**In vivo negative regulation of SARS-CoV-2 receptor, ACE2, by interferons and its genetic control** [version 1; peer review: awaiting peer review]

M. Azim Ansari1-4*, Emanuele Marchi1-3*, Narayan Ramamurthy1-3*, Dominik Aschenbrenner1,3, Sophie Morgan3,5, Carl-Philipp Hackstein2, STOP-HCV consortium, ISARIC-4C Investigators, Shang-Kuan Lin4, Rory Bowden4, Eshita Sharma4, Vincent Pedergnana6, Suresh Venkateswaran7, Subra Kugathasan7, Angela Mo8, Greg Gibson8, Graham Cooke8, John McLauchlan10, John Kenneth Baillie11,12, Sarah Teichmann8, Alexander Mentzer14, Julian Knight4, John A. Todd3,4, Timothy Hinks3,14, Eleanor Barnes1-3, Holm Uhlig1,3,5, Paul Klenerman1-3

1Translational Gastroenterology Unit, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 9DU, UK
2Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine, University of Oxford, Oxford, OX1 3SY, UK
3NIHR Biomedical Research Centre, John Radcliffe Hospital, Oxford, OX3 9DU, UK
4Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK
5Department of Paediatrics, University of Oxford, Oxford, OX3 9DU, UK
6Laboratory MIVEGEC (CNRS, IRD, UM), French National Centre for Scientific Research (CNRS), Montpellier, France
7Department of Pediatrics, Emory University School of Medicine and Children’s health care of Atlanta, Atlanta, USA
8Center for Integrative Genomics, Georgia Institute of Technology, Atlanta, USA
9Department of Infectious Disease, Faculty of Medicine, Imperial College London, London, W2 1PG, UK
10MRC-University of Glasgow Centre for Virus Research, Sir Michael Stoker Building, University of Glasgow, Glasgow, G61 1QH, UK
11Edinburgh Royal Infirmary, 51 Little France crescent, Edinburgh, EH16 4SA, UK
12Genetics and Genomics, Roslin Institute, University of Edinburgh, Edinburgh, EH25 9RG, UK
13Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
14Respiratory Medicine Unit, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 9DU, UK

* Equal contributors

---

**Abstract**

**Background:** Angiotensin I converting enzyme 2 (ACE2) is a receptor for the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and differences in its expression may affect susceptibility to infection. **Methods:** We performed a genome-wide expression quantitative trait loci (eQTL) analysis using hepatitis C virus-infected liver tissue from 190 individuals. **Results:** We discovered that polymorphism in a type III interferon...
gene (IFNL4), which eliminates IFN-λ4 production, is associated with a two-fold increase in ACE2 RNA expression. Conversely, among genes negatively correlated with ACE2 expression, IFN-signalling pathways were highly enriched and ACE2 was downregulated after IFN-α treatment. Negative correlation was also found in the gastrointestinal tract where inflammation driven IFN-stimulated genes were negatively correlated with ACE2 expression and in lung tissue from a murine model of SARS-CoV-1 infection suggesting conserved regulation of ACE2 across tissue and species.

**Conclusions:** We conclude that ACE2 is likely a negatively-regulated interferon-stimulated gene (ISG) and carriage of IFNL4 gene alleles which modulates ISGs expression in viral infection may play a role in SARS-CoV-2 pathogenesis with implications for therapeutic interventions.

**Keywords**
ACE2, IFNL4, HCV, ISG

This article is included in the Wellcome Sanger Institute gateway.

This article is included in the Coronavirus (COVID-19) collection.
Introduction

Entry of coronaviruses into susceptible cells depends on the binding of the spike (S) protein to a specific cell-surface protein and subsequent S protein priming by cellular proteases. Similar to severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1), infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; coronavirus disease 2019 (COVID-19) virus) employs Angiotensin I converting enzyme 2 (ACE2) as a receptor for cellular entry.

Epidemiological studies have indicated that the risk for serious disease and death from COVID-19 is higher in males, in older individuals and those with co-morbidities and it varies across ethnic groups. Host genetic variation is important in determining susceptibility and disease outcome for many infectious diseases and it is likely to be important in determining SARS-COV-2 susceptibility and outcome. Polymorphisms in the host genome could drive differences in ACE2 expression, which may affect susceptibility to SARS-CoV-2 infection and its consequences. Thus a better understanding of ACE2 expression, its regulatory mechanisms in vivo and its association with host genetics, especially during viral infection will provide insights on SARS-CoV-2 pathogenesis and help in repurposing antiviral drugs and development of vaccine strategies.

Additionally host genetic variation could impact on differential immune responses to the infection. The earliest immune defence mechanism activated upon virus invasion is the innate immune system. Virus-induced signalling through innate immune receptors prompts extensive changes in gene expression and it has been shown that host genetics contribute to transcriptional heterogeneity in response to infections.

Therefore, in the context of infectious diseases, it is important to investigate infected tissue to observe infection-triggered immune response heterogeneity and to understand the role of host genetics. For instance in the context of hepatitis C virus (HCV) infection, hepatic interferon (IFN)-stimulated genes (ISGs) induction varies considerably between individuals with some patients showing constant ISGs expression at high levels while others show almost no detectable induction of innate immune system. This differential innate immune activation is strongly associated with the genetic variation in the IFNL locus on chromosome 19q13.2. The causal variant is likely to be the dinucleotide exonic variant rs368234815 in the IFNL4 gene. This variant (ΔG > TT) results in a frameshift, abrogating production of functional IFN-λ4 protein. Lack of production of IFN-λ4 in individuals carrying rs368234815 TT/TT genotype is associated with no or low levels of expression of liver ISGs, higher viral load and paradoxically with higher rates of spontaneous clearance and treatment response to IFN-α and direct-acting antivirals. Lack of production of IFN-λ4 is also associated with better outcome of RNA virus respiratory tract infections (including coronaviruses) in children. This locus has also been associated with viral evolution. It has been shown that IFN-λ4 is highly conserved in mammals and therefore functionally relevant, but in humans, the dinucleotide insertion (rs368234815 TT allele) has a gradient in frequency that rises from Africa (0.29–0.44) to Europe (0.58–0.77) and reaches near fixation in East Asia (0.94–0.97) indicating positive selection has favoured the elimination of IFN-λ4 in humans. The molecular link between the genotype and the phenotype has not been fully defined, in part due to limitations in detection of IFNL proteins and mRNAs in the liver biopsies of patients with chronic HCV infection.

To understand the impact of IFNL locus and other host genetic factors on ACE2 expression in the presence of RNA virus infection, we performed a genome-wide eQTL analysis for ACE2 expression in 190 HCV-infected liver biopsies. We observed that host genetic polymorphism of the IFNL region tagging a variant that eliminates IFN-λ4 production was significantly associated with increase in ACE2 RNA expression. We also observed that increase in age and presence of liver cirrhosis were associated with increased ACE2 expression. Additionally, we identified negative correlation of ACE2 with ISGs in virus infected liver biopsies. We discovered the same pattern in gastrointestinal tract where inflammation driven ISGs were negatively correlated with ACE2 expression in two independent cohorts. The interferon-associated down-regulation of ACE2 was also identified in lung tissue in a murine model of SARS-CoV-1 infection. Furthermore, we detected downregulation of ACE2 transcripts after IFN-α treatment in virus-infected liver biopsies. Due to the conserved pattern of down-regulation of ACE2 in presence of up-regulation of ISGs across tissues, inflammatory responses, infections and species, we conclude that ACE2 is likely a negatively-regulated ISG and the genetic variation in the IFNL locus which modulates ISGs expression in viral infection may potentially play a role in SARS-CoV-2 pathogenesis.

Results

To understand the link between host genetics and ACE2 expression in presence of RNA virus infection, we used genotyped autosomal SNPs in the host genome to undertake a genome-wide eQTL analysis for the expression of ACE2 in 190 virus-infected livers. Due to a dominant effect of IFNL4 locus on hepatic gene expression, we used both additive and dominant genetic models using linear regression and adjusted for population structure by including the first five host genetic principal components (PCs) as covariates. We also added age, sex and liver cirrhosis status as covariates to account for possible confounding. The outcome variable in the linear regression analysis was the expression of ACE2 RNA in log_{10}-transformed transcripts per million (log_{10}(TPM+1)). There was no inflation in the association test statistics (Supplementary Figure 1). We used a P-value threshold of 5x10^{-8} to declare a convincing finding. Across the human genome, in both dominant and additive analysis, the most associated signals were observed for three SNPs in the IFNL locus (Figure 1a, Supplementary Figures 2, 3, 4 and Supplementary Table 1). For all three SNPs the dominant model had lower P-values (rs12980275, \( P_{\text{dom}} = 9.9x10^{-11} \), rs8103142, \( P_{\text{dom}} = 7.8x10^{-10} \), \( P_{\text{add}} = 2.5x10^{-8} \); rs12979860, \( P_{\text{dom}} = 3.9x10^{-9} \), \( P_{\text{add}} = 6.5x10^{-8} \)). In European populations these SNPs are in high linkage disequilibrium (LD) with each other and with the likely causal variant rs368234815 SNP (not typed in
our genotyping array; Supplementary Figure 5; LD between the three associated SNPs and rs368234815 from 1000 genomes study, CEU population, \( r^2_{rs12980275} = 0.90, r^2_{rs8103142} = 1.0, r^2_{rs12979860} = 0.98 \).

To understand the impact of polymorphisms in the IFNL4 gene and other host factors on ACE2 expression in presence of viral infection, we focused on the impact of IFNL4 SNP rs12979860 on ACE2 RNA expression (Figure 1b). This SNP is an IFNL4 intronic SNP and closest to rs368234815 SNP among the three associated SNPs (Supplementary Figure 5), where rs12979860 C allele is in linkage disequilibrium with rs368234815 TT allele (\( r^2 = 0.98 \)). Using the dominant genetic model: C/C versus C/T and T/T genotypes i.e. those that do not produce IFN-\( \lambda \)4 protein and those that do produce IFN-\( \lambda \)4 protein, we

![Image](image_url)

**Figure 1. Impact of host genetics and other factors on ACE2 expression in HCV infected liver.** (a) Manhattan plot of association tests between host genetic variation and ACE2 expression (in log_{10}-transformed transcripts per million, log_{10}(TPM+1)) in virus infected liver biopsies using a dominant genetic model. The dashed line indicates 5x10^{-8} significance threshold. Significant SNPs are indicated and their IDs are shown. (b) Forest plot of the effect sizes of SNP rs12979860 (dominant model C/C versus C/T and T/T genotypes), cirrhosis status, age and sex on ACE2 expression. The black circles indicate the point estimate of the effect sizes and the black lines indicate their 95% confidence interval. (c) Distribution of ACE2 expression stratified by SNP rs12979860 genotypes (dominant model). Black circle shows the mean and the lines indicate its 95% confidence interval. (d) The relationship between ACE2 expression and age. The blue and red lines show the linear regression fit (for C/C and non-C/C genotypes respectively) and the grey area indicates their 95% confidence interval. (e) Distribution of ACE2 expression stratified by host liver cirrhosis status. The black dot and lines indicate the mean and its 95% confidence interval. ACE2 = Angiotensin I converting enzyme 2; TPM = transcripts per million; SNP = single nucleotide polymorphism;
observed a two-fold higher expression of ACE2 (P = 3.9x10^-9, mean expression for C/C = 4.34 TPM, mean expression for non-C/C = 2.03 TPM) in individuals with C/C versus non-C/C genotypes (Figure 1c).

Additionally, we obtained weak evidence that ACE2 expression increased with age (P = 0.04) in both C/C and non-C/C patients (Figure 1d) and patients with cirrhosis had a 1.5 fold higher ACE2 expression relative to non-cirrhotic patients (P = 0.02, mean expression for cirrhotic = 3.86 TPM and mean expression for non-cirrhotic = 2.52 TPM, Figure 1e).

To detect common regulatory mechanisms, biological function and the context of ACE2 expression, we performed correlation analysis accounting for multiple testing to identify genes correlated with ACE2 expression in virus-infected livers. We observed large correlation coefficients (maximum of 0.6 and minimum of -0.5) and detected 1530 genes correlated with ACE2 expression at 1% false discovery rate (FDR) and with correlation coefficients of > 0.3 or < -0.3. Considering separately the genes that were positively correlated (N=1362) and those that were negatively correlated (N=168) with ACE2 expression (Supplementary Tables 2 and 3), we performed a gene set enrichment analysis, observing that genes involved in type I IFN signalling pathways were enriched among genes negatively correlated with ACE2 expression (Figures 2a and 2b, Supplementary Table 4). We also observed that genes involved in extracellular structure organisation were enriched among

![Figure 2.](image-url)

**Figure 2. Negative correlation of transcript levels of ACE2 with interferon-stimulated genes in HCV infected liver biopsies.**

(a) Expression of four representative interferon-stimulated genes (ISG) and their observed negative correlation with ACE2 expression. Pearson's correlation coefficient is shown for each gene. (b) Gene Ontology (GO) gene set enrichment analysis among genes with significant negative correlation with ACE2 expression (at false discovery rate of 1% and correlation coefficient of < -0.3). Only the top ten enriched gene sets are shown (from a total of 22 gene sets), which are all ISGs related pathways. (c) Expression of the four representative ISGs and their observed negative correlation with ACE2 in an independent HCV infected liver biopsy data set. Pearson's correlation coefficient is shown for each gene. (d) ACE2 expression in paired liver biopsy samples from before and after pegylated IFN-α treatment taken at different time points post treatment. (e) Log fold change (log_{10} (post-treatment / pretreatment)) of ACE2 expression after pegylated IFN-α treatment. At all time points we observe a reduction in mean ACE2 expression with the biggest median drop at 16 hours. The reported P-values are from paired t-tests for each time point. Combining all time points, downregulation of ACE2 after IFN-α treatment is highly significant (P = 6.5x10^-5, paired t-test).
genes positively correlated with ACE2 expression (Supplementary Figure 6 and Supplementary Table S5). We used an independent data set of liver biopsies from 28 patients (6 non-HCV infected controls and 22 HCV-infected cases, GSE84346) and replicated concordant correlation signs for 162 of the 168 genes negatively correlated with ACE2 (Figure 2c and Supplementary Table S6). This represents a significant enrichment of concordant correlation signs relative to the null hypothesis of no association between correlation coefficients in the two datasets (P = 2.2 × 10^{-16}, binomial test). In this replication cohort 18 patients had two biopsies taken, one before and another after treatment with pegylated IFN-α at different time points (N_{4hrs} = 4, N_{16hrs} = 3, N_{48hrs} = 3, N_{96hrs} = 3, N_{144hrs} = 5). At all time points we observed a down-regulation of ACE2 expression (Figure 2d), with the biggest median drop at 16 hours post treatment. These reductions were nominally significant at 48 and 144 hours after IFN-α treatment (paired t-test, Figure 2e), but across all time points represent a highly significant down-regulation of ACE2 after IFN-α treatment (P = 6.5 × 10^{-5}, paired t-test).

To further explore the negative correlation of ISGs with ACE2 expression in a known site of SARS-CoV-2 replication, we explored the relationship between ACE2 and ISGs expression in the gastrointestinal (GI) tract in a gene expression study of terminal ileum biopsies in inflammatory bowel disease (IBD) in treatment-naive young donors (RISK cohort, GSE57945). In intestinal biopsies, there was a striking decrease of ACE2 expression with increasing severity of inflammation that was independent of the abundance of transcriptional markers of epithelial identity (Figure 3a and Supplementary Figure 7a).

![Figure 3. Conservation of negative correlation of ACE2 expression with ISGs across tissues, conditions and species.](image)

**Figure 3.** Conservation of negative correlation of ACE2 expression with ISGs across tissues, conditions and species. (a) ACE2 expression in terminal ileum biopsies of IBD patients grouped based on disease severity and histologic assessment of inflammation. The black dots and lines indicate the mean and its 95% confidence interval in each group (Ulcerative Colitis without ileal involvement; iCD = ileal Crohn’s Disease; Micro Infl. = Microscopic Inflammation, Macro Infl. = Macroscopic Inflammation; DU = Deep Ulcers). Data are shown as log_{10}-transformed RPKM relative to the epithelial cell identity metagene (see methods) (b) Expression of genes which are part of the Gene Ontology category “response to type I interferon” (GO ID: 0034340) and which in the liver are negatively correlated with ACE2 (FDR of 1% and correlation coefficient of < -0.3) in the IBD cohort stratified by disease severity. The black dots and lines show the mean and its 95% confidence interval for each group. (c) Negative correlation of representative ISGs with ACE2 in the terminal ileum biopsies of IBD patients. Pearson’s correlation coefficients are shown. (d) Volcano plot of lung tissue differential gene expression pattern induced by SARS-CoV-1 infection in mouse model vs. mock. ACE2 and representative ISGs which are part of the Gene Ontology category “response to type I interferon” (GO ID: 0034340) and which in the liver are negatively correlated with ACE2 are indicated by red dots.
and ISGs had increasing expression with rise in disease activity and were negatively correlated with ACE2 expression (Figures 3b, 3c and Supplementary Figure 7b11). Genes associated with epithelial cell structure and function were enriched among genes that were positively correlated with ACE2 in both liver and intestine, while genes associated with type I interferon signaling pathways were enriched among genes that negatively correlated with ACE2 expression in both tissues (Supplementary Figure 813). These data were supported by analysis of a second independent IBD cohort14 (GSE137344, Supplementary Figure 921).

Since the pattern of gene expression incorporating down-regulation of ACE2 in presence of ISGs was consistent in two models of viral chronic infection and/or inflammation in different tissue, we addressed whether a similar pattern of gene regulation was observed in lung tissue using data from mouse models of SARS-CoV-1 infection15 (GSE59185). Indeed, we observed in SARS-CoV-1 infected lung the same associated down-regulation of ACE2 in the presence of up-regulation of classical ISGs (Figure 3d).

Discussion
To understand the impact of host genetic factors on ACE2 expression in the presence of RNA virus infection, we performed a genome-wide eQTL analysis for ACE2 expression in 190 HCV-infected liver biopsies. Using infected tissue is important, since genetically driven differences in innate immune responses are only likely to be observed when innate immune responses are triggered. We observed that genome-wide host genetic polymorphisms in the IFNL region were significantly associated with ACE2 expression in the presence of viral infection. The likely causal mechanism is the variant rs368234815 [ΔG > TT], which results in a frameshift and abrogates production of IFN-λ416. In the context of HCV infection, production of IFN-λ is associated with high hepatic ISG expression (low ACE2 expression) and low viral load, but paradoxically with lower rates of spontaneous clearance in acute phase of infection and lower rates of response to treatment in the chronic phase of infection. Production of IFN-λ-4 is also associated with worse outcome of RNA virus respiratory tract infection in children16.

Interferon lambda receptor (IFNLRI) is largely restricted to tissues of epithelial origin16,22, therefore, IFN-λ proteins (type III IFN) may have evolved specifically to protect the epithelium. Overall, INF genes lead to a pattern of gene expression which is similar to type I IFN genes, but the time course and pattern of expression may vary10. This has been explored in HCV, where a slower, but sustained impact of IFNL signalling is seen26. In vitro studies have revealed that ISG expression and anti-viral activity induced by recombiant IFNλ4 are comparable to that induced by IFNL326, although the tight regulation of IFNL426 may impact on its ability to induce a rapid antiviral state26,26. However, once established, the IFNλ transcriptional module may also be highly sustained (as seen here and in other HCV cohorts15) and also noted elsewhere, e.g. after childbirth45.

In mice, the type III IFN response is restricted largely to mucosal epithelial tissues, with the lung epithelium responding to both type I and II IFNs17 and intestinal epithelial cells responding exclusively to type III IFNs. Among nonhematopoietic cells, epithelial cells are potent producers of type III IFNs. In mouse models, type III IFNs seem to be the primary type of IFN found in the bronchoalveolar lavage in response to influenza A virus infection and play a critical role in host defence35. The data from the GI tract indicate that this gene expression pattern is conserved amongst tissues, consistent with emerging data36.

To understand the context of ACE2 expression in vivo and detect common regulatory mechanisms in presence of RNA virus infection, we measured the correlation between ACE2 expression and approximately 15000 other genes expressed in HCV-infected liver biopsies. We observed a down-regulation of ACE2 in the presence of up-regulation of classical ISGs. This was replicated in a second cohort of HCV-infected liver biopsies. The same observation was confirmed in the gastrointestinal tract in an inflammatory condition and in lung in a murine model of SARS-CoV-1 infection. This indicates a robust maintenance of the transcriptional downregulation of ACE2 in presence of ISGs up-regulation across infections, inflammatory conditions, tissues and species. We also detected down-regulation of ACE2 in virus-infected liver biopsies after pegylated-IFN-α treatment with the biggest drop at 16 hours post treatment.

We have presented evidence that ACE2 may be negatively regulated by IFNs in vivo. We have also demonstrated that ACE2 expression in presence of RNA virus infection is modulated by genetic variation in the IFNL region. This regulation is likely due to confirmed differential activation of innate immune system in the liver in response to RNA virus infection associated with this region17,21. Therefore, given the prominent role of type III IFNs in defence of epithelial surfaces such as that in the lung from viral infections, we hypothesise that the genetic variation in the IFNL region may also play a role in modulating innate immune responses to SARS-CoV-2 infection. Genetic variation resulting in production of IFN-λ-4 is associated with high ISGs level and downregulation of ACE2 which may limit the ability of SARS-CoV-2 and other related coronaviruses to enter cells, but may, if sustained, also have impacts on inflammation and interfere with lung tissue repair37,38. Indeed ACE2- mice suffer from enhanced disease following virus infection of the lung through an angiotensin-driven mechanism39.

These data are derived from an in vivo assessment and the downregulation of ACE2 is consistent across conditions, tissues and species. The data are also potentially consistent with up-regulation of ACE2 seen in early time points by IFN-α in vitro40. However, we note that measuring ACE2 expression using cell cultures stimulated by type I and III IFNs we observed large amounts of variability between cell lines and also between donors when using primary bronchial epithelial cells (Supplementary Figures 10, 11 and 1213). The likely explanation for the difference is that the regulation of this physiologic receptor in an in vivo setting is distinct from
studies in vitro, but the full kinetics of this need further study during natural infection.

This study is relevant to the expression of ACE2 during SARS-CoV-2 infection. Although we did not study this directly in the respiratory tract, such studies should be performed to confirm these data. Furthermore the overall impact of IFNL4 polymorphism on the clinical course should be assessed, especially given the very variable distribution of IFNL4 alleles in different ethnic groups\(^{10,41}\). Finally, the genetic data add weight to the idea of a careful exploration of IFN-λ pathways in therapy for SARS-CoV-2\(^ {42}\).

**Methods**

**Boson patient cohort**

For this study, we used patient data from the BOSON cohort that has been described elsewhere in details\(^ {43}\). In summary, The BOSON study is a phase 3 randomized open-label trial to determine the efficacy and safety of treatment with sofosbuvir, with and without pegylated IFN-α, in treatment-experienced patients with cirrhosis and HCV genotype 2 infection and treatment-naive or treatment-experienced patients with HCV genotype 3 infection. All patients provided written informed consent before undertaking any study-related procedures. The BOSON study protocol was approved by each institution’s review board or ethics committee before study initiation. The study was conducted in accordance with the International Conference on Harmonisation Good Clinical Practice Guidelines and the Declaration of Helsinki (clinical trial registration number: NCT01962441).

RNA extraction, library prep, sequencing and mapping for the BOSON cohort

Liver biopsy samples were available for 198 patients. Total RNA was extracted from patient liver biopsies at baseline (pre-treatment) using RNeasy mini kits (Qiagen, 74104). Briefly, liver biopsy samples were mechanically disrupted in the presence of lysis buffer and homogenized using a QIAshredder (Qiagen, 79654). Tissue lysates were then centrifuged (8000 g for 1 minute) and clarified supernatants were transferred into new microcentrifuge tubes (pellets were discarded). Next, 350 μL volume of 70% ethanol was added to the lysates and samples were mixed by gentle vortexing. 700 μL of sample was then transferred into RNeasy spin columns (Qiagen, 74104; with 2 mL collection tubes) and centrifuged at 10000 rpm for 15 seconds. Column flow-through was discarded. DNase (Qiagen, 79254) digestion was subsequently performed to eliminate any contamination from genomic DNA. 80μL of DNase I solution (10μL DNase I stock + 70 μL Buffer RDD) was added directly to RNeasy spin columns and incubated at room temperature for 15 minutes. Following DNase incubation, the columns were washed with 350 μL of Buffer RW1 and centrifuged at 10000 rpm for 15 seconds. Flow-through was discarded and 500μL of Buffer RPE was added to the spin columns. Columns were then centrifuged again at 10000 rpm for 15 seconds and flow-through was discarded. An additional 500 μL of Buffer RPE was added to the spin columns and additional columns were centrifuged at 10000 rpm for 2 minutes. Finally, spin columns were transferred into new microcentrifuge tubes and 30 mL of RNase-free water was added directly to the column membrane. Columns were then centrifuged at 10000 rpm for 1 minute to elute the RNA.

RNA yield was quantified using a NanoDrop spectrophotometer. Selected samples were also run on an Agilent TapeStation system to assess RNA quality and purity. Library preparation from purified RNA samples was performed using the Smart-Seq2 protocol\(^ {44}\), used along with previously described indexing primers during amplification (see Additional file 5 of \(^ {45}\)).

High-throughput RNA sequencing of prepared libraries was performed on the Illumina HiSeq 4000 platform to 75bp PE at the Wellcome Center for Human Genetics (Oxford, UK). Reads were trimmed for Nextera, Smart-seq2 and Illumina adapter sequences using skewer-v0.1.125\(^ {46}\). Trimmed read pairs were mapped to human genome GRCh37 using HISAT2 version 2.0.0-beta\(^ {47}\). Uniquely mapped read pairs were counted using featureCounts\(^ {48}\), subread-1.5.0\(^ {49}\), using exons annotated in ENSEMBL annotations, release 75. Mapping QC metrics were obtained using picard-tools-1.92 CollectRnaSeqMetrics.jar. Three samples were excluded after QC checks due to low sequencing depth which left 195 samples for analysis. Genes were filtered using the criteria of having a count per million (CPM) of 1.25 in at least 10 samples to remove low expressed genes. Function cpm from edgeR\(^ {50}\) version 3.20.9 was used to calculate the CPM values. After removing low expressed genes we were left with 14882 genes. To normalise for library size and gene length, transcripts per million (TPM) values were calculated from unique mapped read counts and \(\log_{10}(TPM+1)\) was used in the analysis.

For the replication cohort\(^2\) (GSE84346), the read counts were downloaded from GEO. Gene expression data for 46 liver biopsy samples from 28 individuals were available for this data set. 22 individuals had chronic HCV infection while 6 individuals did not have HCV infection and were enrolled as controls. Among the 22 individuals with chronic HCV infection, 18 individuals were treated with pegylated interferon-alpha and a second biopsy was taken post-treatment. Genes were filtered using the criteria of having a count per million (CPM) of 1.25 in at least 10 pre-treatment samples to remove low expressed genes. After removing low expressed genes we were left with 14661 genes. To normalise for library size and gene length, transcripts per million (TPM) values were calculated from unique mapped read counts and \(\log_{10}(TPM+1)\) was used in the analysis.

**Host genotyping**

Host genome-wide genotyping was performed on 567 patients from the BOSON cohort as described previously\(^ {16}\). Briefly, genomic DNA was extracted from buffy coat using Maxwell RSC Buffy Coat DNA Kit (Promega, AS1540) as per the manufacturer’s protocol and quantified using Qubit (ThermoFisher). DNA samples from patients were genotyped using the Affymetrix UK Biobank array\(^ {49}\). After quality control and filtering of the human genotype data, approximately 330,000 common SNPs with minor allele frequency greater than 5%
were available for analysis. Both liver RNA transcriptomic and human genome-wide SNP data were obtained on a total of 190 patients of mainly White self-reported ancestry infected with HCV subtype 3a.

Statistical analysis

To test for association between autosomal human SNPs and ACE2 expression in $\log_{10}(\text{TPM}+1)$ unit, we performed linear regression using PLINK\textsuperscript{31} version 1.9 using additive and dominant genetic models adjusted for the human population structure by adding the first five genetic principal components as covariates. We also added host cirrhosis status, age and sex as covariates to the analysis. To assess the impact of age per 10 years increase, we divided the age by 10 before adding it as a covariate. For 190 patients both host genome-wide genotyping data and hepatic ACE2 expression data were available. We used a significance threshold of 5x10^{-4}. Investigating the impact of rs12979860 on ACE2 expression (in $\log_{10}(\text{TPM}+1)$ unit) we used linear regression with a dominant genetic model (C/C versus C/T and T/T genotypes) with the same exact covariates as in the GWAS model described above.

For the BOSON Cohort, $\log_{10}(\text{TPM}+1)$ values were calculated and used to estimate Pearson’s correlation coefficient between ACE2 and all other genes. The qvalue (version 2.10.0) package in R was used to calculate false discovery rate. We used FDR of 1% and correlation coefficient of >0.3 or < -0.3 to decide on genes significantly correlated with ACE2. To test for enrichment we used enrichGO function from the clusterProfiler (version 3.15.2) package\textsuperscript{32}, limiting the analysis to GO “biological process” class and maximum gene set size of 500.

An independent liver biopsy dataset\textsuperscript{4} (GSE84346) was used to replicate the negative correlation of ACE2 with 168 genes found in the BOSON cohort. We used the $\log_{10}(\text{TPM}+1)$ values from the 28 baseline liver biopsy samples and calculated Pearson’s correlation coefficient between ACE2 and the 168 genes. Assuming a null hypothesis of no association between correlation coefficients signs in the two data sets, one can use a binomial test to assess this null hypothesis where the probability of a negative correlation sign was estimated from the BOSON cohort by dividing the number of genes with negative correlation with ACE2 (3558) divided by total number of genes (14881) tested against ACE2. To test the hypothesis of down-regulation of hepatic ACE2 expression when treated with pegylated interferon-alpha, we used the 18 individuals with liver biopsy samples taken before and after the treatment. We used one-sided paired t-test to perform hypothesis testing for downregulation of ACE2 for each of the time points and across all time points.

RISK cohort

The RISK study is an observational prospective cohort study with the aim to identify risk factors that predict complicated course in pediatric patients with Crohn’s disease as previously described\textsuperscript{35}. The RISK study recruited treatment-naïve patients with a suspected diagnosis of Crohn’s disease. The Paris modification of the Montreal classification were used to classify patients according to disease behaviour (non-complicated B1 disease (non-stricturing, non-penetrating disease); complicated disease, composed of B2 (stricturing) and/or B3 (penetrating) behaviour) as well as disease location (L1, ileal only, L2, colonic only, L3, ileocolonic and L4, upper gastrointestinal tract). 322 samples were investigated with ileal RNA-seq. Individuals without ileal inflammation were classified as non-IBD controls. Patients with Crohn’s disease were followed over a period of 3 years. Patients were largely of European (85.7%) and African (4.1%) ancestry. RPKM expression values for the RISK cohort\textsuperscript{50} were retrieved from GEO (GSE57945). The dataset was filtered to (n=19,556) genes that had an expression value ≥ 0.1 in >10% of the patients.

Statistical analysis of RISK cohort

To account for the potential loss of epithelial cells contribution to gene expression a metagene score was generated based on the average expression of epithelial identity genes\textsuperscript{32}. RPKM data were transformed and presented as: RPKM+1 epithelial cell metagene. For the ISG metagene score, we used 17 genes that were significantly negatively correlated with ACE2 in the liver and were part of the GO term “response to type I interferon” (GO ID: 0034340). For the intersection of ACE2 correlated gene expression, genes were ranked based on their Pearson’s correlation coefficient with ACE2 for each patient subgroup. Intersection lists of ACE2 expression positively (Pearson’s correlation coefficient > 0.5) and negatively (Pearson’s correlation coefficient < -0.5) correlated genes were extracted (positive correlation: n = 2067; negative correlation: n = 2264). BOSON liver ACE2 expression and RISK ACE2 expression positively and negatively correlated gene sets were intersected based on Entrez gene identifiers using Cytoscape (version 3.7.1) and visualized using the Cytoscape Venn and Euler Diagrams (Version 1.0.3) plugin. Functionally grouped networks of terms and pathways were analysed using Cytoscape (version 3.7.1) ClueGO (version 2.5.6) and CluePedia (version 1.5.6) plug-in\textsuperscript{41}. The analysis was performed by accessing the Gene Ontology Annotation (GOA) Database for Biologic processes, Cellular components, Immune system processes and Molecular function, the Reactome pathways database and the KEGG database. Only pathways with an adjusted enrichment p-value ≤ 0.05 were considered (Two-sided hypergeometric test, Benferroni step down p-value correction). GO terms were grouped based on the highest significance when more than 50% of genes or terms were shared. The filtered RISK gene expression data (n=19,556; expression value ≥ 0.1 in >10% of the patients) served as reference gene set.

Resources for statistical analysis and data visualization:

- Prism version 8.0 (GraphPad Software)
- Excel for Mac Version 15.32 (Microsoft)
- Cytoscape 3.7.1 (https://cytoscape.org/)
- Cytoscape 3.7.1 plugin ClueGO (Version 2.5.6)
- Cytoscape 3.7.1 plugin CluePedia (version 1.5.6)
- Cytoscape 3.7.1 plugin Venn and Euler Diagrams (Version 1.0.3)
- Morpheus (https://software.broadinstitute.org/morpheus/)
- R (Version 3.6.1)
- RStudio (Version 1.2.5001)
GENESIS cohort

GENESIS is funded by the National Institute of Diabetes and Digestive and Kidney Diseases and managed by Emory University for the recruitment of self-identified African American subjects with IBD. We used a subset of 195 GENESIS cohort subjects with ileal transcriptomic profiles as an additional replication cohort to test for negative correlation of ACE2 expression with interferon-stimulated genes expression. Pearson correlation tests between normalized expression values for ACE2 and four ISGs confirmed that this pattern of negative correlation is also observable in a cohort enriched for African American ancestry. This dataset includes 158 IBD patients along with 37 controls. Subjects with ileal inflammation were included as IBD, while non-IBD controls did not have ileal inflammation. This dataset is enriched for African American ancestry (70%), and gender was equally distributed. Full descriptions of age, sex, race, disease status and other phenotypic information are available in a prior publication.

Additionally, ileal transcriptomic profiles sequenced on the NextSeq 550 platform are available in the GEO repository (GSE57945) for all subjects.

Data analysis for SARS-CoV-1 mouse model

The mouse lung tissue microarray data were downloaded from GEO using the accession number GSE59185. In case of multiple probes per gene, they were collapsed into a single feature, which resulted in 21217 features. Three lung tissue samples were from mice infected with wild type virus and three lung tissue samples were from mock infection. We then used LIMMA (version 3.34.9) to perform differential gene expression analysis between these two conditions. The log fold change and the p-values were used to make a volcano plot. The genes highlighted as red on the volcano plot are 17 genes which were significantly negatively correlated with ACE2 in the human liver and were part of the GO term “response to type I interferon” (GO ID: 0034340).

Cell line and human primary bronchial epithelial cell culture

Cell lines used in this study were hepatocyte cell lines Huh7, Huh7.5 (Gift from Charlie Rice), HepaRG, HepG2, human kidney cell line HEK 293T, Human lung fibroblast cell line MRC5, human colorectal tumor cells CaCO2 and Vero E6. All cell lines except the HepaRG cell line were cultured in DMEM media at 37°C, 5% CO2. HepaRG cell line was cultured at 37°C, 5% CO2 in Williams’ E medium (Gibco-Invitrogen) supplemented with glucose (2 mM), penicillin/streptomycin (50 U/mL), gentamycin (20 μg/mL), insulin bovine (5 μg/mL, Roche-Boehringer-Manheim, France), hydrocortisone hemisuccinate (7x10^{-3} M, Roche-Boehringer-Manheim) and FCS (10% selected, non-decomplemented, Fetaclone II-Hyclone-PERBIO France). The cells were allowed to differentiate in the presence of 2% (v/v) DMSO to the medium. Cells were stimulated with 100ng of IFNα2 (Cat # NBP2-34971-20ug), IFNL3 (Cat # 5259-IL-025) or IFNL4 (Cat # 9165-IF-025) all from Biotecnica-R&D systems, Abingdon, UK.

RNA isolation and quantification

RNA was extracted from cell lysates using the RNAlater Mini Kit (Qiagen, 74104). RNA concentration was determined with a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, MA, USA) at 260 nm.

cDNA synthesis and RT-qPCR analysis

cDNA was reverse-transcribed from template RNA either using a two-step reverse transcription using AppScript cDNA synthesis kit (Appleton Woods, ARP601). Samples were incubated at 42°C for 30 mins, followed by 85°C for 10 mins to inactivate reverse transcriptase. All RT-qPCR reactions were either performed using the Roche Light Cycler 480 instrument using AppProbe reagents (Appleton Woods, ARP305). Cycling conditions are as follows: 95°C 5 minutes; amplification 94°C 10 seconds, 58°C 30 seconds, 72°C 10 seconds, for 50 cycles followed by final cooling to 40°C for 10 seconds. Primers were designed using the Roche Universal Probe library system. Relative gene expression was calculated using the comparative cycle threshold method normalised to expression of the housekeeping gene GAPDH and expressed relative to a mock treated sample.

Data availability

Underlying data

The liver gene expression read counts are submitted to Gene Expression Omnibus (GEO) under accession number GSE149601. The raw FASTQ files are deposited in the European Genome-phenome Archive under the accession code EGAS00001004996. Human genotype data underlying this manuscript are deposited in the European Genome-phenome Archive under accession code EGAS00001002324.

Due to patient privacy concerns the gene expression FASTQ files and the human genotype data can only be accessed by making an application to the data access committee. The Information on access to the study data is available at [http://www.stop-hcv.ox.ac.uk/data-access](http://www.stop-hcv.ox.ac.uk/data-access).
Extended data
Figshare: Extended data for the paper “In vivo negative regulation of SARS-CoV-2 receptor, ACE2, by interferons and its genetic control”. https://doi.org/10.6084/m9.figshare.c.5284667.v1

This project contains the following extended data:
- Supplementary tables in CSV and xlsx formats
- Supplementary Figures

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

Acknowledgements
The authors would like to thank Gilead Sciences for the provision of samples and data from the BOSON clinical study for use in these analyses. We thank HCV Research UK (funded by the Medical Research Foundation) for their assistance in handling and coordinating the release of samples for these analyses, and the ISARIC-4C investigators for their discussion and contributions to this study. The authors thank Jan Rehwinkel for his contributions to the manuscript preparation and David Klenerman (University of Cambridge) for his contribution to the high throughput sequencing platform that underpins this work.

References


