HGGA, Volume *6*

Supplemental information

HCV- and HBV-mediated liver cancer converge on similar transcriptomic landscapes and immune profiles

Elizabeth S. Borden, Annika Jorgensen, Heini M. Natri, Karen Taraszka Hastings, Kenneth H. Buetow, and Melissa A. Wilson

Table of contents:

Figure S1: Quality control of sequencing data. Mean quality scores for each base position across the read. The four samples in red correspond to the two paired reads for tumor and normal from individual RK203. All other samples had sufficient quality to be included in the subsequent analyses.

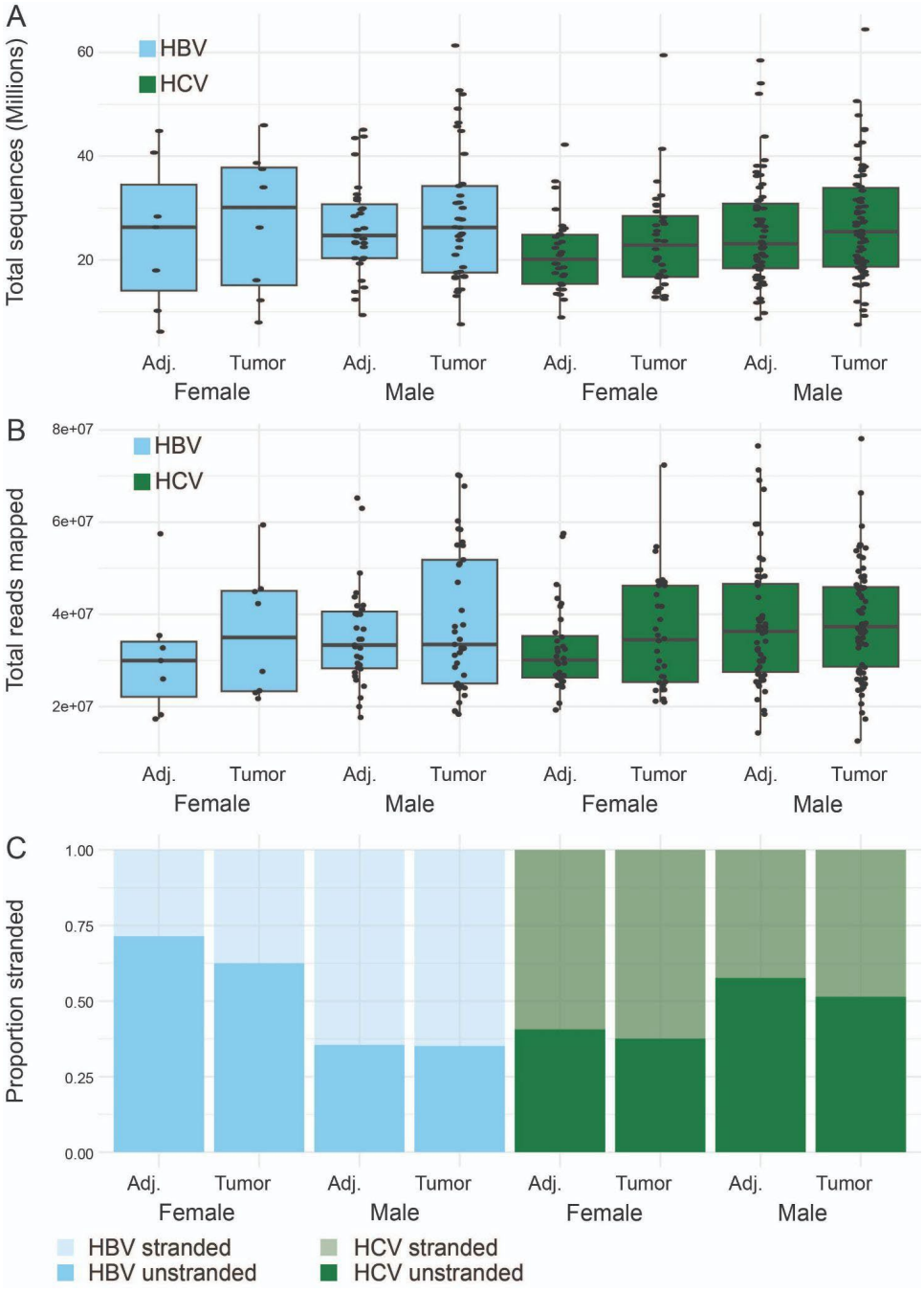

Figure S2: Comparison of sequencing depth and library type by group. Comparison of sequencing metrics separated by viral status, sex, and tumor vs. adjacent tissue. (A) Total number of sequenced reads. (B) Total number of mapped reads. Significance was tested for the total number of sequenced reads and the total number of mapped reads with a Kruskal-Wallis rank sum test and no significant difference was found. (C) The proportion of samples that had a stranded or unstranded library type. Significance was tested for the proportion of samples that had a stranded or unstranded library type using a chi-squared test and no significant difference was found.

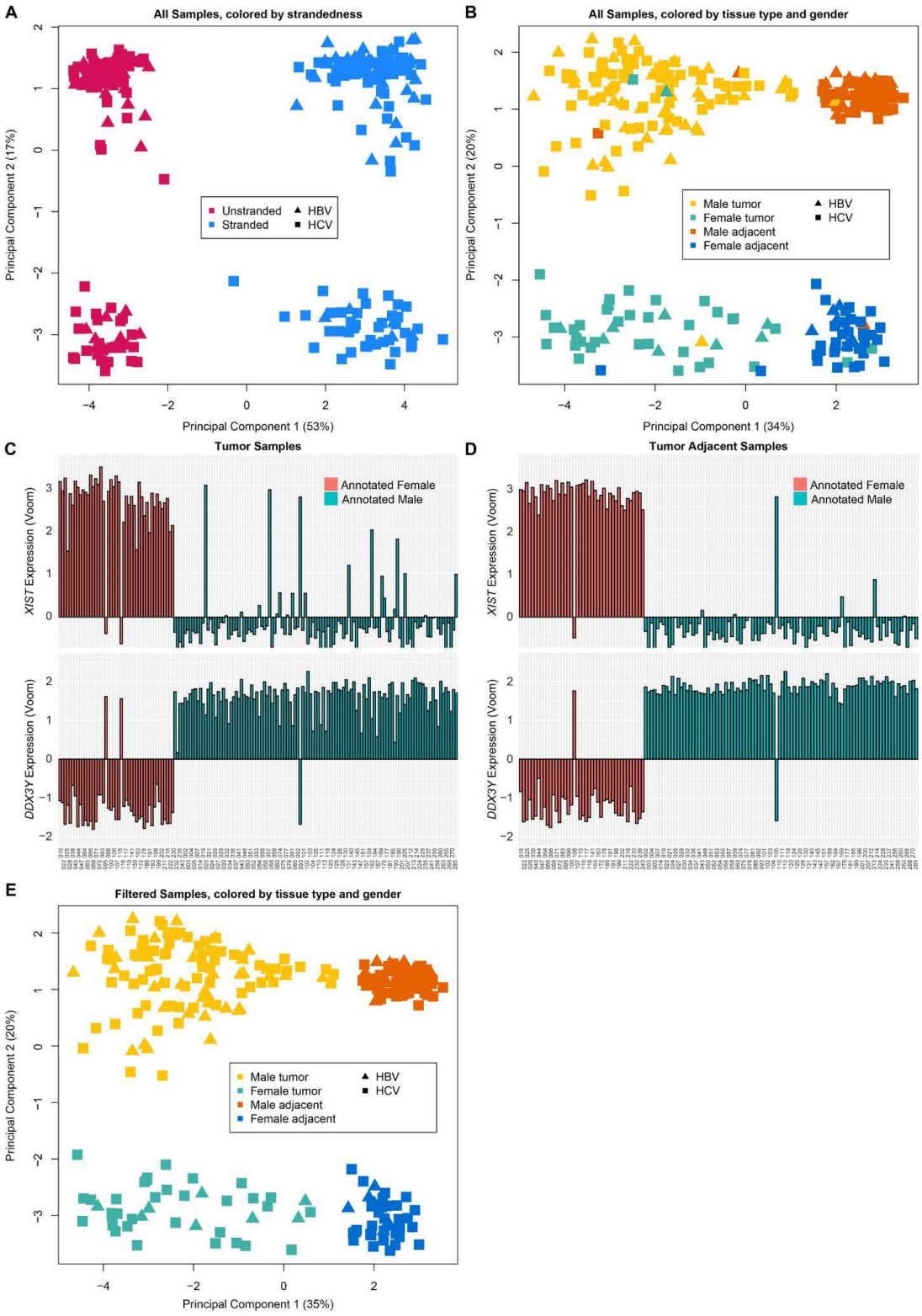

Figure S3: Quality control of all samples on tumor type and sex. MDS plot on the top 25 most variable genes, colored by **(A)** library type and **(B)** tumor status and sex after removal of library type batch effect. Plot of expression of *XIST* and *DDX3Y* across all **(C)** tumor samples and **(D)** tumor-adjacent samples. RK106, RK135, and RK105 were

removed from subsequent analyses due to likely mislabeled sex supported by the MDS plots and expression of *XIST* and *DDX3Y*. RK066, RK096, RK113, and RK116 were removed due to the proximity of the paired sample suggesting that there may be cross-contamination of the samples. Finally, RK169 and RK065 had tumor and tumor-adjacent samples that were in opposite clusters on the MDS plot, therefore, these samples were relabeled to be consistent with their observed clusters. **(E)** MDS plot on the top 25 most variable genes colored by tumor status and sex.

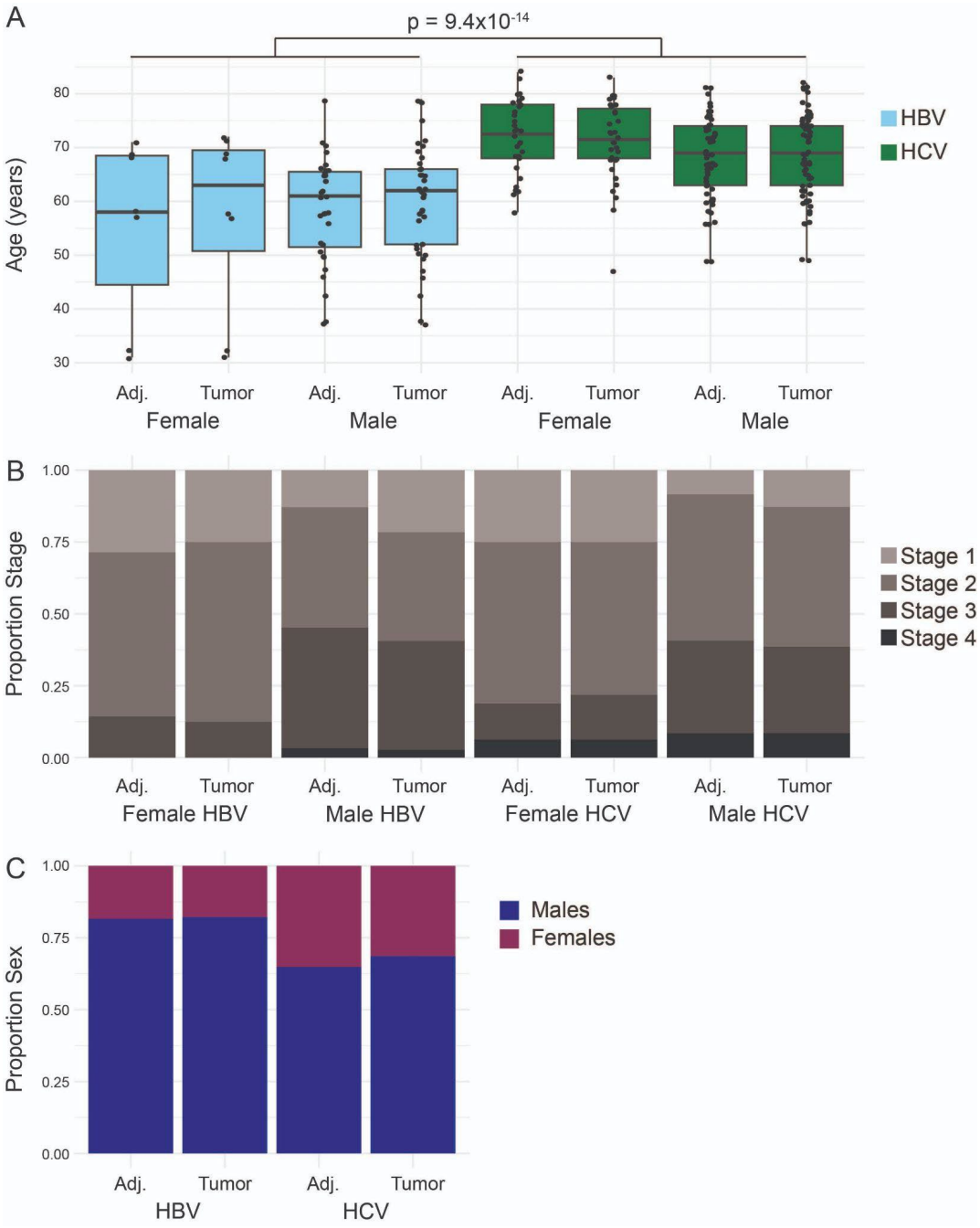

Figure S4: Comparison of demographics by group. Comparison of demographics by viral status, sex, and tumor vs. adjacent tissue. (A) Age in years. Significance tested between HBV and HCV with a Wilcoxon rank sum test. Significance was tested within HBV and HCV samples with a Kruskal-Wallis rank sum test and no significant differences were identified (B) Proportion of samples at each tumor stage. (C) Proportion of samples from male and female individuals. Significance was tested for the proportion of tumors at each stage and the proportions of samples from each sex using a chi-squared test and no significant differences were identified.

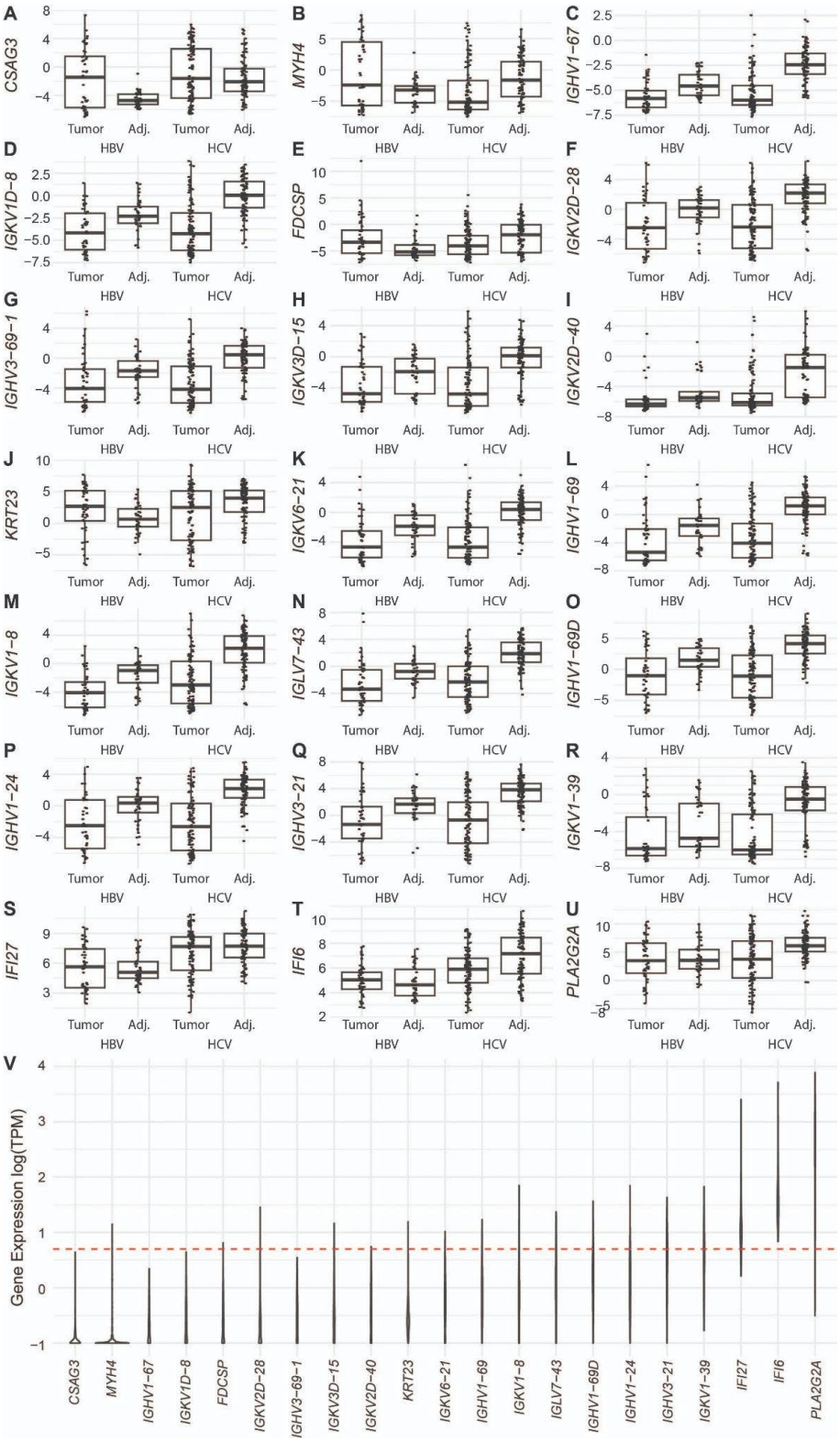

Figure S5: Expression of top differentially expressed genes from HBV tumor-adjacent to HCV tumor-adjacent tissue. (A-U) Boxplots demonstrating the *voom* normalized expression of each gene upregulated in HCV vs. HBV

tumor-adjacent samples across tumor and tumor-adjacent, HBV and HCV samples. **(V)** Violin plot of the expression of genes upregulated in HCV tumor-adjacent tissue compared to HBV tumor-adjacent tissue in healthy liver tissue from the GTEx dataset. Expression in GTEx is normalized as transcripts per million (TPM) and is displayed on a log base 10 scale. The horizontal red dashed line corresponds to 1 TPM of gene expression.

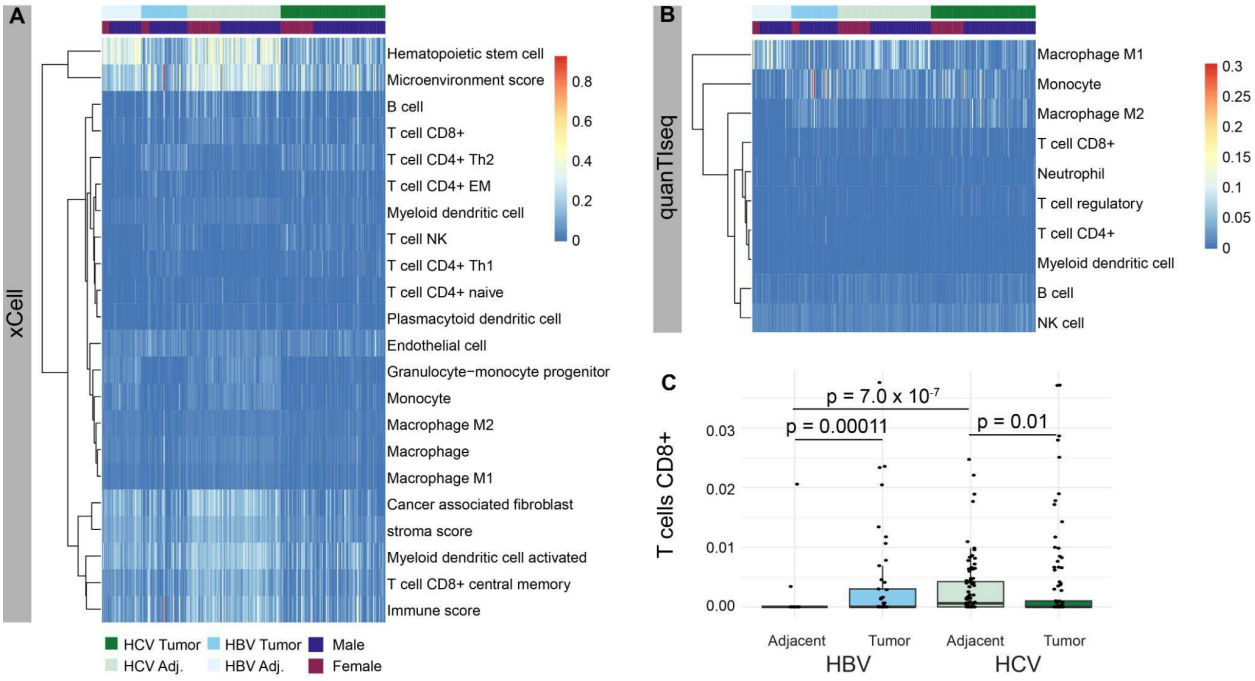

Figure S6: Full xCell and quanTIseq results across HBV and HCV samples. Heatmap of all immune cells identified with **(A)** xCell and **(B)** quanTIseq. Each column represents a single sample and the annotation bars across the top separate HCV tumor, HCV adjacent, HBV tumor, HBV adjacent, and male and female samples. Adj.; adjacent. **(C)** Comparison of quanTIseq quantified CD8+ T cells in tumor and tumor-adjacent tissue segregated by etiology. The bold line indicates the median and the upper and lower limits of the boxes indicate the 75th and 25th percentiles, respectively. The lower and upper whiskers indicate the minimum and maximum. Dots outside of the box and whiskers indicate outliers. Significance tested with a Wilcoxon rank sum test.

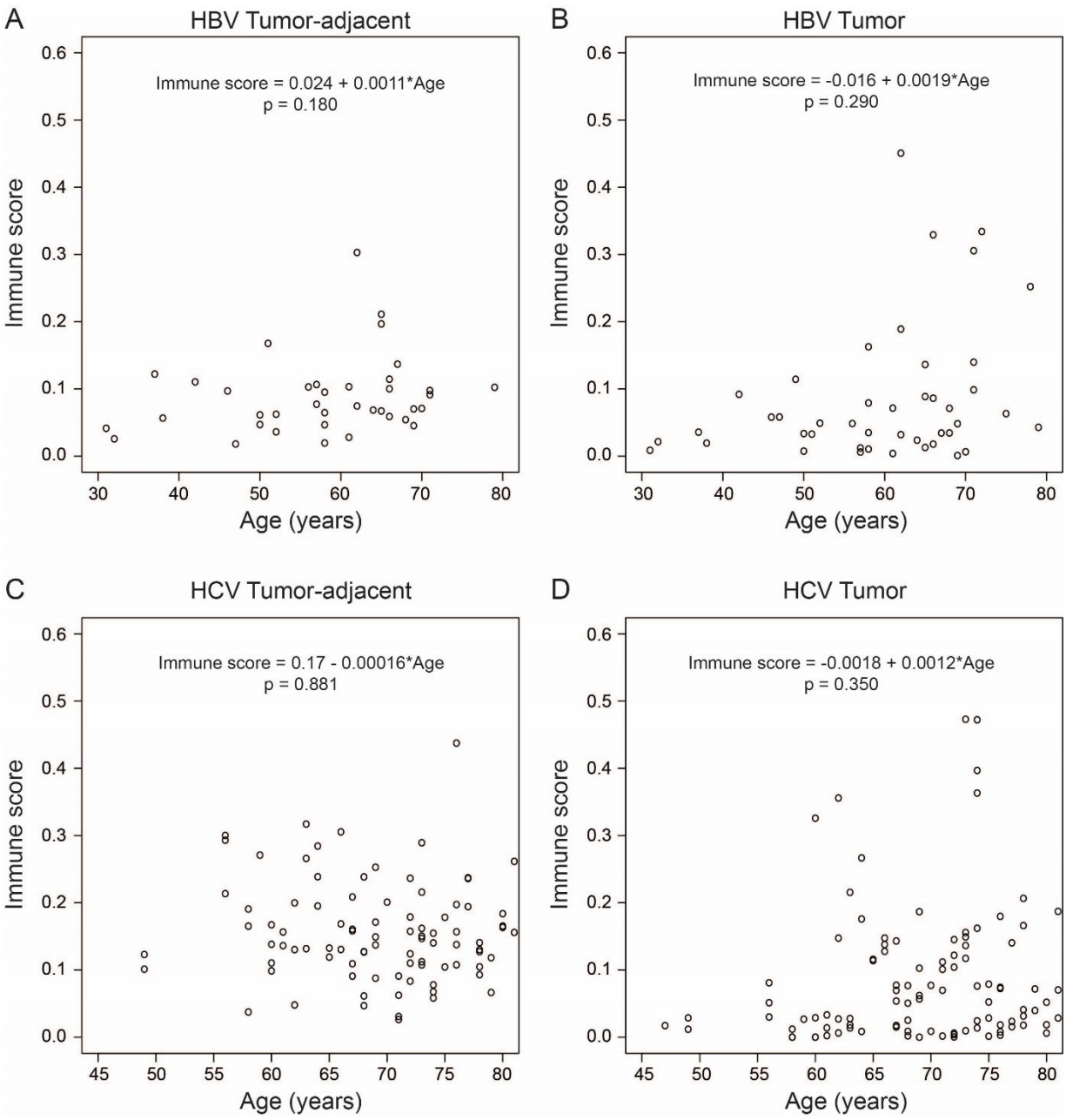

Figure S7: No association of immune infiltration and age. xCell immune score vs. age for (A) HBV tumor-adjacent samples, (B) HBV tumor samples, (C) HCV tumor-adjacent samples, and (D) HCV tumor samples. Linear regression models were fit to determine if there was a significant association for any of the four groups and the regression equation and p-value are provided.

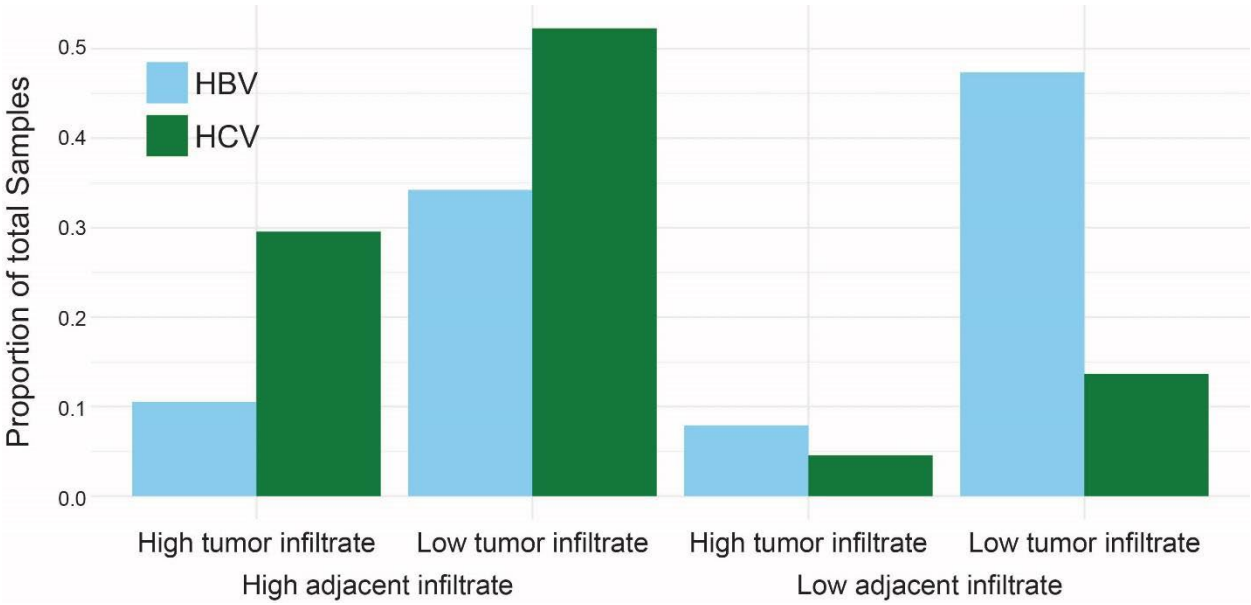

Figure S8: Comparison of infiltrate in tumor and adjacent matched samples. Comparison of the percent of HBV and HCV samples that fall into one of four categories of immune infiltration, high adjacent and tumor infiltration, high adjacent and low tumor infiltration, low adjacent and high tumor infiltration, and low adjacent and low tumor infiltration. High infiltration is defined as any infiltration above the median of all samples and low is defined as any infiltration below the median of all samples.

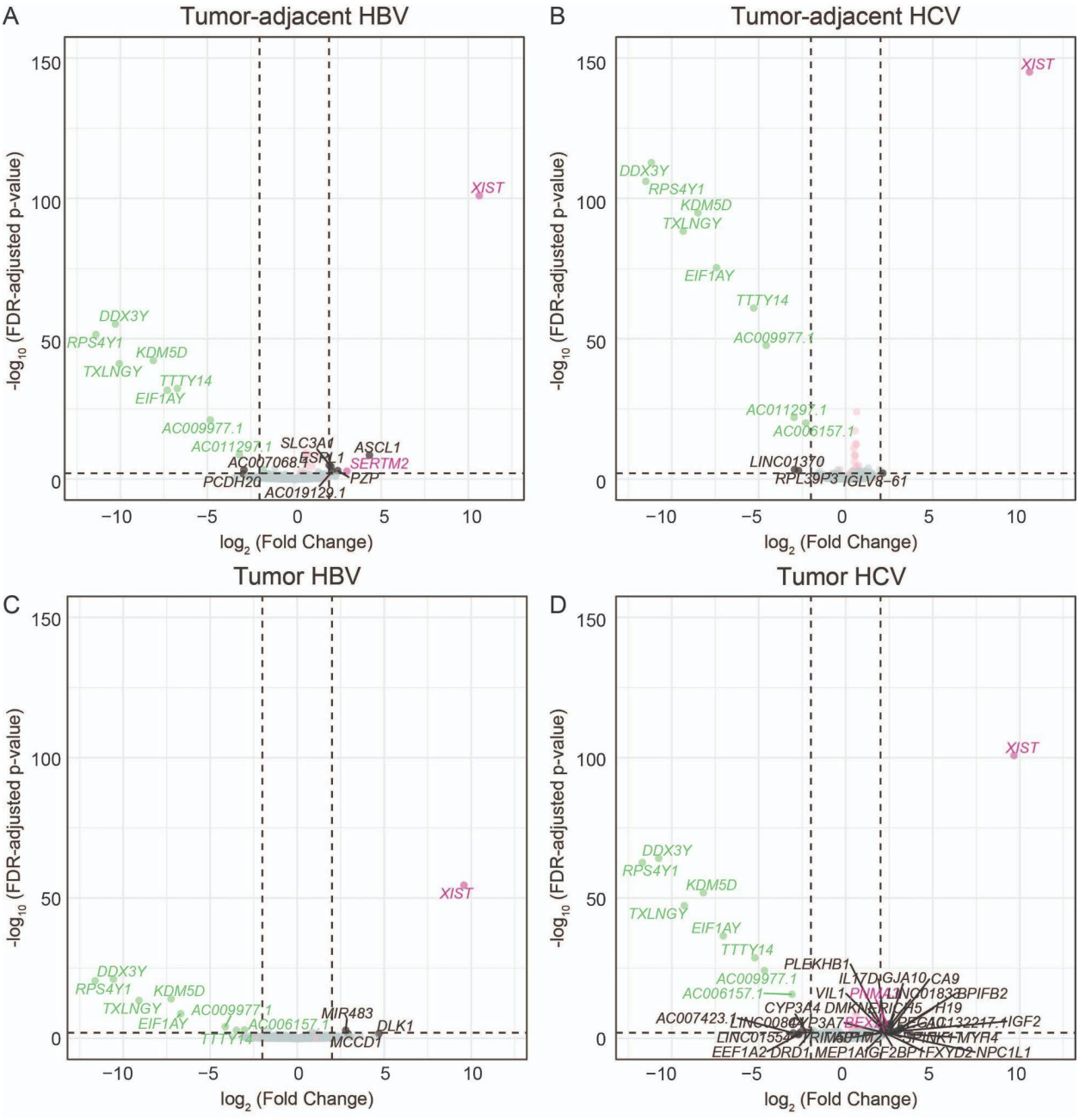

Figure S9: Differential expression in male and female samples. A-D) Volcano plots of differentially expressed genes from **(A)** male HBV tumor-adjacent:female HBV tumor-adjacent samples, **(B)** male HCV tumor-adjacent:female HCV tumor-adjacent samples, **(C)** male HBV tumor:female HBV tumor samples, and **(D)** male HCV tumor:female HCV tumor samples. X-linked genes are indicated in pink, Y-linked in green, and autosomal in black.

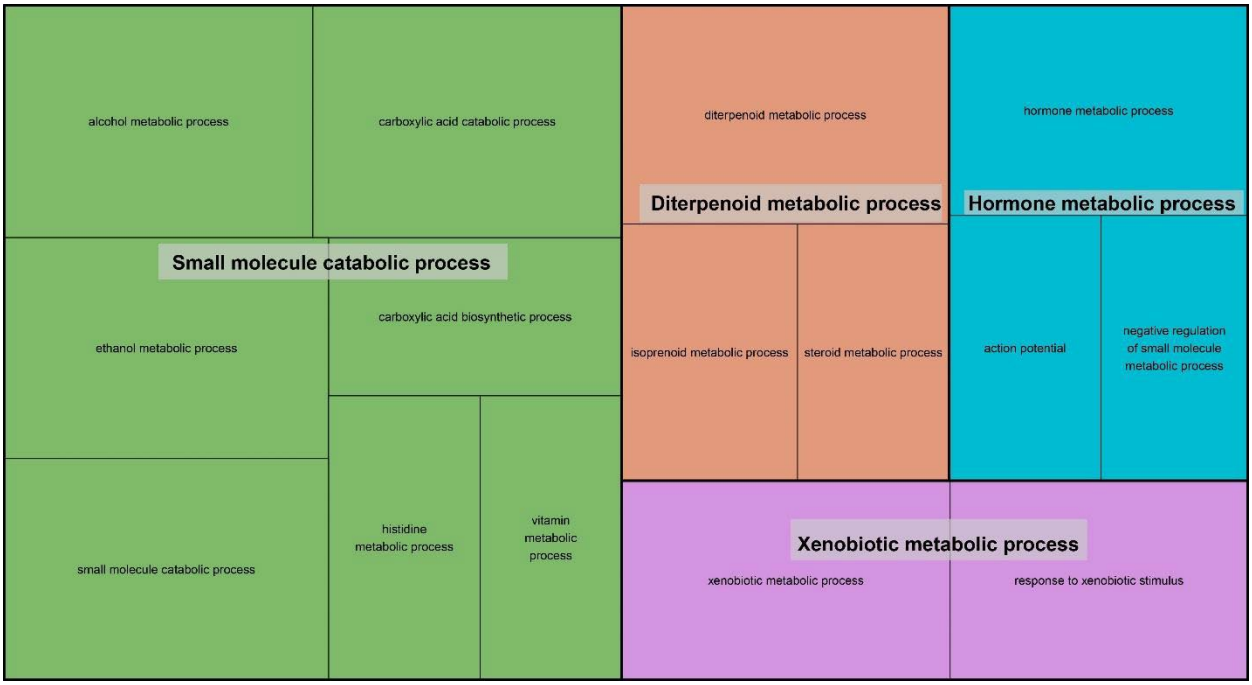

Figure S10: Pathways enriched in male HCV tumor:female HCV tumor comparison. Treemap visualization of GO enrichment analysis from all differentially expressed genes in the male HCV tumor:female HCV tumor comparison. The sizes of the boxes reflect the magnitude of the false-discovery adjusted p-value for the GO enrichment term.