

Supplementary materials for manuscript: “**Negative regulation of ACE2 by interferons *in vivo* and its genetic control**” by Ansari et al.

Supplementary Tables:

Supplementary Table 1: ACE2 negatively correlated genes. Only genes at 1% FDR and correlation coefficient of < -0.3 are shown (168 genes). To Calculate correlation, $\log_{10}(\text{TPM}+1)$ values calculated from RNASeq data was used. Correlation was measured using Pearson’s correlation coefficient as implemented in R function cor.test. Please see the attached csv and Excel files.

Supplementary Table 2: ACE2 positively correlated genes. Only genes at 1% FDR and correlation coefficient of > 0.3 are shown (1362 genes). To Calculate correlation, $\log_{10}(\text{TPM}+1)$ values calculated from RNASeq data was used. Correlation was measured using Pearson’s correlation coefficient as implemented in R function cor.test. Please see the attached csv and Excel files.

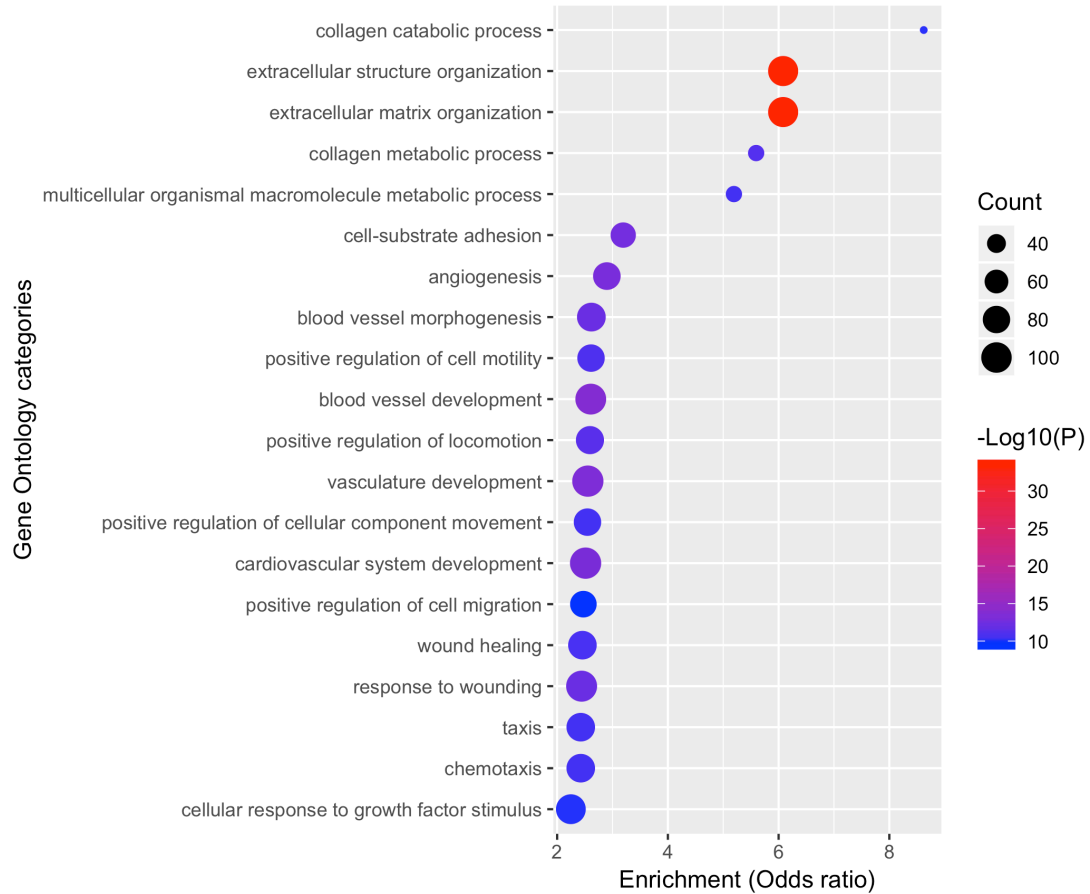
Supplementary Table 3: Gene set enrichment analysis of genes negatively correlated with ACE2 (at FDR of 1% and correlation coefficient of <-0.3). We used Gene Ontology (GO) gene sets from the “biological process” class with maximum gene set size of 500. Only GO categories that are significant at 1% FDR are included in this table. Please see the attached csv and Excel files.

Supplementary Table 4: Gene set enrichment analysis of genes positively correlated with ACE2 (at FDR of 1% and correlation coefficient of <-0.3). We used Gene Ontology (GO) gene sets from the “biological process” class with maximum gene set size of 500. Only GO categories that are significant at 1% FDR are included in this table. Please see the attached csv and Excel files.

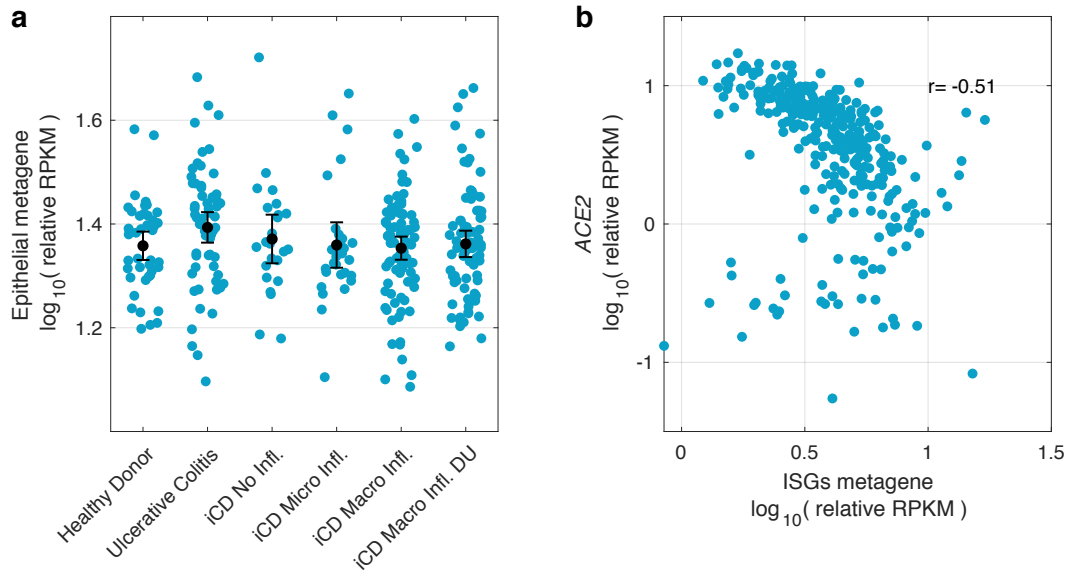
Supplementary Table 5: Measuring correlation in an independent cohort of HCV infected liver biopsies. Negative correlation was replicated for 162 genes out of 168 genes reported in **Supplementary table 1**. The results are not available for two genes as in this cohort as they were filtered out in the filtering step. Please see the attached csv and Excel files.

Supplementary Table 6: SNPs with p-values of less than 1×10^{-5} in at least one of the dominant or additive genome-wide eQTL analysis for ACE2 expression. Please see the attached csv and Excel files.

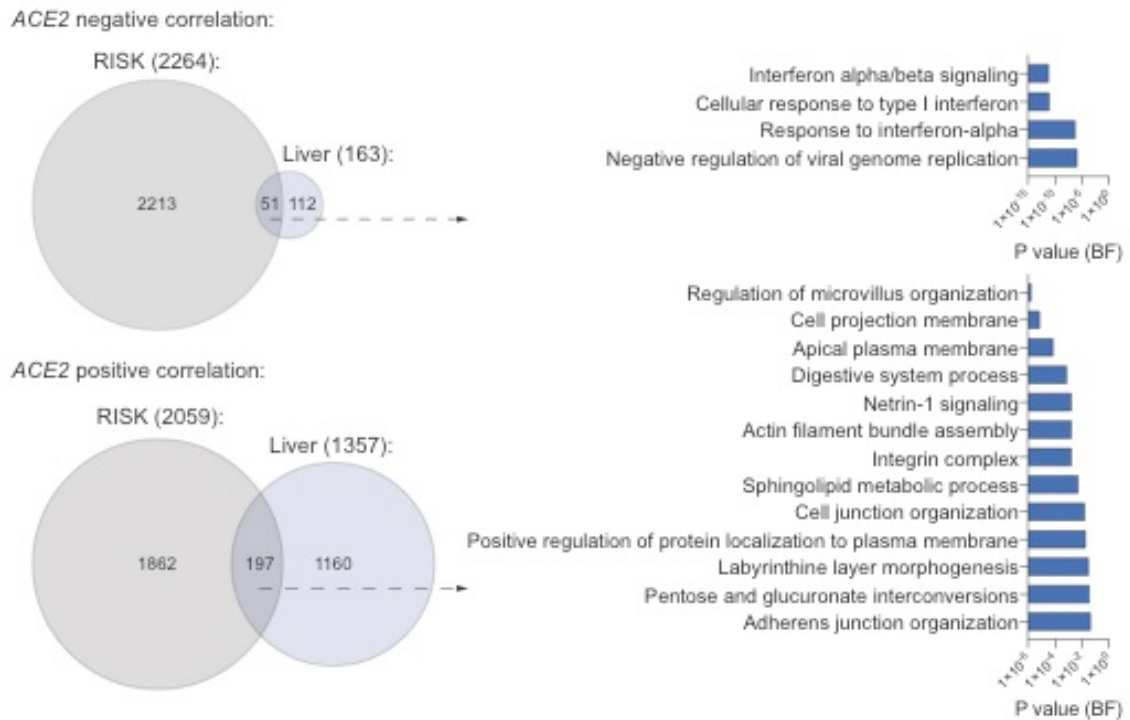
Supplementary Figures



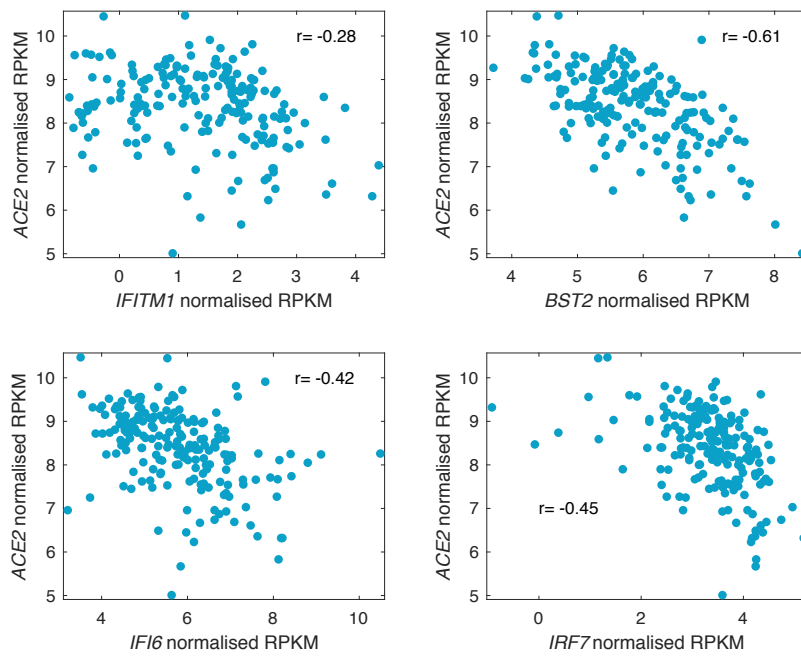
Supplementary Figure 1: Gene set enrichment analysis of genes whose expression are positively correlated (at FDR of 1% and correlation coefficient of > 0.3) with *ACE2* expression using Gene Ontology (GO) database. There are 189 GO categories that are significantly enriched among the positively correlated genes at 1% FDR. This figure only shows the top 20 most significant GO categories. For the full list see **Supplementary Table 4**.



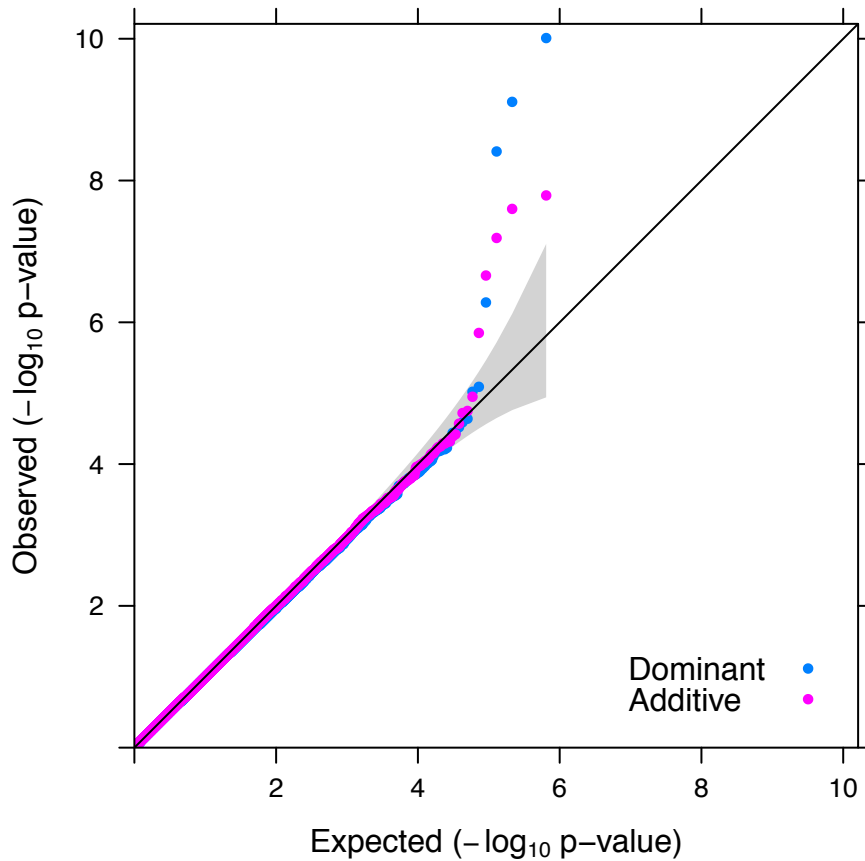
Supplementary Figure 2: (a) xCell-derived epithelial metagene expression in terminal ileum biopsies transcriptome of the RISK cohort. Patients were grouped based on disease severity and histologic assessment of inflammation (Ulcerative Colitis without ileal involvement; iCD = ileal Crohn’s disease; Micro. = Microscopic Inflammation, Macro. = Macroscopic Inflammation; DU = Deep Ulcers). (b) *ACE2* expression against ISGs metagene score (ISGs metagene was the list of genes that were negatively correlated with *ACE2* in liver and were part of GO category “response to type I interferon”, GO ID: 0034340). Pearson’s correlation coefficients are shown.



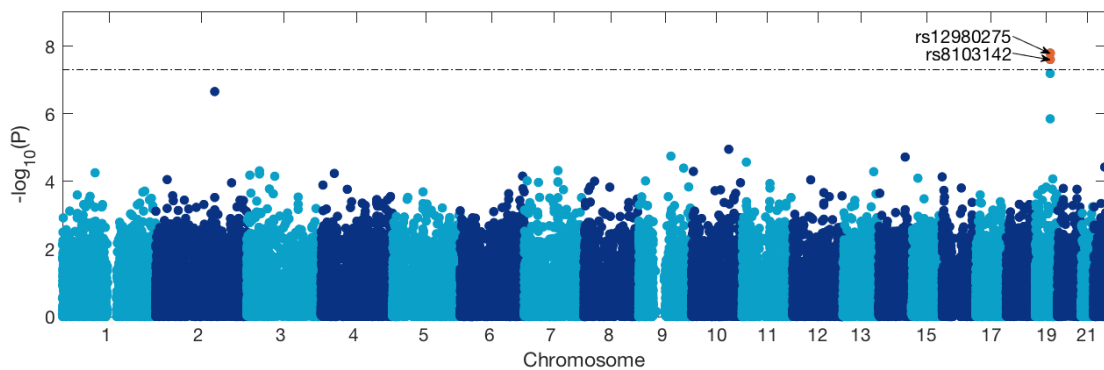
Supplementary Figure 3: Overlap of *ACE2* negatively and positively correlated genes in the RISK cohort intestinal transcriptome (Pearson's correlation coefficient > 0.5 and < -0.5 ; q-value < 0.01) and the liver gene expression data (Pearson's correlation coefficient > 0.3 and < -0.3 ; q-value < 0.01), and enrichment of Gene Ontology gene sets for the intersection of genes.



Supplementary Figure 4: Expression (normalised RPKM) and correlation of representative interferon stimulated genes (ISG) and ACE2 in the Genesis cohort. Pearson's correlation coefficients are shown.



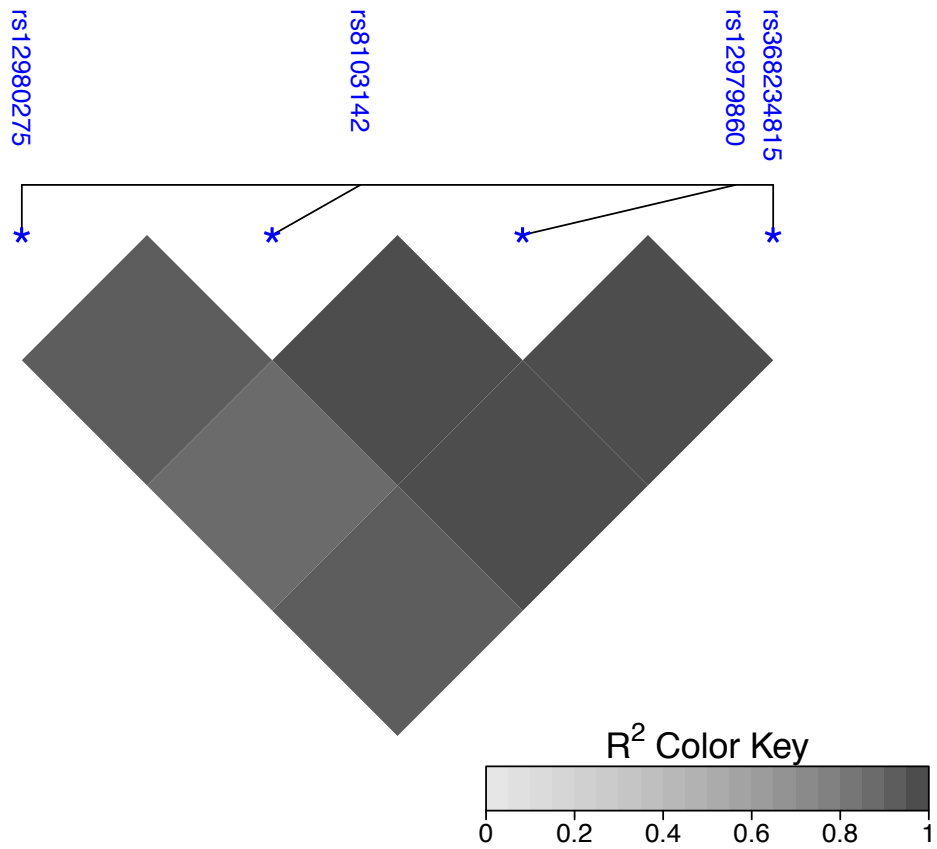
Supplementary Figure 5: QQplot for the *ACE2* genome-wide eQTL analysis for both additive and dominant genetic models. There is no inflation of the p-values which indicates host population structure is accounted for.



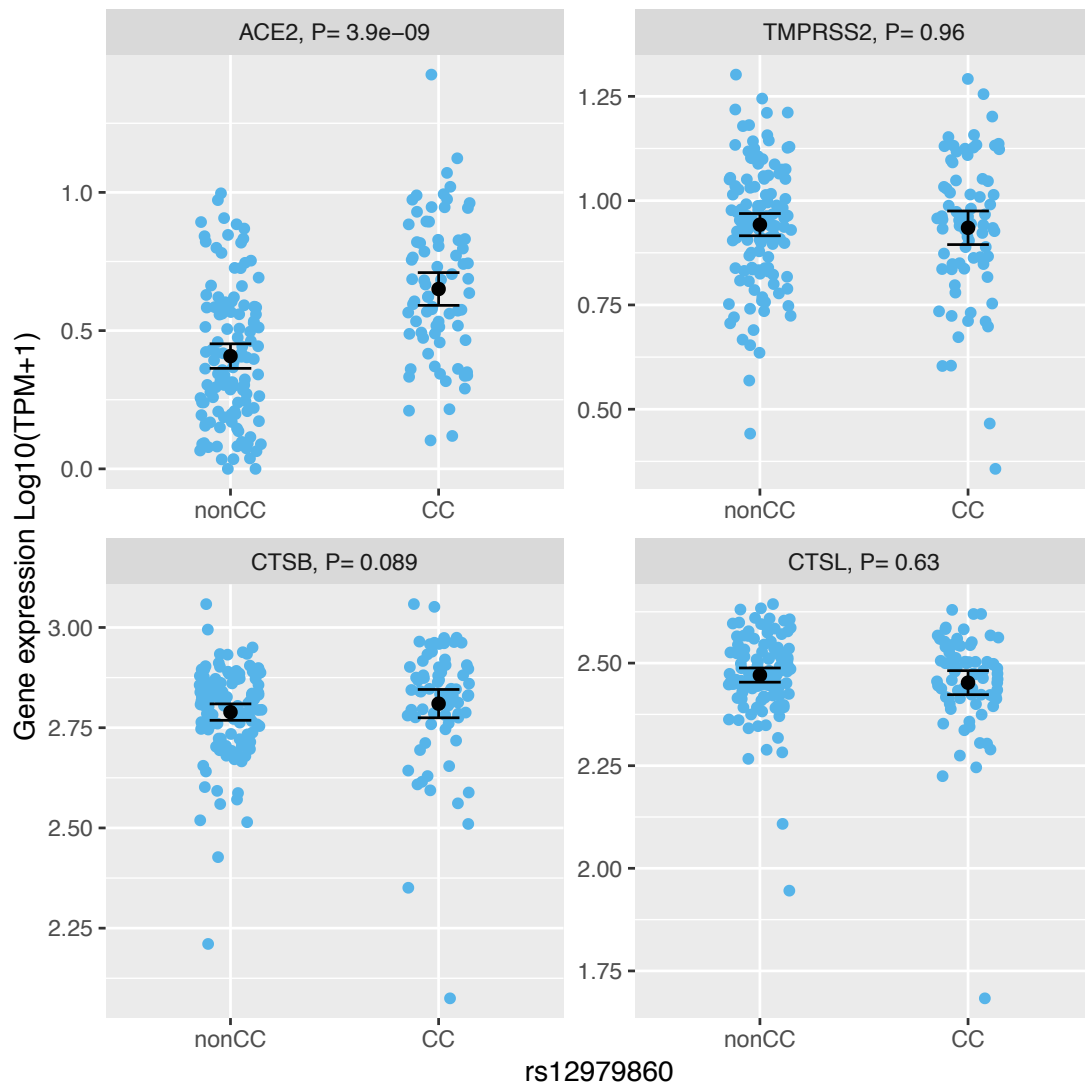
Supplementary Figure 6: Manhattan plot of *ACE2* genome-wide eQTL analysis using additive genetic model for liver biopsy samples from 190 HCV infected patients.

Pairwise LD

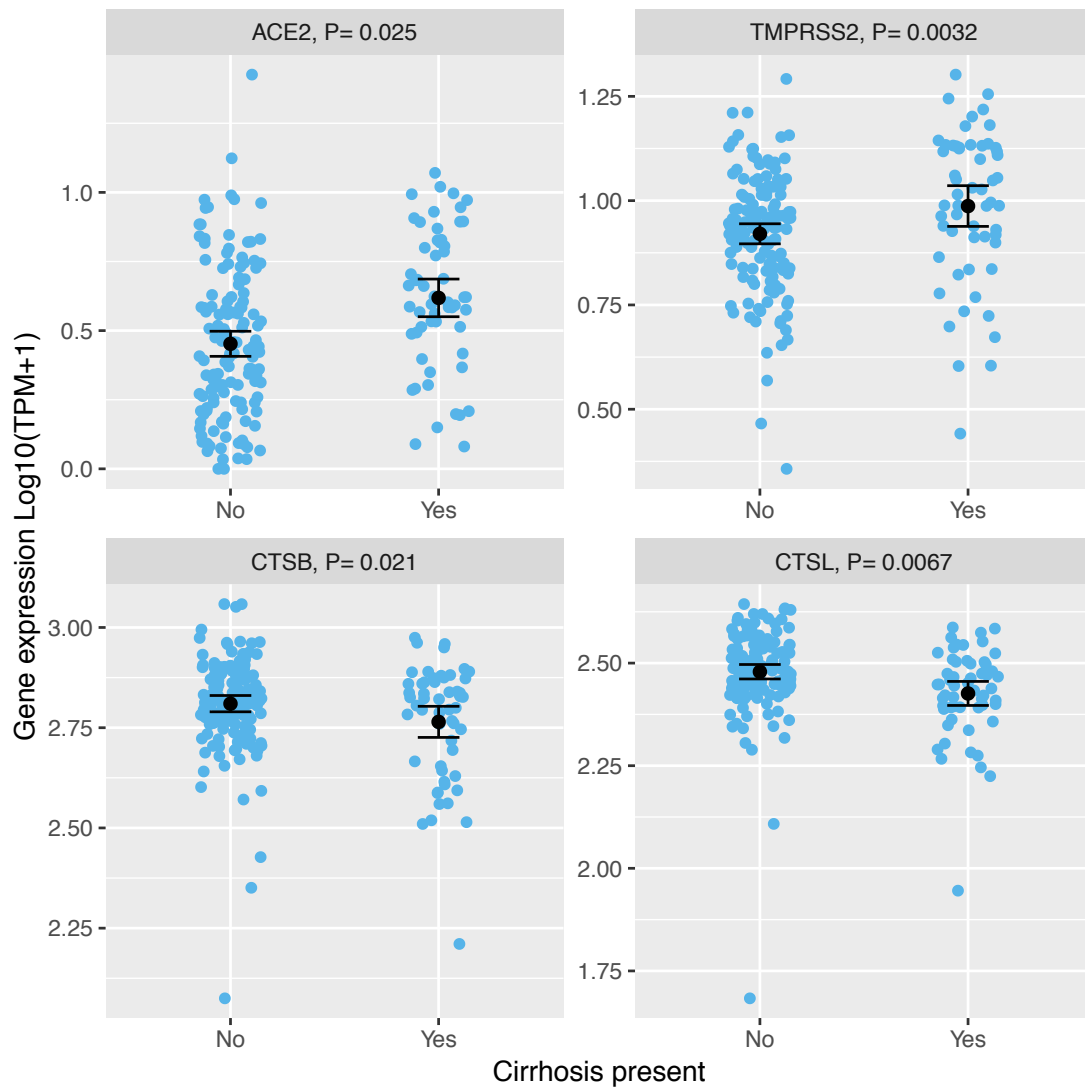
Physical Length:7.4kb



Supplementary Figure 9: Pairwise LD plot between the three typed SNPs which are significantly associated with *ACE2* expression and the likely causal SNP rs368234815 (not present on our genotyping array) from the 1000 genome project in the CEU population (Utah residents with Northern and Western European ancestry). The line represents the physical distance between SNPs.



Supplementary Figure 10: Expression of SARS-CoV-2 receptor (*ACE2*) and three other genes (*TMPRSS2*, *CTSB* and *CTSL*) that may be required for the virus entry into the cell. Distribution of gene expression stratified by *IFNL4* SNP rs12979860 genotypes using a dominant model (C/C vs C/T and T/T). The black dot and line indicates the mean and its 95% confidence interval. The p-value of association test is shown in the title of each panel. The estimated p-value is from a linear regression where log10(TPM+1) of expression was used as the dependent variable and the covariates included were cirrhosis status, age, sex and the first five host genetic principal components. TPM: transcripts per million.



Supplementary Figure 11: Expression of SARS-CoV-2 receptor (*ACE2*) and three other genes (*TMPRSS2*, *CTSB* and *CTSL*) that may be required for entry into the cell. Distribution of gene expression stratified by patient cirrhosis status. The black dot and line indicates the mean and its 95% confidence interval. The p-value of the association test is shown in the title of each panel. The estimated p-value is from a linear regression where $\log_{10}(\text{TPM}+1)$ of expression was used as the dependent variable and the explanatory variable included were cirrhosis status, age, sex and the first five host genetic principal components. TPM: transcripts per million.