.3* | 15.100292 | 7.22E-03 |
| *CCDC163P* | 14.133518 | 1.03E-02 |
| *CENPQ** | 14.243691 | 1.03E-02 |
| *RP4-736L20.3* | 14.142785 | 1.03E-02 |

| | | |
|---|---|---|
| *F12** | 13.823228 | 1.05E-02 |
| *GRK6** | 13.92507 | 1.05E-02 |
| *LRRC23** | 13.902588 | 1.05E-02 |
| *RP11-105C19.2* | 13.577894 | 1.15E-02 |
| *CFHR3* | 13.16418 | 1.37E-02 |
| *RP4-608O15.3* | 13.044288 | 1.41E-02 |
| *CMA1** | 11.929339 | 2.29E-02 |
| *RP13-753N3.3** | 12.001046 | 2.29E-02 |
| *RP11-669E14.4* | 11.937499 | 2.29E-02 |
| *WHAMM** | 11.580894 | 2.59E-02 |
| *RRP12** | 11.538845 | 2.59E-02 |
| *CPXM2** | 11.518611 | 2.59E-02 |
| *RP11-80A15.1** | 11.223818 | 2.70E-02 |
| *BAATP1* | 11.314561 | 2.70E-02 |
| *FRMPD2B** | 11.377566 | 2.70E-02 |
| *RP11-153K11.3** | 11.258503 | 2.70E-02 |
| *AC114730.7** | 11.069887 | 2.71E-02 |
| *RP11-459O16.8* | 11.068885 | 2.71E-02 |
| *ULK4P2** | 11.108146 | 2.71E-02 |
| *TRG-AS1* | 10.96346 | 2.80E-02 |
| *RP11-421M1.8** | 10.769889 | 2.97E-02 |
| *RP11-434C1.2** | 10.739028 | 2.97E-02 |
| *FAM110D* | 10.717104 | 2.97E-02 |
| *TMPRSS11D* | 10.580786 | 3.13E-02 |
| *DDX11* | 10.472235 | 3.24E-02 |
| *CTD-2270P14.1* | 10.390323 | 3.32E-02 |
| *AGMO** | 10.305048 | 3.40E-02 |
| *SPSB2* | 9.904202 | 4.05E-02 |

| | | |
|---|---|---|
| *LARS** | 9.936718 | 4.05E-02 |
| *ZG16B** | 9.800817 | 4.10E-02 |
| *DCTN5** | 9.801029 | 4.10E-02 |
| *CCDC146* | 9.774758 | 4.10E-02 |
| *FCRL3* | 9.62148 | 4.37E-02 |
| *CFHR1* | 9.357763 | 4.43E-02 |
| *LDAH* | 9.444584 | 4.43E-02 |
| *RPL23AP7** | 9.502224 | 4.43E-02 |
| *LINC01252* | 9.43441 | 4.43E-02 |
| *RP11-1260E13.2** | 9.441852 | 4.43E-02 |
| *NDUFA6-AS1** | 9.337829 | 4.43E-02 |
| *CYP2D6** | 9.363136 | 4.43E-02 |
| *LINC01277** | 9.411621 | 4.43E-02 |
| *SLC25A1P5** | 9.211757 | 4.67E-02 |
| *RP11-738E22.3* | 9.142993 | 4.77E-02 |
| *RP11-711K1.8* | 9.071118 | 4.88E-02 |
| *CYP2D7* | 9.043546 | 4.88E-02 |

Cox's proportional hazard ratio: likelihood ratio $\chi^2$, FDR-adjusted *p*-value threshold 0.05. Genes that were not associated with joint survival are indicated with an asterisk.

# Discussion

## Sex-specific patterns of gene expression in liver and HCC

It is well established that patterns of gene expression vary between the sexes. Previous studies have confounded these differences with those which may be etiologically important in cancer. For example, Yuan et al. previously reported extensive sex-biased signatures in gene expression in HCC and other strongly sex-biased cancers [12]. From the results presented here, it is possible to

186 distinguish the differences detected in comparisons of male and female HCC from those
187 reflecting normal sex-differences.

188 We report etiologically meaningful differences in gene expression between male and female HCC
189 cases. In addition to X- and Y-linked genes, 8 autosomal genes were expressed in a sex-biased
190 way in HCC tumors. Three of these genes - all of which were underexpressed in male HCC
191 compared to female - are of particular interest in the context of HCC: Notch-regulating *DTX1* has
192 been identified as a putative tumor suppressor gene in head and neck squamous cell carcinoma
193 [26]. *ATF5* is highly expressed in the liver but downregulated in HCC [27]. *ATF5* is known to
194 inhibit hepatocyte proliferation [28] and re-expression of *ATF5* in HCC inhibits proliferation and
195 induces G2/M arrest of the cell cycle [28]. *ATF5* also may act as a negative regulator of IL-1β
196 signaling pathway in hepatocytes and thus act as a regulator of IL-1β-mediated immune response
197 [29]. *CD24* has a crucial role in T cell homeostasis and autoimmunity [30]. In breast cancer, high
198 *CD44*/*CD24* ratio is an indicator for malignancy [31]. The opposing roles of *CD24* expression in
199 cancer and autoimmune diseases raise interesting questions on the role of sex differences in
200 immunity underlying female prevalence in autoimmunity [32] and male prevalence in cancer.
201 More studies are needed to better understand the differential regulation of immune functions
202 between the sexes, and how these differences contribute to the observed biases in disease
203 occurrence.

204 In both sex-specific analyses, 25 X-linked genes were detected including the MAGE family
205 (Melanoma-associated antigen), the XAGE (X Antigen) family members, the PAGE (Prostate-
206 associated) family, and the SXX (synovial sarcoma X) family. While most of the 25 X-linked
207 genes exhibited similar patterns between the sexes, *PAGE4* was overexpressed only in male
208 HCC. *PAGE4* is expressed in the human fetal prostate during development as well as malignant
209 prostate and prostate cancer, but not in healthy adult prostate [33]. The role of *PAGE4* in prostate
210 cancer has been studied in considerable detail, and it has been indicated as a promising target for

211 therapies [33]. However, its role in other cancers has not been thoroughly explored. MAGE,
212 PAGE, and SSX genes are members of a large family of cancer testis antigens (CTAs). CTAs are
213 expressed in tumors, but not in normal tissue, with the exception of testis and placenta, and
214 recognized by cytotoxic T lymphocytes, making them ideal targets for immunotherapeutic
215 approaches [34]. *MAGE* expression has been associated with response to checkpoint inhibition
216 therapy in metastatic melanoma [35]. Given the observed differences here, CTA expression may
217 prove valuable in determining the potential response for checkpoint inhibition therapy in HCC
218 [36].

219 Sex-differences in the pathway and network activation strongly indicate that male and female
220 HCC are driven by distinct functional pathways. Males and females differed in oncogenic
221 processes with females showing RAS pathway enrichment while males showed PI3K/AKT
222 pathway enrichment. Additionally, female DEGs were highly enriched in genes involved in
223 innate immunity Toll-like receptor (TLR) signaling (Table S9), while male DEGs exhibited high
224 enrichment of genes involved in adaptive immunity and Notch signaling (Table S8). TLR
225 signaling has a crucial role in the regulation of the immune system by evoking an inflammatory
226 response [37]. Interestingly, TLR signaling may activate Notch signaling [38], and Notch
227 signaling may feedback modulate TLR signaling pathway to modulate inflammatory response
228 through extracellular signal-regulated kinase 1/2-mediated nuclear factor κB activation [39]. Both
229 of these processes regulate macrophage functions and are likely to have a major role in cancer
230 progression by modulating the inflammatory tumor microenvironment [37–39]. TLR and Notch
231 signaling pathways are notable targets for anti-cancer and anti-metastasis therapies [40,41].
232 Furthermore, while both sexes exhibited an enrichment of genes involved in cell cycle, apoptosis
233 and p53 signaling, only males showed an activation of pathways regulating natural killer cell
234 cytotoxicity, and females showed an activation of pathways involved in NF-kB (nuclear factor
235 kappa-light-chain-enhancer of activated B cells) activation (Fig 2).

236 **Sex-specific regulatory effects of germline variants**

237 We detected 24 genes under germline regulatory control in male HCC only (Fig 3B). Functional
238 annotations of these male HCC specific eGenes provide insight into possible regulatory
239 mechanisms contributing to the observed male-bias in HCC. Protein O-glucosyltransferase 1
240 (*POGLUT1*) was found to be under germline regulation in male HCC, but not in female HCC or
241 in joint analysis of both sexes (Fig 3D). The eVariant associated with *POGLUT1* is located on a
242 promoter region of its target (Table S14). *POGLUT1* is an enzyme that is responsible for O-
243 linked glycosylation of proteins. Glycosylation is a major type of post-translational modification,
244 during which proteins are decorated with oxygen (O) or nitrogen (N) atoms of amino acids. These
245 modifications may critically alter the proteins' folding, localization, and function [42]. Altered
246 glycosylation of proteins has been observed in many cancers [43,44], including liver cancer
247 [45,46]. Cell surface receptor-mediated growth and apoptosis have been proposed as mechanisms
248 explaining the role of O-glycosylation in cancer [47]. Interestingly, altered glycosylation of tumor
249 proteins may also allow the tumor cells to evade natural killer cell immunity [48].

250 *POGLUT1* is an essential regulator of Notch signaling and is likely involved in cell fate and
251 tissue formation during development. Genes involved in Notch signaling were also found to be
252 expressed in a sex-biased way in HCC tumors and highly enriched in DEGs detected in tumor vs.
253 tumor adjacent normal comparison of male samples, pointing to a possible role of *Notch* in the
254 development of HCC in males. The role of Notch signaling in various biological processes has
255 been thoroughly studied in a variety of organisms and tissues, particularly in *Drosophila* [49].
256 Notch signaling is of particular interest in the context of HCC, as it is involved in liver
257 development [50,51], the development of sexually dimorphic traits [52], and tumorigenesis [49].
258 The role of Notch signaling in HCC may differ between early and late-stage tumors and among
259 molecular subtypes, and further studies are necessary to understand the possible oncogenic
260 properties of Notch among HCC subtypes and between the sexes.

261 Chromatin states are a major determinant of sex-biased gene expression and sex-specific
262 regulatory functions in mouse liver [53] and in human peripheral blood mononuclear cells [19].
263 Other plausible mechanisms underlying sex-specific regulatory effects are differential
264 transcription factor activity and hormonal regulation. Further studies are needed to explore the
265 role of these mechanisms in male and female HCC.

266 **Regulatory mechanisms underlying HCC risk variants**

267 Genetic variants in the Human Leukocyte Antigen (HLA) region have been found to be
268 associated with hepatitis B related HCC in Asian populations [22,25]. While some of these risk
269 variants were found to be associated with HLA expression in normal liver tissue [25] or
270 differentially expressed between HCC tumors and tumor adjacent tissue [22], direct evidence of
271 regulatory mechanisms connecting the risk variants to HCC was lacking. We discovered
272 regulatory variants tightly linked with known HCC risk loci associated with HLA expression in
273 HCC tumors (Table 1). Polymorphisms in the HLA region may influence immune responses and
274 oncogenesis by altering the epitope binding properties and/or expression levels of HLA
275 molecules. Loss of HLA expression, associated with an immune escape by avoiding T-cell
276 recognition, is commonly found in malignant cells [54]. Tumor immune escape is one of the
277 hallmarks of cancer, and it has been shown to have negative effects on the clinical outcome of
278 cancer immunotherapies [55]. HLA expression is critical in tumor rejection, and HLA re-
279 expression has been recognized as a major potential target in the development of
280 immunotherapies [54]. HLA allele frequencies and LD relationships among these variants vary
281 between population, and HLA polymorphisms are likely partially responsible for the ethnic
282 disparities in HBV persistence among ethnicities [56], and thus likely contribute to the observed
283 biases in HCC occurrence between populations. Interestingly, sex modulates the degree to which
284 HLA molecules propagate the selection and expansion of T cells [57]. Sex-specific regulatory
285 functions discovered in this study may partially underlie the sex-biases in diseases with immune

286 system involvement. HLA associated shaping of the immune system points to a promising target
287 for future studies to develop more targeted therapies accounting for patients' germline genetic
288 makeup.

289 **Genes under germline regulatory control alter patient survival in HCC**

290 In addition to the discovery of HCC risk variants, characterizations of patterns and levels of gene
291 expression in HCC tumors have led to the discovery of biomarkers associated with HCC
292 etiologies, disease progression and outcome [58–60]. We discovered genes under germline
293 regulatory control altering overall survival in the TCGA LIHC cohort, 15 of which showed
294 differential survival effects between the sexes (Tables 2 and 3). Some of the 15 male-specific
295 survival associated eGenes have been found to affect disease progression and outcome in other
296 cancer types, indicating a pan-cancer role of these genes. In breast cancer, lower Coagulation
297 Factor XII (*F12*) expression is a favorable prognostic marker [61]. Interestingly, low *F12*
298 expression was associated with poor survival among males in the TCGA LIHC cohort (Fig 4). In
299 some population studies, variants in Cytochrome P450 2D6 (*CYP2D6*) have been found to be
300 associated with recurrence in breast cancer patients treated with Tamoxifen ([62] but see [63] and
301 [64]), and it is highly expressed in the liver, where its expression may affect drug clearance,
302 efficacy, and safety [65]. G protein-coupled receptor kinase 6 (*GRK6*) deficiency promotes
303 angiogenesis, tumor progression, and metastasis in murine models of human lung cancer [66].
304 Variations in the miRNA binding sites of the Zymogen Granule Protein 16B (*ZG16B*) gene are
305 associated with prognosis in patients with colorectal cancer [67]. Variants at the Dynactin Subunit
306 5 (*DCTN5*) gene are associated with ovarian cancer mortality [68].

307 Furthermore, several of the male-specific survival associated eGenes have been shown to be
308 involved in HCC and other liver disorders. Family With Sequence Similarity 99 Member A
309 (*FAM99A*) has previously been shown to be downregulated in HCC [69]. In our study, low

310 *FAM99A* expression indicated poor survival in males (Table 3). Carboxypeptidase X, M14
311 Family Member 2 (*CPXM2*) has been identified as a driver gene in HCC based on patterns of
312 somatic alterations [70]. Alkylglycerol Monooxygenase (*AGMO*) has been shown to be
313 hypermethylated and transcriptionally repressed in HCC in comparison to cancer-free liver tissue
314 [71]. Mutations in the Leucyl-TRNA Synthetase (*LARS*) gene have been identified as a cause of
315 infantile hepatopathy [72].

316 In summary, we discovered differential genetic and regulatory functions in HCC tumors between
317 the sexes. By integrating genotype and gene expression data, we identified putative regulatory
318 mechanisms underlying known HCC risk loci and discovered genes under germline regulatory
319 control modulating the overall survival of patients with HCC. Our results highlight the role of
320 differential immune functions in cancer sex-bias and germline regulatory variants associated with
321 these functions. This work provides a framework for future studies on sex-biased cancers.

## 322 Materials and Methods

### 323 Data

324 GTEx (release V6p) whole transcriptome (RNAseq) data (dbGaP accession #8834) were
325 downloaded from dbGaP. TCGA LIHC Affymetrix Human Omni 6 array genotype data, whole
326 exome sequence (WES) and RNAseq data (dbGaP accession #11368) were downloaded from
327 NCI Genomic Data Commons [73]. FASTQ read files were extracted from the TCGA LIHC
328 WES BAM files using *XYAlign* [74]. We used *FastQC* [75] to assess the WES and RNAseq
329 FASTQ quality. Reads were trimmed using *TRIMMOMATIC IlluminaClip* [76], with the
330 following parameters: seed mismatches 2, palindrome clip threshold 30, simple clip threshold 10,
331 leading quality value 3, trailing quality value 3, sliding window size 4, minimum window quality
332 30 and minimum read length of 50.

333 **Read mapping and read count quantification**

334 Reads were mapped to custom sex-specific reference genomes using *HISAT2* [77]. To avoid miss
335 mapping of reads along the sex chromosomes, female samples were mapped to the human
336 reference genome GRCh38 with the Y-chromosome hard-masked. Male samples were mapped to
337 the human reference genome with Y-chromosomal pseudoautosomal regions hard-masked. Gene-
338 level counts from RNAseq were quantified using *Subread featureCounts* [78]. Reads overlapping
339 multiple features (genes or RNA families with conserved secondary structures) were counted for
340 each feature.

341 **Germline variant calling**

342 BAM files were processed according to Broad Institute *GATK* (*Genome Analysis Toolkit*) best
343 practices [79–81]: Read groups were added with *Picard Toolkit*'s *AddOrReplaceReadGroups* and
344 optical duplicates marked with *Picard Toolkit*'s *MarkDuplicates* (v.2.18.1,
345 http://broadinstitute.github.io/picard/). Base quality scores were recalibrated with *GATK*
346 (v.4.0.3.0) *BaseRecalibrator.* Germline genotypes were called from whole blood Whole Exome
347 Sequence samples from 248 male and 119 female HCC cases using the scatter-gather method
348 with *GATK HaplotypeCaller* and *GenotypeGVCFs* [79]. Affymetrix 6.0 array genotypes of
349 matching samples were lifted to GRCh38 using the UCSC *LiftOver* tool [82] and converted to
350 VCF. Filters were applied to retain variants with a minimum quality score >30, minor allele
351 frequency >10%, minor allele count >10, and no call rate <10% across all samples.

352 **Cellular content of tumor samples**

353 To examine the cellular heterogeneity of tumor samples, we performed cell type enrichment
354 analysis for 64 immune and stromal cell types using *xCell* [83].

355 **Filtering of gene expression data**

356 FPKM (Fragments Per Kilobase of transcript per Million mapped reads) expression values for
357 each gene were obtained using *EdgeR* [84]. Each expression dataset was filtered to retain genes
358 with mean FPKM>0.5 and read count of >6 in at least 10 samples across all samples under
359 investigation to avoid falsely inferring differential expression between two groups where both are
360 functionally not transcribed.

361 **Differential expression analysis**

362 For differential expression (DE) analysis, filtered, untransformed read count data were quantile
363 normalized and logCPM transformed with *voom* [85]. From the TCGA LIHC dataset, paired
364 tumor and tumor adjacent samples were available for 22 females and 28 males. From the GTEx
365 liver dataset, 22 female and 28 male samples were randomly selected to be used in the DE
366 analysis. A multi-factor design with sex and tissue type as predictor variables were used to fit the
367 linear model. To make comparisons both between and within subjects in the paired tumor and
368 tumor adjacent samples, *duplicateCorrelation* was used to treat the individual as a random effect.

369 Differentially expressed genes (DEGs) between comparisons were identified using the
370 *limma*/*voom* pipeline [85]. Linear modeling was conducted using the *LmFit* and *contrast.fit*
371 functions. DEGs between sexes and tissue types were identified using contrast designs (e.g., all
372 males for tumor versus all females for tumor, or all tumor samples versus all tumor adjacent
373 samples within the same sex). In all comparisons, DEGs were identified by computing empirical
374 Bayes statistics with *eBayes*, with an FDR adjusted *p*-value threshold of 0.01 and an absolute $\log_2$
375 fold-change ($\log_2$FC) threshold of 2.

376 **Enrichment of biological functions and canonical pathways**

377 We utilized the Gene Ontology Consortium's web tool [86] to analyze gene lists for enrichment
378 in regards to molecular function, biological processes, and cellular functions. Fisher's exact test
379 was used to calculate enrichment scores. We also used the *NetworkAnalyst [87]* for enrichment
380 and network-based analyses of gene lists in regards to pathway relationships (KEGG and
381 Reactome pathways). An FDR-adjusted *p*-value threshold of 0.01 was used to select significantly
382 enriched GO terms and canonical pathways.

383 **Accounting for technical confounders and population structure**

384 Gene expression values are affected by genetic, environmental, and technical factors, many of
385 which may be unknown or unmeasured. Technical confounding factors introduce sources of
386 variance that may greatly reduce the statistical power of association studies, and even cause false
387 signals [88]. Thus, it is necessary to account for known and unknown technical confounders. This
388 is often achieved by detecting a set of latent confounding factors with methods such as principal
389 component analysis (PCA), Probabilistic Estimation of Expression Residuals (PEER) [89] or
390 Surrogate Variable Analysis (SVA) [88]. These known and surrogate variables are then used as
391 covariates in downstream analyses, or expression data is adjusted using a regression model or
392 other approach e.g. *ComBat*, which utilizes an Empirical Bayes method [90]. In eQTL studies, the
393 number of latent factors to adjust for is often selected to maximize the number of significant
394 associations. However, such an approach may remove biologically relevant sources of variation
395 and increase the number of false positive associations. Introducing too many covariates to linear
396 models may also cause a problem of overfitting. Furthermore, confounding factor methods may
397 model the effects of trans-acting broad impact eQTL as confounding variation [91,92].

398 Our goal was to remove the technical sources of variation while protecting the biologically
399 relevant variation. Because of missing values in technical covariate data, we were unable to adjust
400 the expression data directly for known covariates. We identified 50 PEER factors in the tumor

401 data. Some of these latent factors were associated with known clinical phenotypes or technical
402 covariates (Fig S4). Weights of the PEER factors that did not strongly correlate with biologically
403 relevant phenotypes were used as covariates in the eQTL analysis. To select these PEER factors,
404 FDR adjusted *p*-value threshold of 0.01 based on linear regression for continuous covariates and
405 ANOVA for categorical covariates were used. To avoid overfitting, the number of PEER factors
406 to include was limited to 10.

407 We used the R package *SNPRelate* [93] to perform PCA on the germline genotype data. We
408 accounted for population structure by applying the first three genotype PCs as covariates in the
409 eQTL analysis.

410 **eQTL analysis**

411 Germline genotypes and tumor gene expression data from 248 male and 119 female donors were
412 available for use in the eQTL analysis. Filtered count data was normalized by fitting the FPKM
413 values of each gene and sample to the quantiles of the normal distribution. *cis*-acting eQTLs were
414 detected with *QTLtools* v.1.1 [94]. We used the permutation pass with 10,000 permutations to get
415 adjusted *p*-values of associations between the phenotypes and the top WES and array variants in
416 *cis*. FDR adjusted *p*-values were calculated to correct for multiple phenotypes tested. An FDR-
417 adjusted *p*-value threshold of 0.01 was used to select significant associations.

418 **Annotating genomic locations of putative regulatory variants**

419 We used the R package *Annotatr* to annotate the genomic locations of eVariants [95]. Variant
420 sites were annotated for promoters, 5'UTRs, exons, introns, 3'UTRs, CpGs (CpG islands, CpG
421 shores, CpG shelves), and putative regulatory regions based on ChromHMM [96] annotations.

422 **Overlap of eQTL loci and HCC GWAS loci**

423 Due to the lack of genome-wide GWAS summary statistics, we were unable to test for
424 colocalization of GWAS and eVariants. To overcome this, we used top HCC GWAS variants and
425 top eVariant to detect high confidence colocalization events based on linkage. Locations of 32
426 GWAS loci associated with HCC were obtained from the NHGRI-EBI GWAS catalog [97]. To
427 find instances where known GWAS loci and eQTL loci are overlapping or tightly linked, we
428 searched for loci with $r^2>0.8$ of the GWAS risk locus. Linked loci were obtained with the
429 *LDLink*'s *LDProxy* tool [98].

430 **Survival analyses**

431 We used the R package *survival* to test for differences in survival between different eVariant
432 genotypes. To detect possible differences in early and late survival, we used the *surv_pvalue*
433 function with log-rank, Tarone-Ware and Fleming-Harrington (p=1, q=1) tests. For genotype
434 comparisons, the *pairwise_survdiff* function from the R package *survminer* was used to calculate
435 *p*-values between all pairwise comparisons. To test for the effect of eGene expression on patient
436 survival, we utilized Cox's proportional hazards regression model with the function *coxph.* FDR
437 adjusted *p*-values were calculated with the *p.adjust* function from the *stats* R package.

438 **Acknowledgments**

439 This study was supported by ASU Center for Evolution and Medicine postdoctoral fellowship for
440 HMN, ASU School of Life Sciences startup funds for MAW and ASU Center for Evolution and
441 Medicine Venture funds.

442 **Citations**

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

445 2. Regitz-Zagrosek Vera. Sex and gender differences in health. EMBO Rep. 2012;13: 596–
446 603.

447 3. Jaillon S, Berthenet K, Garlanda C. Sexual Dimorphism in Innate Immunity. Clin Rev
448 Allergy Immunol. 2017; doi:10.1007/s12016-017-8648-x

449 4. Klein SL. The effects of hormones on sex differences in infection: from genes to
450 behavior. Neurosci Biobehav Rev. 2000;24: 627–638.

451 5. Klein SL, Flanagan KL. Sex differences in immune responses. Nat Rev Immunol.
452 2016;16: 626–638.

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

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

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

457 9. Petrick JL, Braunlin M, Laversanne M, Valery PC, Bray F, McGlynn KA. International
458 trends in liver cancer incidence, overall and by histologic subtype, 1978-2007. Int J Cancer.
459 2016;139: 1534–1545.

460 10. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer
461 statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in
462 185 countries. CA Cancer J Clin. 2018; doi:10.3322/caac.21492

463 11. Tang R, Liu H, Yuan Y, Xie K, Xu P, Liu X, et al. Genetic factors associated with risk of
464 metabolic syndrome and hepatocellular carcinoma. Oncotarget. 2017;8: 35403–35411.

465 12. Yuan Y, Liu L, Chen H, Wang Y, Xu Y, Mao H, et al. Comprehensive Characterization
466 of Molecular Differences in Cancer between Male and Female Patients. Cancer Cell. 2016;29:
467 711–722.

468 13. Agarwal R, Cao Y, Hoffmeier K, Krezdorn N, Jost L, Meisel AR, et al. Precision
469 medicine for hepatocelluar carcinoma using molecular pattern diagnostics: results from a
470 preclinical pilot study. Cell Death Dis. 2017;8: e2867.

471 14. Yin L, Cai Z, Zhu B, Xu C. Identification of Key Pathways and Genes in the Dynamic
472 Progression of HCC Based on WGCNA. Genes . 2018;9. doi:10.3390/genes9020092

473 15. Wen J, Song C, Liu J, Chen J, Zhai X, Hu Z. Expression quantitative trait loci for
474 TNFRSF10 influence both HBV infection and hepatocellular carcinoma development. J Med
475 Virol. 2016;88: 474–480.

476 16. Ma S, Yang J, Song C, Ge Z, Zhou J, Zhang G, et al. Expression quantitative trait loci for
477 PAX8 contributes to the prognosis of hepatocellular carcinoma. PLoS One. 2017;12: e0173700.

478 17. Tian T, Song C, Pu Z-N, Ge Z-J, Yu C-X, Liu J-B, et al. Expression quantitative trait loci
479 for PVT1 contributes to the prognosis of hepatocellular carcinoma. HRMAGAZINE. 2018;4: 27.

480 18. Chow JC, Yen Z, Ziesche SM, Brown CJ. Silencing of the mammalian X chromosome.
481 Annu Rev Genomics Hum Genet. 2005;6: 69–92.

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

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

486 21. Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R, Hosono N, et al. Genome-wide
487 association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. Nat
488 Genet. 2011;43: 455–458.

489 22. Jiang D-K, Sun J, Cao G, Liu Y, Lin D, Gao Y-Z, et al. Genetic variants in STAT4 and
490 HLA-DQ genes confer risk of hepatitis B virus-related hepatocellular carcinoma. Nat Genet.
491 2013;45: 72–75.

492 23. Li S, Qian J, Yang Y, Zhao W, Dai J, Bei J-X, et al. GWAS identifies novel susceptibility
493 loci on 6p21.32 and 21q21.3 for hepatocellular carcinoma in chronic hepatitis B virus carriers.
494 PLoS Genet. 2012;8: e1002791.

495 24. Lin Y-Y, Yu M-W, Lin S-M, Lee S-D, Chen C-L, Chen D-S, et al. Genome-wide
496 association analysis identifies a GLUL haplotype for familial hepatitis B virus-related
497 hepatocellular carcinoma. Cancer. 2017;123: 3966–3976.

498 25. Sawai H, Nishida N, Khor S-S, Honda M, Sugiyama M, Baba N, et al. Genome-wide
499 association study identified new susceptible genetic variants in HLA class I region for hepatitis B
500 virus-related hepatocellular carcinoma. Sci Rep. 2018;8: 7958.

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

504 27. Pascual M, Gómez-Lechón MJ, Castell JV, Jover R. ATF5 is a highly abundant liver-
505 enriched transcription factor that cooperates with constitutive androstane receptor in the
506 transactivation of CYP2B6: implications in hepatic stress responses. Drug Metab Dispos.
507 2008;36: 1063–1072.

508 28. Gho JW-M, Ip W-K, Chan KY-Y, Law PT-Y, Lai PB-S, Wong N. Re-expression of
509 transcription factor ATF5 in hepatocellular carcinoma induces G2-M arrest. Cancer Res. 2008;68:
510 6743–6751.

511 29. Abe T, Kojima M, Akanuma S, Iwashita H, Yamazaki T, Okuyama R, et al. N-terminal
512 hydrophobic amino acids of activating transcription factor 5 (ATF5) protein confer interleukin 1β
513 (IL-1β)-induced stabilization. J Biol Chem. 2014;289: 3888–3900.

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

516 31. Li W, Ma H, Zhang J, Zhu L, Wang C, Yang Y. Unraveling the roles of CD44/CD24 and
517 ALDH1 as cancer stem cell markers in tumorigenesis and metastasis. Sci Rep. 2017;7: 13856.

518 32. Whitacre CC. Sex differences in autoimmune disease. Nat Immunol. 2001;2: 777–780.

519 33. Kulkarni P, Dunker AK, Weninger K, Orban J. Prostate-associated gene 4 (PAGE4), an
520 intrinsically disordered cancer/testis antigen, is a novel therapeutic target for prostate cancer.
521 Asian J Androl. 2016;18: 695–703.

522 34. Mazzone R, Zwergel C, Mai A, Valente S. Epi-drugs in combination with
523 immunotherapy: a new avenue to improve anticancer efficacy. Clin Epigenetics. 2017;9: 59.

524 35. Shukla SA, Bachireddy P, Schilling B, Galonska C, Zhan Q, Bango C, et al. Cancer-
525 Germline Antigen Expression Discriminates Clinical Outcome to CTLA-4 Blockade. Cell.
526 2018;173: 624–633.e8.

527 36. Iñarrairaegui M, Melero I, Sangro B. Immunotherapy of Hepatocellular Carcinoma: Facts
528 and Hopes. Clin Cancer Res. 2018;24: 1518–1524.

529 37. Gu J, Liu Y, Xie B, Ye P, Huang J, Lu Z. Roles of toll-like receptors: From inflammation
530 to lung cancer progression. Biomed Rep. 2018;8: 126–132.

531 38. Palaga T, Buranaruk C, Rengpipat S, Fauq AH, Golde TE, Kaufmann SHE, et al. Notch
532 signaling is activated by TLR stimulation and regulates macrophage functions. Eur J Immunol.
533 2008;38: 174–183.

534 39. Zhang Q, Wang C, Liu Z, Liu X, Han C, Cao X, et al. Notch signal suppresses Toll-like
535 receptor-triggered inflammatory responses in macrophages by inhibiting extracellular signal-
536 regulated kinase 1/2-mediated nuclear factor κB activation. J Biol Chem. 2012;287: 6208–6217.

537 40. Li L, Tang P, Li S, Qin X, Yang H, Wu C, et al. Notch signaling pathway networks in
538 cancer metastasis: a new target for cancer therapy. Med Oncol. 2017;34: 180.

539 41. Braunstein MJ, Kucharczyk J, Adams S. Targeting Toll-Like Receptors for Cancer
540 Therapy. Target Oncol. 2018;13: 583–598.

541 42. Shental-Bechor D, Levy Y. Effect of glycosylation on protein folding: a close look at
542 thermodynamic stabilization. Proc Natl Acad Sci U S A. 2008;105: 8256–8261.

543 43. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. Nat
544 Rev Cancer. 2015;15: 540–555.

545 44. Stowell SR, Ju T, Cummings RD. Protein glycosylation in cancer. Annu Rev Pathol.
546 2015;10: 473–510.

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

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

551 47. Zhu W, Leber B, Andrews DW. Cytoplasmic O-glycosylation prevents cell surface
552 transport of E-cadherin during apoptosis. EMBO J. 2001;20: 5999–6007.

553 48. Tsuboi S, Sutoh M, Hatakeyama S, Hiraoka N, Habuchi T, Horikawa Y, et al. A novel
554 strategy for evasion of NK cell immunity by tumours expressing core2 O-glycans. EMBO J.
555 2011;30: 3173–3185.

556 49. Ntziachristos P, Lim JS, Sage J, Aifantis I. From fly wings to targeted cancer therapies: a
557 centennial for notch signaling. Cancer Cell. 2014;25: 318–334.

558 50. McDaniell R, Warthen DM, Sanchez-Lara PA, Pai A, Krantz ID, Piccoli DA, et al.
559 NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling
560 pathway. Am J Hum Genet. 2006;79: 169–173.

561 51. Oda T, Elkahloun AG, Pike BL, Okajima K, Krantz ID, Genin A, et al. Mutations in the
562 human Jagged1 gene are responsible for Alagille syndrome. Nat Genet. 1997;16: 235–242.

563 52. Deng H, Jasper H. Sexual Dimorphism: How Female Cells Win the Race. Curr Biol.
564 2016;26: R212–5.

565 53. Sugathan A, Waxman DJ. Genome-wide analysis of chromatin states reveals distinct
566 mechanisms of sex-dependent gene regulation in male and female mouse liver. Mol Cell Biol.
567 2013;33: 3594–3610.

568 54. Garrido F, Aptsiauri N, Doorduijn EM, Garcia Lora AM, van Hall T. The urgent need to
569 recover MHC class I in cancers for effective immunotherapy. Curr Opin Immunol. 2016;39: 44–
570 51.

571 55. Beatty GL, Gladney WL. Immune escape mechanisms as a guide for cancer
572 immunotherapy. Clin Cancer Res. 2015;21: 687–692.

573 56. Akcay IM, Katrinli S, Ozdil K, Doganay GD, Doganay L. Host genetic factors affecting
574 hepatitis B infection outcomes: Insights from genome-wide association studies. World J
575 Gastroenterol. 2018;24: 3347–3360.

576 57. Schneider-Hohendorf T, Görlich D, Savola P, Kelkka T, Mustjoki S, Gross CC, et al. Sex
577 bias in MHC I-associated shaping of the adaptive immune system. Proc Natl Acad Sci U S A.
578 2018;115: 2168–2173.

579 58. Hass HG, Jobst J, Scheurlen M, Vogel U, Nehls O. Gene expression analysis for
580 evaluation of potential biomarkers in hepatocellular carcinoma. Anticancer Res. 2015;35: 2021–
581 2028.

582 59. Meng C, Shen X, Jiang W. Potential biomarkers of HCC based on gene expression and
583 DNA methylation profiles. Oncol Lett. 2018;16: 3183–3192.

584 60. Ferrín G, Aguilar-Melero P, Rodríguez-Perálvarez M, Montero-Álvarez JL, de la Mata
585 M. Biomarkers for hepatocellular carcinoma: diagnostic and therapeutic utility. Hepat Med.
586 2015;7: 1–10.

587 61. Todd JR, Ryall KA, Vyse S, Wong JP, Natrajan RC, Yuan Y, et al. Systematic analysis
588 of tumour cell-extracellular matrix adhesion identifies independent prognostic factors in breast
589 cancer. Oncotarget. 2016;7: 62939–62953.

590 62. Zembutsu H, Nakamura S, Akashi-Tanaka S, Kuwayama T, Watanabe C, Takamaru T, et
591 al. Significant Effect of Polymorphisms in CYP2D6 on Response to Tamoxifen Therapy for
592 Breast Cancer: A Prospective Multicenter Study. Clin Cancer Res. 2017;23: 2019–2026.

593 63. Hertz DL, Kidwell KM, Hilsenbeck SG, Oesterreich S, Osborne CK, Philips S, et al.
594 CYP2D6 genotype is not associated with survival in breast cancer patients treated with
595 tamoxifen: results from a population-based study. Breast Cancer Res Treat. 2017;166: 277–287.

596 64. Lu J, Li H, Guo P, Shen R, Luo Y, Ge Q, et al. The effect of CYP2D6 *10 polymorphism
597 on adjuvant tamoxifen in Asian breast cancer patients: a meta-analysis. Onco Targets Ther.
598 2017;10: 5429–5437.

599 65. Stevens JC, Marsh SA, Zaya MJ, Regina KJ, Divakaran K, Le M, et al. Developmental
600 changes in human liver CYP2D6 expression. Drug Metab Dispos. 2008;36: 1587–1593.

601 66. Raghuwanshi SK, Smith N, Rivers EJ, Thomas AJ, Sutton N, Hu Y, et al. G protein-
602 coupled receptor kinase 6 deficiency promotes angiogenesis, tumor progression, and metastasis. J
603 Immunol. 2013;190: 5329–5336.

604 67. Kang BW, Lee SJ, Lee YJ, Kim JG, Chae YS, Sohn SK, et al. Genetic variations in
605 miRNA binding site of TPST1 and ZG16B associated with prognosis for patients with colorectal
606 cancer. J Clin Orthod. American Society of Clinical Oncology; 2013;31: 3553–3553.

607 68. Goode EL, Chenevix-Trench G, Hartmann LC, Fridley BL, Kalli KR, Vierkant RA, et al.
608 Assessment of hepatocyte growth factor in ovarian cancer mortality. Cancer Epidemiol
609 Biomarkers Prev. 2011;20: 1638–1648.

610 69. Esposti DD, Hernandez-Vargas H, Voegele C, Fernandez-Jimenez N, Forey N, Bancel B,
611 et al. Identification of novel long non-coding RNAs deregulated in hepatocellular carcinoma
612 using RNA-sequencing. Oncotarget. 2016;7: 31862–31877.

613 70. Ling S, Hu Z, Yang Z, Yang F, Li Y, Lin P, et al. Extremely high genetic diversity in a
614 single tumor points to prevalence of non-Darwinian cell evolution. Proc Natl Acad Sci U S A.
615 2015;112: E6496–505.

616 71. Udali S, Guarini P, Ruzzenente A, Ferrarini A, Guglielmi A, Lotto V, et al. DNA
617 methylation and gene expression profiles show novel regulatory pathways in hepatocellular
618 carcinoma. Clin Epigenetics. 2015;7: 43.

619 72. Casey JP, McGettigan P, Lynam-Lennon N, McDermott M, Regan R, Conroy J, et al.
620 Identification of a mutation in LARS as a novel cause of infantile hepatopathy. Mol Genet Metab.
621 2012;106: 351–358.

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

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

627 75. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data. In:
628 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Internet]. 2010. Available:
629 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

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

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

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

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

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

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

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

647 83. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity
648 landscape. Genome Biol. 2017;18: 220.

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

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

653 86. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11:
654 expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool
655 enhancements. Nucleic Acids Res. 2017;45: D183–D189.

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

658 88. Leek JT, Storey JD. Capturing Heterogeneity in Gene Expression Studies by Surrogate
659 Variable Analysis. PLoS Genet. 2007;3: e161.

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

90. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8: 118–127.

91. Fusi N, Stegle O, Lawrence ND. Joint Modelling of Confounding Factors and Prominent Genetic Regulators Provides Increased Accuracy in Genetical Genomics Studies. PLoS Comput Biol. 2012;8: e1002330.

92. Goldinger A, Henders AK, McRae AF, Martin NG, Gibson G, Montgomery GW, et al. Genetic and Nongenetic Variation Revealed for the Principal Components of Human Gene Expression. Genetics. 2013;195: 1117.

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

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

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

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

97. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017;45: D896–D901.

98. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31: 3555–3557.

99. Lee M-H, Huang Y-H, Chen H-Y, Khor S-S, Chang Y-H, Lin Y-J, et al. Human leukocyte antigen variants and risk of hepatocellular carcinoma modified by hepatitis C virus genotypes: A genome-wide association study. Hepatology. 2017; doi:10.1002/hep.29531

# Figure Captions

**Fig 1. Sex-differences in gene expression in healthy liver, HCC tumor adjacent, and HCC tumor tissue.** X-linked genes are indicated in red, Y-linked in blue, and autosomal genes in black. An FDR-adjusted *p*-value threshold of 0.01 and absolute $\log_2$(FC) threshold of 2 were used to select significant genes. Multiple transcripts of the long non-coding RNA *XIST* are independently expressed.

695 **Fig 2. Expression and regulatory networks underlying male and female HCC. A:** Volcano
696 plots of DEGs in tumor vs. tumor adjacent comparisons in the sex-specific analyses and in joint
697 analysis of both sexes. X-linked genes are indicated as red, Y-linked as blue, and autosomal as
698 black. Significant genes were selected based on an FDR-adjusted *p*-value threshold of 0.01 and
699 absolute $\log_2$(FC) threshold of 2. **B:** Venn-diagram of the overlap of DEGs in the sex-specific
700 analyses and in joint analysis of both sexes. **C**: Overlap of functional pathways and canonical
701 networks enriched in males, females, and in the joint analysis of both sexes. **C**: Examples of sex-
702 specific and sex-shared pathways.

703 **Fig 3. Sex-specific genetic effects on gene expression. A:** QQ-plot of eQTL associations in the
704 joint analysis of both sexes (grey), male HCC (blue) and female HCC (red). **B:** Overlap of
705 eGenes detected in joint and stratified analyses. **C:** Genomic annotations of eVariants detected in
706 joint and stratified analyses. CDS: coding sequences. **D:** An example of a male-specific eQTL.
707 *POGLUT1* expression is modulated by a germline variant in *cis* in male HCC, but not in female
708 HCC nor in joint analysis of both sexes.

709 **Fig 4. An example of a gene under germline regulatory control with differential effects on**
710 **survival between the sexes. A:** Expression of *F12* in female and male HCC. **B-D:** Effect of *F12*
711 expression on survival in joint analysis of both sexes (**B**), female (**C**), and male (**D**). Cox's
712 proportional hazard ratio, FDR-adjusted *p*-value threshold 0.01. For visualization, samples with
713 normalized expression value >1 were selected as high expression and <-1 as low expression
714 samples.

715 # Supporting Information

716 **S1 Fig. Expression of homologous X and Y linked genes in the different sexes and tissues.**
717 For females, expression of X-linked gene is plotted. For males, combined expression of X- and
718 Y-linked genes is shown.

719 **S2 Fig. Absolute $\log_2$-fold changes of DEGs detected from tumor vs. tumor-adjacent**
720 **comparisons in the unstratified analysis of both sexes and in sex-specific analyses.** Absolute
721 log2-fold changes are given across all samples, female samples, and male samples.

722 **S3 Fig. Cell content of study samples based on cell-type enrichment analysis from bulk-**
723 **RNAseq data.**

724 **S4 Fig. PEER factors associated with continuous and categorical sample and**

725 **patient covariates. A-B:** $R^2$ and FDR-adjusted p-values between PEER factors and continuous
726 technical and biological covariates. **C-D:** ANOVA test statistic and FDR-adjusted *p*-values
727 between PEER factors and categorical technical and biological covariates.

728 **S1 Table. Sex-biased gene expression in GTEx liver tissue samples.** N males = 28, N females
729 = 22. Empirical *p*-value based on sex-label shuffling with 1000 permutations.

730 **Table S2. Sex-biased gene expression in TCGA LIHC tumor adjacent samples.** N males =
731 28, N females = 22. Empirical *p*-value based on sex-label shuffling with 1000 permutations.

732 **Table S3. Sex-biased gene expression in TCGA LIHC tumor samples.** N males = 28, N
733 females = 22. Empirical *p*-value based on sex-label shuffling with 1000 permutations.

734 **Table S4. Differentially expressed genes between matched tumor and tumor-adjacent**
735 **samples in joint analysis of male and female samples.** Each tissue N = 50, N female donors =
736 22, N male donors = 28.

737 **Table S5. Differentially expressed genes between matched male tumor and tumor-adjacent**
738 **samples.** N of sample pairs = 28.

739 **Table S6. Differentially expressed genes between matched female tumor and tumor-**
740 **adjacent samples.** N of sample pairs = 22.

741 **Table S7. Enriched GO terms and canonical networks in tumor vs. tumor-adjacent DEGs in**
742 **the joint analysis of both sexes.**

743 **Table S8. Enriched GO terms and canonical networks in male-specific tumor vs. tumor-**
744 **adjacent DEGs.**

745 **Table S9. Enriched GO terms and canonical networks in female-specific tumor vs. tumor-**
746 **adjacent DEGs.**

747 **Table S10. *cis*-eQTLs detected in joint analysis of male and female tumors.**

748 **Table S11. *cis*-eQTLs detected in male tumors.**

749 **Table S12. *cis*-eQTLs detected in female tumors.**

750 **Table S13. Genomic annotations of regulatory variants detected in joint analysis of both**
751 **sexes.**

752 **Table S14. Genomic annotations of regulatory variants detected in the male-specific**
753 **analysis.**

754 **Table S15. Genomic annotations of regulatory variants detected in the female-specific**
755 **analysis.**

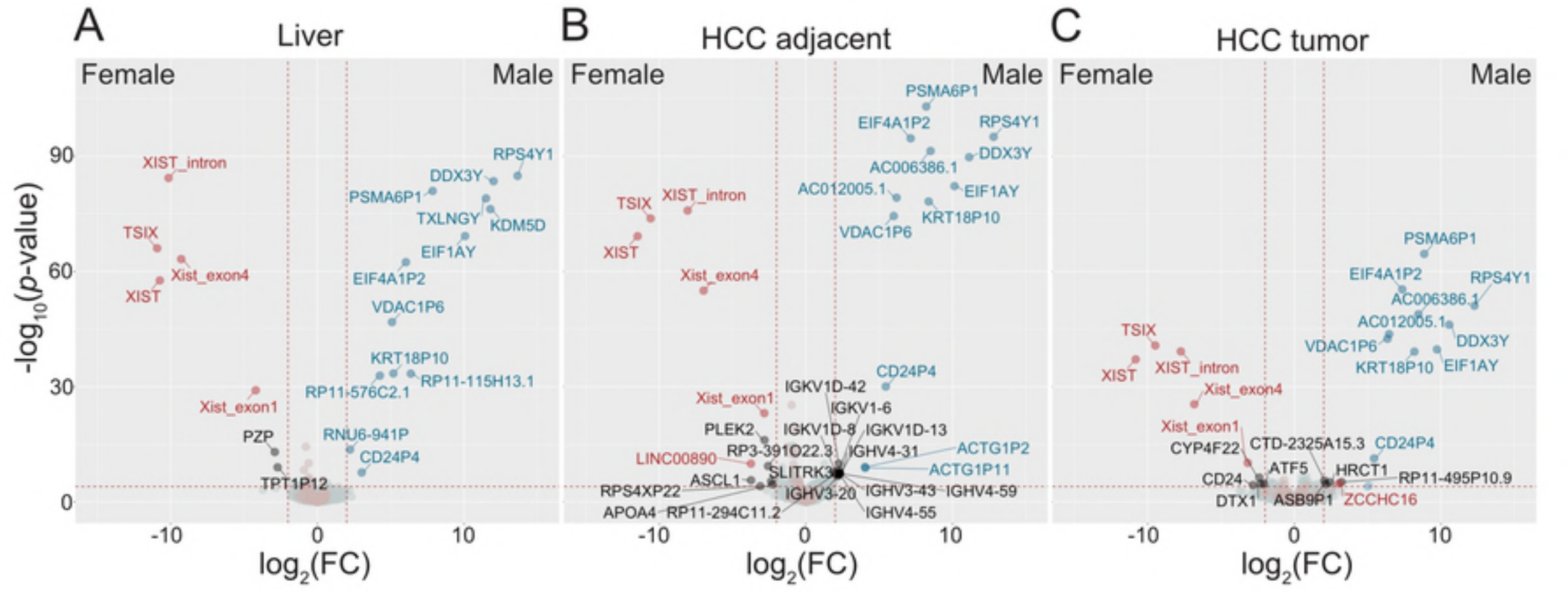
A
Liver
Female
Male
XIST_intron
TSIX
Xist_exon4
XIST
Xist_exon1
PZP
TPT1P12
RPS4Y1
DDX3Y
PSMA6P1
TXLNGY
KDM5D
EIF1AY
EIF4A1P2
VDAC1P6
KRT18P10
RP11-115H13.1
RP11-576C2.1
RNU6-941P
CD24P4
B
HCC adjacent
Female
Male
PSMA6P1
EIF4A1P2
RPS4Y1
DDX3Y
AC006386.1
AC012005.1
EIF1AY
KRT18P10
VDAC1P6
TSIX
XIST_intron
XIST
Xist_exon4
CD24P4
IGKV1D-42
IGKV1-6
Xist_exon1
PLEK2
IGKV1D-8
IGKV1D-13
RP3-391O22.3
IGHV4-31
ACTG1P2
LINC00890
SLITRK3
ACTG1P11
ASCL1
RPS4XP22
IGHV3-20
IGHV3-43
IGHV4-59
APOA4
RP11-294C11.2
IGHV4-55
C
HCC tumor
Female
Male
PSMA6P1
EIF4A1P2
RPS4Y1
AC006386.1
AC012005.1
VDAC1P6
DDX3Y
KRT18P10
EIF1AY
TSIX
XIST
XIST_intron
Xist_exon4
Xist_exon1
CYP4F22
CTD-2325A15.3
CD24P4
ATF5
HRCT1
CD24
RP11-495P10.9
DTX1
ASB9P1
ZCCHC16
$-\log_{10}(p\text{-value})$
90
60
30
0
-10
0
10
$\log_2(FC)$

Fig1

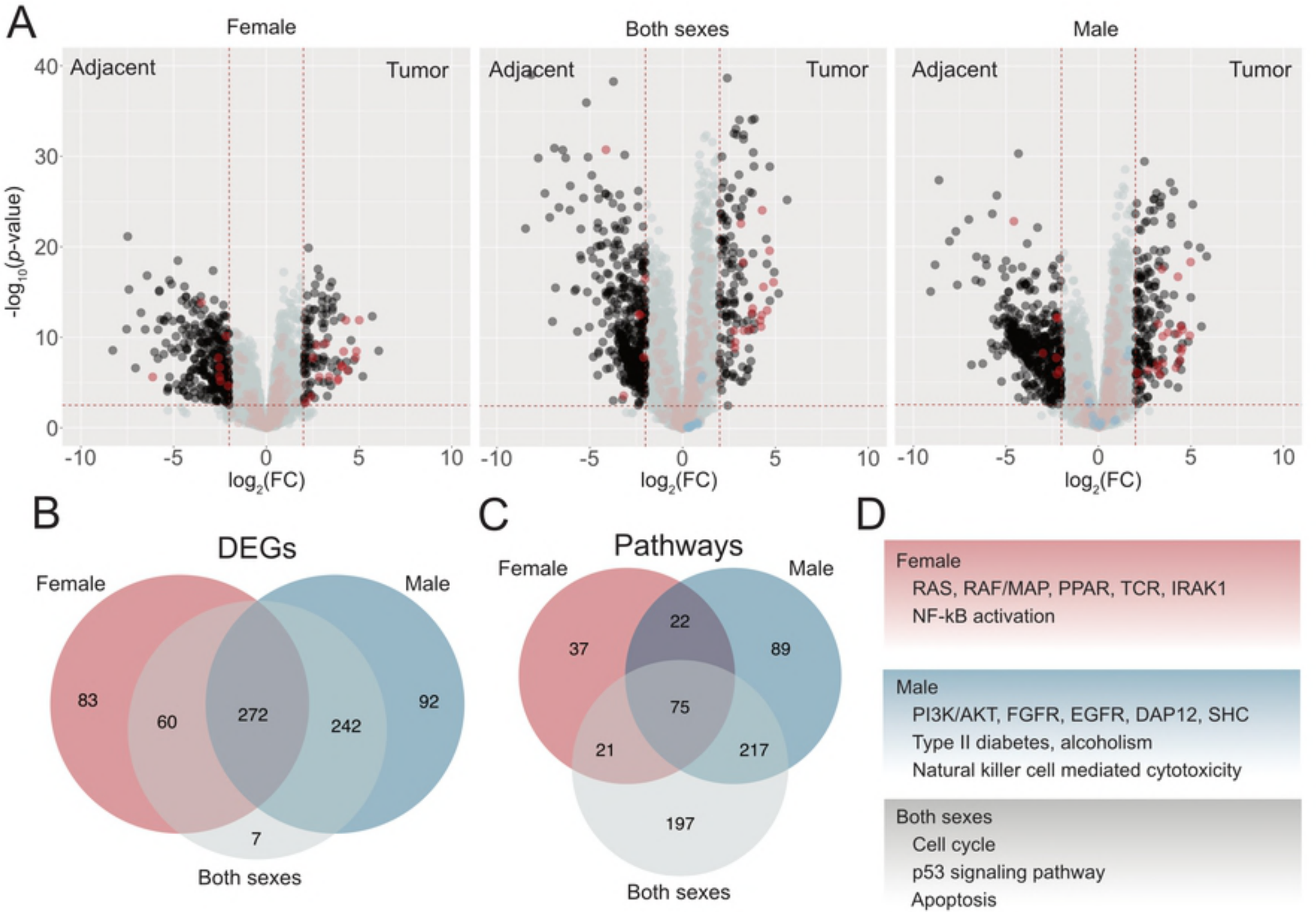
A
Female
Both sexes
Male
Adjacent
Tumor
-log10(p-value)
log2(FC)
B
DEGs
Female
Male
Both sexes
83
60
272
242
92
7
C
Pathways
Female
Male
Both sexes
37
22
89
75
21
217
197
D
Female
RAS, RAF/MAP, PPAR, TCR, IRAK1
NF-kB activation
Male
PI3K/AKT, FGFR, EGFR, DAP12, SHC
Type II diabetes, alcoholism
Natural killer cell mediated cytotoxicity
Both sexes
Cell cycle
p53 signaling pathway
Apoptosis

Fig2

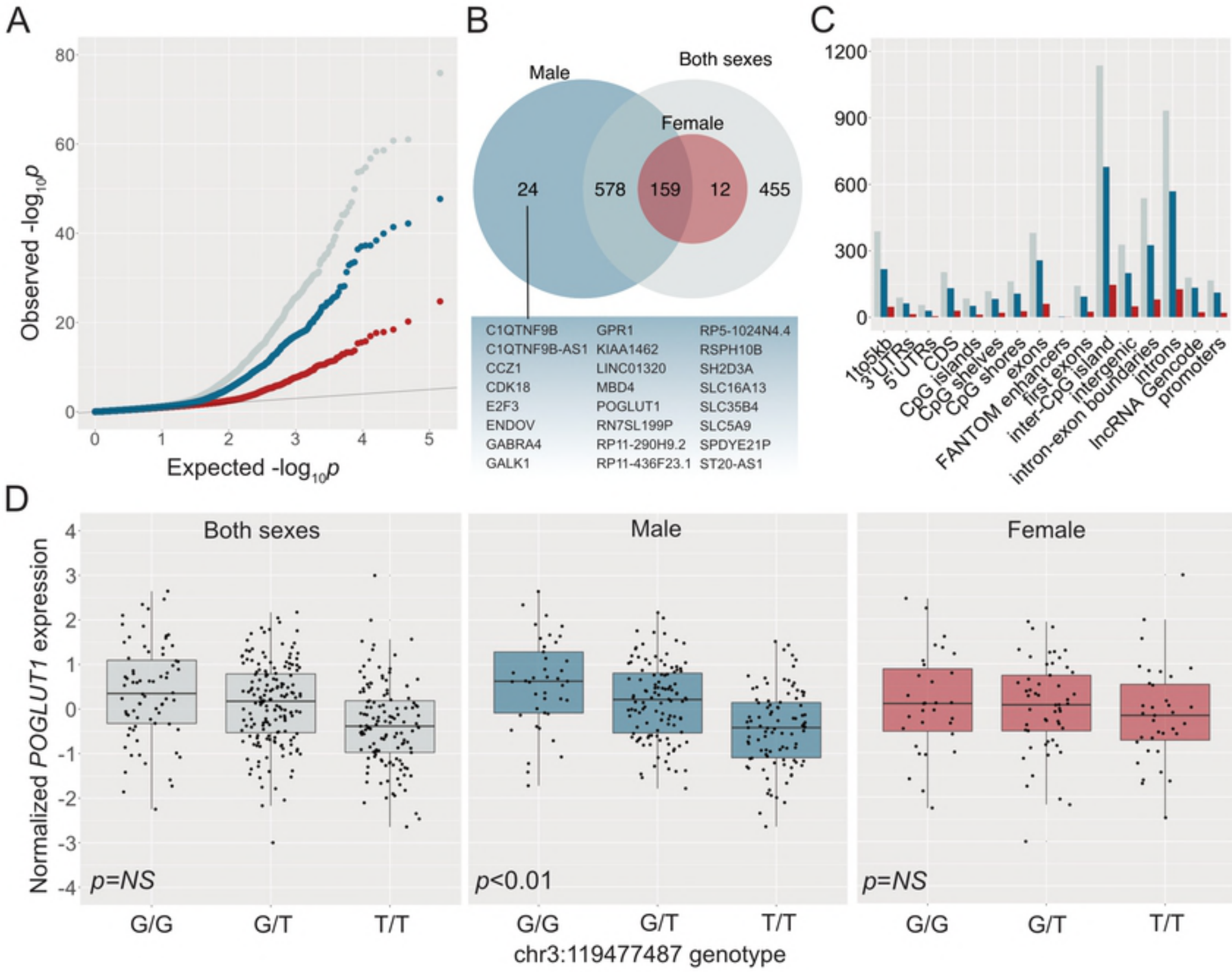
A
Observed -log10p
Expected -log10p
B
Male
Both sexes
Female
24
578
159
12
455
C1QTNF9B
C1QTNF9B-AS1
CCZ1
CDK18
E2F3
ENDOV
GABRA4
GALK1
GPR1
KIAA1462
LINC01320
MBD4
POGLUT1
RN7SL199P
RP11-290H9.2
RP11-436F23.1
RP5-1024N4.4
RSPH10B
SH2D3A
SLC16A13
SLC35B4
SLC5A9
SPDYE21P
ST20-AS1
C
1to5kb
3'UTRs
5'UTRs
CDS
CpG islands
CpG shelves
CpG shores
exons
FANTOM enhancers
first exons
inter-CpG island
intergenic
intron-exon boundaries
introns
lncRNA Gencode
promoters
D
Normalized POGLUT1 expression
Both sexes
Male
Female
p=NS
p<0.01
p=NS
G/G
G/T
T/T
chr3:119477487 genotype

Fig3

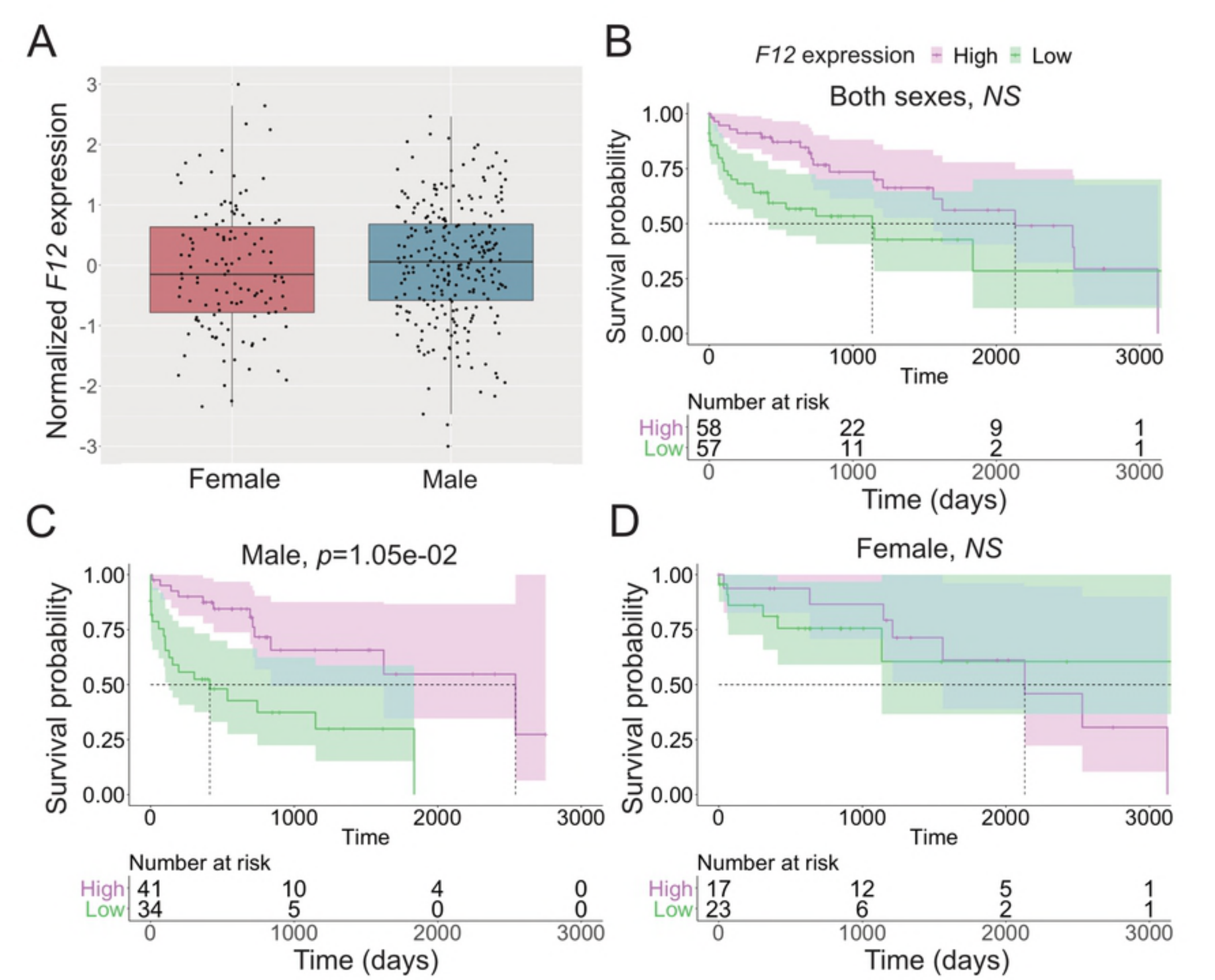
A
Normalized F12 expression
3
2
1
0
-1
-2
-3
Female
Male
B
F12 expression High Low
Both sexes, NS
Survival probability
1.00
0.75
0.50
0.25
0.00
0
1000
2000
3000
Time
Number at risk
High 58 22 9 1
Low 57 11 2 1
Time (days)
C
Male, p=1.05e-02
Survival probability
Time
Number at risk
High 41 10 4 0
Low 34 5 0 0
Time (days)
D
Female, NS
Survival probability
Time
Number at risk
High 17 12 5 1
Low 23 6 2 1
Time (days)

Fig4