A gene locus that controls expression of *ACE2* in virus infection

M. Azim Ansari[1,2,3,4]*, Emanuele Marchi[1,2,3]*, Narayan Ramamurthy[1,2,3]*, Dominik Aschenbrenner[1,3], Carl-Philipp Hackstein[2], STOP-HCV consortium, ISARIC-4C Investigators, Rory Bowden[4], Eshita Sharma[4], Vincent Pedergnana[5], Suresh Venkateswaran[6], Subra Kugathasan[6], Angela Mo[7], Greg Gibson[7], John McLauchlan[8], Eleanor Barnes[1,2,3], John Kenneth Baillie[9,10], Sarah Teichmann[11], Alex Mentzer[4], John Todd[4], Julian Knight[4], Holm Uhlig[1,3,12], Paul Klenerman[1,2,3]*

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

*These authors contributed equally

Correspondence to:

M. Azim Ansari & Paul Klenerman, Peter Medawar Building for Pathogen Research, University of Oxford, Oxford OX1 3SY

Tel: +441865281885
Fax: +441865281236
Email: ansari@well.ox.ac.uk or paul.klenerman@ndm.ox.ac.uk

Abstract

The SARS-CoV-2 pandemic has resulted in widespread morbidity and mortality globally, but with widely variable outcomes. The development of severe disease and mortality is higher in older individuals, males and those with other co-morbidities, and may vary across ethnic groups. However, so far, no host genetic factor has been clearly associated with susceptibility and development of severe disease. To understand the impact of host genetics on expression of *ACE2* (SARS-CoV-2 receptor) during RNA virus infection we performed a GWAS for *ACE2* expression in HCV-infected liver tissue from 195 individuals. We discovered that polymorphisms in the host *IFNL* region which control expression of *IFNL3* and *IFNL4* modulate *ACE2* expression. *ACE2* expression was regulated additionally by age, with a subsidiary effect of co-morbidity. The *IFNL* locus controlled expression of a gene network incorporating many well-known interferon-stimulated genes which anti-correlated with *ACE2* transcript levels. The same anti-correlation was found in the gastrointestinal tract, a site of SARS-CoV-2 replication where inflammation driven interferon-stimulated genes are negatively correlated with *ACE2* expression. The interferon dependent regulation of *ACE2* was identified in a murine model of SARS-CoV-1 suggesting conserved regulation of *ACE2* across species. Polymorphisms in the *IFNL* region, as well as age, may impact not only on classical antiviral responses but also on *ACE2* with potential consequences for clinical outcomes in distinct ethnic groups and with implications for therapeutic interventions.

Introduction

Severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) which results in coronavirus disease 2019 (COVID-19) is a positive-stranded RNA virus that causes a severe respiratory syndrome in a subset of infected individuals and has led to widespread global mortality. The genome of SARS-CoV-2 shares about 80% sequence similarity with SARS-CoV and 96% sequence similarity with bat coronavirus Bat CoV RaTG13[1].

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 SARS-CoV-1, infection by SARS-CoV-2 employs *ACE2* as a receptor for cellular entry[2]. Viral entry also depends on *TMPRSS2* protease activity, and cathepsin B/L activity may be able to substitute for *TMPRSS2*[2].

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[3–5]. Host genetic variation is important in determining susceptibility and disease outcome for many infectious diseases[6] and it is likely to be important in determining SARS-COV-2 susceptibility and development of severe disease after infection. Such insights would be important in understanding pathogenesis, repurposing antiviral drugs and vaccine development.

The earliest immune defence mechanism activated upon virus invasion is the innate immune system[7]. Virus-induced signalling through innate immune receptors prompts extensive changes in gene expression which are highly effective in resisting and controlling pathogens and subsequently prompt the activation of inflammatory and or antiviral immune effectors involved in pathogen clearance[8]. It has been shown that host genetics contributes to transcriptional heterogeneity in response to infections[9], which underlies some of the differences in innate immune responses observed between individuals and the varying susceptibility to infection[10]. Therefore in the context of infectious diseases, it is of paramount importance to investigate infected tissue to understand how host genetics may impact on gene expression.

To understand the impact of host genetics on expression of *ACE2* in the presence of viral infection, we first used HCV-infected liver biopsies from 195 individuals and performed host genome-wide genotyping and liver transcriptomics. Performing a genome-wide association study (GWAS) for *ACE2* expression, we found that *ACE2* expression is modulated by host genetic variation in the *IFNL* region on chromosome 19. We also observed that increase in age is significantly associated with increase in *ACE2* expression and that expression of interferon pathway signalling genes are negatively correlated with *ACE2* expression. This pattern is conserved across tissues, infections and species.

Results

We used genotyped autosomal SNPs in the host genome to undertake a GWAS, where the trait of interest was the expression of *ACE2* in virus-infected liver, performing more than 300,000 association tests. We performed linear regression assuming an additive model, and adjusting for population structure by including the first five host genetic principal components (PCs) as covariates. There was no inflation in the association test statistics (**Supplementary Figure 1**). We used a false discovery rate (FDR) of 5% to decide on significance.

Across the human genome, the most significant associations were observed for three SNPs (all in linkage disequilibrium[11] in European populations, **Supplementary Figure 2** and **3**) in the *IFNL* locus on chromosome 19 (**Figure 1a** and **Supplementary Table 1**, min p-value = 8.1×10^{-8} for SNP rs12980275).

Host genetic variations in this region have previously been associated with HCV infection outcomes such as viral load, spontaneous clearance and treatment outcome, as well as viral evolution[10,12,13]. The causal variant is likely to be the dinucleotide exonic variant rs368234815 in *IFNL4*[14]. This variant [ΔG > TT] causes a frameshift, abrogating production of functional IFN-λ4 protein. This variant is not directly typed on our genotyping array, however it is in high linkage disequilibrium with SNP rs12979860 [T > C] (r^2 = 0.975 CEU population, 1000 Genomes dataset, rs12979860 T allele is in linkage with rs368234815 ΔG allele) which is an intronic SNP in the *IFNL4* gene and was directly typed on our genotyping array and is one of the three SNPs significantly associated with *ACE2* expression. The allele frequency varies substantially between populations globally, with the rs12979860 CC genotype (which protects against HCV) most enriched in East Asian populations and the non-CC genotypes (CT and TT) strongly enriched amongst those of African origin[15] (**Supplementary Figure 4**).

To further understand the impact of polymorphisms in the host *IFNL4* gene on *ACE2* expression in presence of viral infection, we investigated the impact of *IFNL4* SNP rs12979860 on expression of *ACE2* in HCV-infected liver. We used linear regression using a dominant genetic model: CC vs CT and TT genotypes (i.e. those that do not produce IFN-λ4 protein and those that do) including age, gender and cirrhosis status as covariates (**Figure 1b**). We observed significantly higher expression of *ACE2* ($P=1.5 \times 10^{-9}$) in individuals with CC vs. non-CC genotypes (**Figure 1c**)**.** Additionally, we observed that *ACE2* expression increased with age (P=0.006) in both CC and non-CC patients (**Figure 1d**). Patients with cirrhosis tend to have higher *ACE2* expression although this association was marginally not significant (P=0.056, **Figure 1b**).

The *IFNL4* gene itself is polymorphic and a common amino acid substitution (coded by the SNP rs117648444 [G > A]) in the IFN-λ4 protein, which changes a proline residue at position 70 (P70) to a serine residue (S70), reduces its antiviral activity in vitro[16,17]. Patients harbouring the impaired IFN-

λ4-S70 variant display lower hepatic interferon-stimulated gene (ISG) expression levels, which is associated with increased HCV clearance following acute infection and a better response to IFN-based therapy, compared to patients carrying the more active IFN-λ4-P70 variant[16]. After imputing and phasing *IFNL4* rs368234815 and rs117648444 we observed three haplotypes: TT/G (IFN-λ4-Null); ΔG/G (IFN-λ4-P70) and ΔG/A (IFN-λ4-S70). HCV-infected patients were classified into three groups according to their predicted ability to produce IFN-λ4 protein: (i) no IFN-λ4 (two allelic copies of IFN-λ4-Null, N = 69), (ii) IFN-λ4–S70 (two copies of IFN-λ4-S70 or one copy of IFN-λ4-S70 and one copy of IFN-λ4-Null, N = 21), and (iii) IFN-λ4-P70 (at least one copy of IFN-λ4-P70, N = 92). Analysis of the IFN-λ4 predicted patient groups revealed that IFN-λ4–S70 group had the expected impact on ACE2 expression i.e. lower levels of *ACE2* expression relative to the IFN-λ4-Null group (P=0.022), and higher levels relative to the (more stimulatory) IFN-λ4-P70 group although the effect was marginally not significant, P=0.087, **Supplementary Figure 5**).

We also investigated the impact of the *IFNL4* SNP rs12979860 genotypes on *TMPRSS2*, *CTSB* and *CTSL* genes which may also be needed for viral entry[2]. We observed that these three genes had much higher expression levels in the liver relative to *ACE2*, however their expression was not significantly associated with SNP rs12979860 genotypes (**Supplementary Figure 6**). Both *CTSB* and *CTSL* genes had significantly lower expression in patients with cirrhosis ($P_{CTSB}=1.2x10^{-5}$, $P_{CTSL}=8.8x10^{-8}$), while *TMPRSS2* had higher level of expression in cirrhotic patients; however this was not significant (P=0.11, **Supplementary Figure 7**).

We also performed correlation analysis accounting for multiple testing to identify genes correlated with *ACE2* in virus-infected livers. We observed large correlation coefficients (maximum of 0.51) and detected 591 genes significantly correlated with *ACE2* expression at 1% FDR and with correlation coefficients of > 0.3 or < -0.3. Considering separately the genes that were positively correlated and those that were negatively correlated with *ACE2* expression (**Supplementary Tables 2** and **3**), we performed a gene set enrichment analysis, observing that genes involved in type I interferon signalling pathways were enriched among genes negatively correlated with *ACE2* expression (**Figure 1e** and **1f**). These genes overlap strongly with those induced by *IFNL*[7]. We also observed that genes involved in extracellular structure organisation were enriched among genes positively correlated with *ACE2* expression (**Supplementary Figure 8**).

To understand the impact of the *IFNL* locus on the overall gene expression in the virus-infected liver, we used our liver transcriptome data and tested for association between SNP rs12979860 and gene expression data using a dominant genetic model (CC vs. CT and TT genotypes). At 1% FDR, SNP rs12979860 was an eQTL for 583 genes. Genes involved in type I interferon signalling were enriched among genes that were upregulated in non-CC individuals relative to the CC individuals. Genes involved in B and T cell mediated immunity were enriched among genes upregulated in CC individuals relative to non-CC individuals (**Supplementary Figures 9** and **10**).

To further explore the anti-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, examining the levels of *IFN*-regulated genes in a gene expression study of terminal ileum biopsies in inflammatory bowel disease (IBD) in treatment-naive young donors (RISK cohort[18]). 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[19] (**Figure 2a, Supplementary Figure 11a**) and genes anti-correlated with *ACE2* had increasing expression with rise in disease activity (**Figures 2b** and **Supplementary Figure 11b**)**.** 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 signalling pathways were enriched among genes that negatively correlated with *ACE2* expression in both tissues (**Figure 2c** and **Supplementary Figure 11c**).

We repeated this analysis to define the impact of *IFNL4* polymorphisms on gene expression in a second cohort of IBD patients enriched for those of African-American ethnicity[20] (**Supplementary Figure 11d**). This analysis confirmed the clear anti-correlation of *ACE2* expression with ISGs. Consistent with the absence of viral infection, there was no association seen between *IFNL4* genotype and *ACE2* expression in this IBD disease cohort (P=0.4).

Since the pattern of gene expression incorporating downregulation of *ACE2* was consistent in two models of chronic infection and/or inflammation in different sites, we addressed whether a similar pattern of gene regulation was observed in lung tissue using data from mouse models of SARS infection[21] (GSE59185). Indeed we observed in SARS-CoV-1 infected lung a similar enrichment of *ACE2* regulating genes as observed in human liver. There was a strong correlation of gene regulation measured by GSEA analysis and furthermore we observed the same associated down-regulation of *ACE2* in the presence of up-regulation of classical ISGs (**Figure 2e** and **f**)**.**

Discussion

To understand the host genetic factors that drive *ACE2* expression in the presence of RNA virus infection, we performed a genome-wide association analysis, for *ACE2* expression in infected liver. 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 host genetic polymorphisms in the *IFNL* region modulate *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-λ4 protein. Although our initial observation was made in patients with HCV infection and the liver tissue, given the robust maintenance of the transcriptional pattern in the GI tract and

in murine models of SARS, it seems likely to be relevant to SARS-CoV-2 pathogenesis.

Interferon lambda receptor (*IFNLR1*) is largely restricted to tissues of epithelial origin[22,23], therefore, IFN-λ proteins may have evolved specifically to protect the epithelium. Overall, *INFL* genes lead to a pattern of gene expression which is similar to type I interferon genes, but the time course and pattern of expression may vary[7]. This has been explored in HCV, where a slower, but sustained impact of *IFNL* signalling is seen[24]. *In vitro* studies have revealed that ISG expression and anti-viral activity induced by recombinant *IFNL4* are comparable to that induced by *IFNL3*[25]. Importantly, however, the tight regulation of *IFNL4*[26] means its ability to respond and induce a rapid antiviral state may be limited[27] as seen both *in vitro*, and *in vivo*[28]. However, once established, the *IFNL4* transcriptional module may also be highly sustained (as seen here and in other HCV cohorts[16]) and also noted elsewhere, e.g. after childbirth[29].

In mice, the type III IFN response is restricted largely to mucosal epithelial tissues, with the lung epithelium responding to both type I and III IFNs[30] 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 defence[31]. Intriguingly, in humans, the *IFNL4* polymorphism identified here is associated with the outcome of RNA virus respiratory tract infections in children, with the non-CC variant showing a poorer outcome[32]. The data from the GI tract indicate that this gene expression pattern is conserved amongst tissues, consistent with emerging data[33]. Inflammatory signals may act to sustain the triggering of ISGs and sustain downregulation of *ACE2*. Of note for inflammatory bowel disease, loss of *ACE2* in the ileum impacts on secretion of antimicrobial peptides and the local microbiome[34].

Downregulation of *ACE2* itself may limit the ability of coronaviruses to enter cells, but may, if sustained, have impacts on inflammation. Indeed *ACE2*$^{-/-}$ mice suffer from enhanced disease following virus infection of the lung through an angiotensin-driven mechanism[35]. In other settings, the nonCC genotype may provide a more limited early response[26] and lead to more sustained activation of the *IFNL* pathway[29] than CC genotype **(Supplementary Figure 12** and **13)** although overall the issue of which genotype might be protective in COVID-19 remains open. We also note the impact of ageing, which blunts this response in both genotypes. The mechanism for this requires further study, as does the impact of gender. These data and model are also consistent with transient upregulation of *ACE2* seen in early time points by IFNα *in vitro*[36] but the full kinetics need further study *in vitro* and *in vivo*.

This study provides an orthogonal investigation of SARS-CoV-2 induction of *ACE2*. Although we did not study this directly in the respiratory tract, such studies should be urgently performed to confirm these data – ideally in

epithelial tissue, where the model suggested above can be tested. Furthermore the overall impact of this polymorphism on the clinical course should be assessed, especially given the very variable distribution of *IFNL4* alleles in different ethnic groups, which may in turn reflect selection earlier in human evolution[10,15]. Finally, the strong genetic data add weight to the idea of a careful exploration of *IFNL* pathways in therapy for SARS-CoV-2[37].

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. The authors would also like to thank HCV Research UK (funded by the Medical Research Foundation) for their assistance in handling and coordinating the release of samples for these analyses. This work was funded by a grant from the Medical Research Council (MR/K01532X/1 – STOP-HCV Consortium). The work was supported by Core funding to the Wellcome Centre for Human Genetics provided by the Wellcome Trust (203141/Z/16/Z). This work was also supported by the Medical Research Council [grant number MC_PC_19059] and a strategic award from the Wellcome Trust (211276/Z/18/Z – WSSS). PK, is supported as a Wellcome Trust Senior Investigator (WT 109965MA) and an NIHR Senior Investigator. PK is affiliated to the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emerging and Zoonotic Infections at University of Liverpool in partnership with Public Health England (PHE), in collaboration with Liverpool School of Tropical Medicine and the University of Oxford. EB was funded by the Medical Research Council UK, the Oxford NIHR Biomedical Research Centre and is an NIHR Senior Investigator. The work was also supported by the NIHR Biomedical Research Centre, Oxford. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England. 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.

Methods

Boson cohort:

Boson patient cohort:

For this study, we used patient data from the BOSON cohort that has been described elsewhere[38]. 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). Briefly, liver biopsy samples were mechanically disrupted in the presence of lysis buffer and homogenized using a QIAshredder. Tissue lysates were then centrifuged and clarified supernatants were transferred into new microcentrifuge tubes (pellets were discarded). Next, 1 volume of 70% ethanol was added to the lysates and samples were mixed by gentle vortexing . 700uL of sample was then transferred into RNeasy spin columns (with 2mL collection tubes) and centrifuged at 10000 rpm for 15 seconds. Column flow-through was discarded. DNase digestion was subsequently performed to eliminate any contamination from genomic DNA. 80uL of DNase I solution (10uL DNase I stock + 70uL 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 350uL of Buffer RW1 and centrifuged at 10000 rpm for 15 seconds. Flow-through was discarded and 500uL 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 500uL of Buffer RPE was added to the spin columns and columns were centrifuged at 10000 rpm for 2 minutes. Finally, spin columns were transferred into new microcentrifuge tubes and 30uL 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[39], used along with previously described indexing primers during amplification[40].

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[41]. Trimmed read pairs were mapped to human genome GRCh37 using HISAT2 version 2.0.0-beta[42]. Uniquely mapped read pairs were counted using featureCounts[43], subread-1.5.0[44], 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 of 1.25 in at least 10 samples to remove low expressed genes. Function cpm from edgeR[45] version 3.20.9 was used to calculate the counts per million (CPM) values. Two samples had zero CPM values for *ACE2* gene and they were set to the minimum *ACE2* expression (in CPM) observed across all the samples. Log10 of CPM was used for all the analysis.

Host genotyping and imputation

Genomic DNA was extracted from buffy coat using Maxwell RSC Buffy Coat DNA Kit (Promega) as per the manufacturer's protocol and quantified using Qubit (Thermofisher). DNA samples from patients were genotyped using the Affymetrix UK Biobank array, as described elsewhere[12]. Both liver RNA trascriptomic 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. 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. The Phasing and imputation was performed using SHAPEIT2[46] and IMPUTE2[47] version 2.3.1 using default settings and the 1000 Genomes Phase III dataset as a reference population[48]. After QC, the imputation data for 182 patients were of high quality. Imputation quality for samples included for rs117648444 and rs368234815 variants (information 0.974 and 0.994 respectively and certainty 0.995 and 0.997 respectively). All patients were also independently genotyped for SNP rs12979860 as described previously[38].

Statistical analysis

To test for association between human SNPs and *ACE2* expression (in log10(CPM)), we performed linear regression using PLINK[49] version 1.9 using an additive genetic model adjusted for the human population structure (first five PCs). For 190 patients both host genome-wide genotyping data and liver RNAseq data was available. We used the qvalue package in R to calculate false discovery rate in this analysis and 5% FDR as significance threshold.

To test for association between *ACE2* expression and host SNP rs12979860 (dominant genetic model (CC vs. CT and TT genotypes)), cirrhosis status, sex and age we used multivariate linear regression as implemented in R. To test for association between *ACE2* expression and IFN-λ4 predicted patient groups (IFN-λ4-Null, IFN-λ4-S70 and IFN-λ4-P70), we used linear regression as implemented in R and added cirrhosis status, sex and age as covariates to the analysis.

Log(CPM) data as calculated by cpm function from edgeR package was used to calculate Pearson's correlation coefficient against *ACE2* expression. The qvalue package was used to calculate false discovery rate. To filter out significant genes we used FDR of 1% and correlation coefficient of >0.3 or <-0.3 for positively and negatively correlated genes. To test for enrichment we used enrichGO function from the clusterProfiler package[50]. We only investigated gene sets in "biological process" GO hierarchy.

To test for association between SNP rs12979860 genotypes (CC vs. CT and TT) and expression of all genes, we used LIMMA package with voom transformation[51]. Cirrhosis status, sex, age, race and batch number were included as covariate to account for possible confounders. The gene set

enrichment analysis was performed using ernichGO function from the clusterProfiler package.

RISK cohort:

RNA isolation cDNA synthesis

Cell lysates were homogenized with a QIAshredder column (Qiagen, Crawley, UK) and RNA extracted with the RNEasy Mini Kit (Qiagen, Crawley, UK) following manufacturers instruction. cDNA was reverse-transcribed from template RNA either using a two-step reverse transcription using AppScript cDNA synthesis kit (Appleton Woods).

RISK cohort classification and data analysis.
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[52]. The RISK study recruited treatment-naive 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[52] 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.

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[19]. RPKM data were transformed and presented as: RPKM+1/epithelial cell metagene. For the intersection of *ACE2* correlated gene expression, genes were ranked based on their pearson correlation coefficient to *ACE2* for each patient subgroup. Intersected lists of *ACE2* expression positively (pearson correlation coefficient > 0.5) and negatively (pearson correlation coefficient < -0.5) correlated genes were extracted (positive correlation: n = 2067; negative correlation: n = 2264). Liver *ACE2* expression and RISK *ACE2* expression correlated and anti-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 (http://apps.cytoscape.org/apps/vennandeulerdiagrams). Functionally grouped networks of terms and pathways were analysed using the Cytoscape (version 3.7.1) ClueGO (version 2.5.6) and CluePedia (version 1.5.6) plug-in[53].
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 (https://reactome.org/) and the KEGG database (https://www.genome.jp/kegg/pathway.html). Only pathways with an adjusted enrichment p-value ≤ 0.05 were considered (Two-sided hypergeometric test, Bonferroni 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 Eluler Digrams (Version 1.0.3)
Morpheus (https://software.broadinstitute.org/morpheus/)
R (Version 3.6.1)
RStudio (Version 1.2.5001)

Specific statistical tests applied in this study are described in the respective figure legends. The level of statistically significant difference was defined as p ≤ 0.05.

GENESIS cohort:

Ileal biopsies from 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[54]. We used a subset of 195 GENESIS cohort subjects with ileal transcriptomic profiles as an additional replication cohort to test for anti-correlation of ACE2 expression with IFN gene expression. Pearson correlation tests between normalized expression values for ACE2 and four IFN genes confirmed that this pattern of anti-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, gender, race, disease status and other phenotypic information are available in a prior publication[20]. Additionally, ileal transcriptomic profiles sequenced on the NextSeq 550 platform are available in the GEO repository (GSE57945) for all subjects.

RNA pre-processing and data analysis for SARS mouse model

RNA-Seq read counts from each samples were aligned to human reference hg38 using STAR[55] and HISAT[56] alignment algorithms and gene read counts

were generated for each mapped sample using *featureCounts* program[43]. Low expressed features were filtered out prior logCPM transform of the read counts matrix, normalization and further gene filtering by expression variability (IQR > 0.75) and annotation were applied retaining 5817 genes to following analysis. *EdgeR*[57] and *limma*[51] packaged were used to perform differential gene expression analysis of in-home generated and public available data. Gene set enrichment analysis (GSEA[58]) was employed to measure co-expression of our lists of gene correlations and differentially expression analysis in public available data from GEO.

Figures

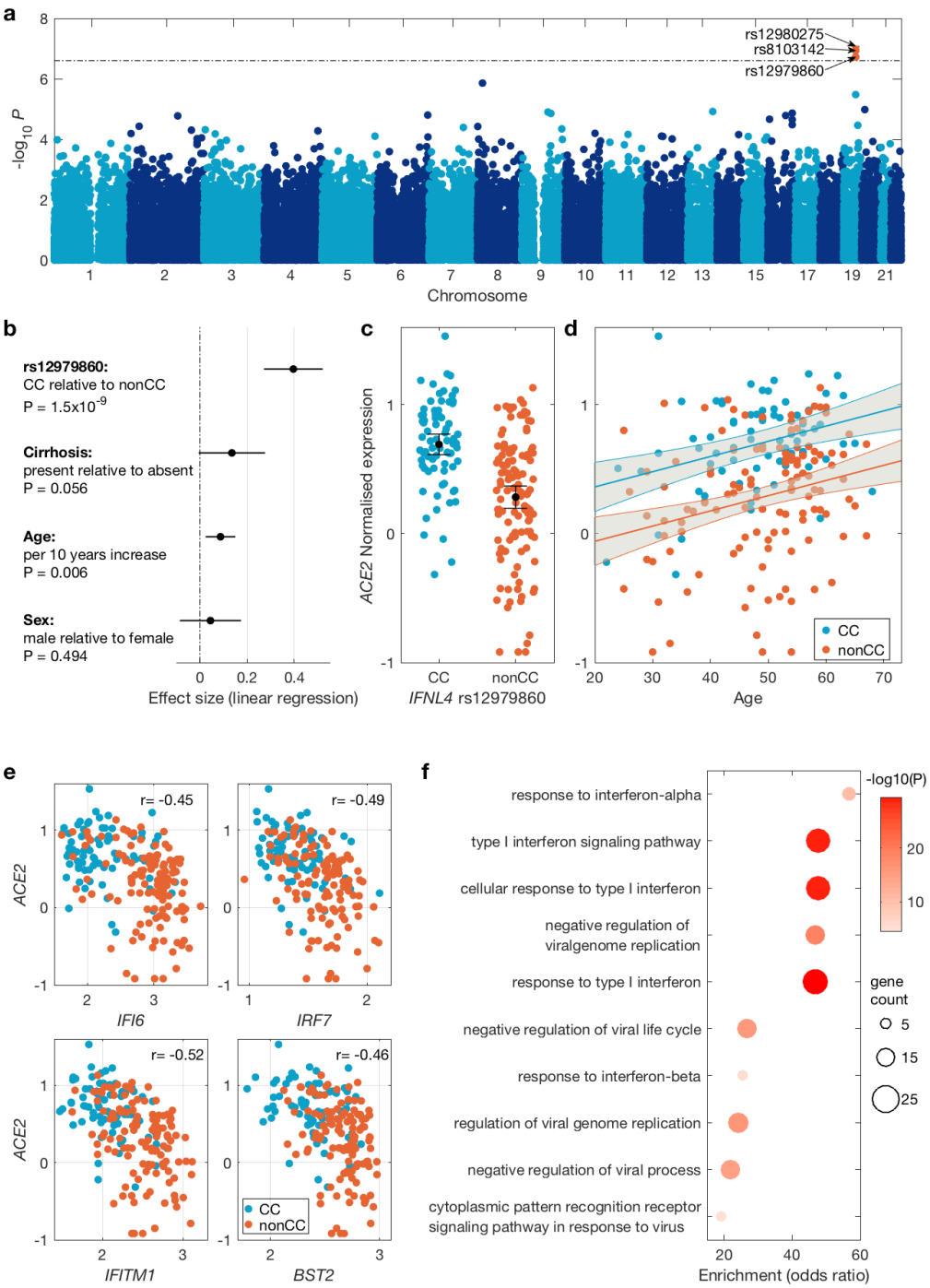

Figure 1: Impact of host genetics on *ACE2* expression in presence of virus. (**a**) Manhattan plot of association between host genetic variation and *ACE2* expression in virus infected liver biopsies using an additive model. The dashed line indicates 5% FDR level. Significant SNPs are coloured red and their ID is shown. (**b**) Forest plot of the effect sizes of SNP rs12979860 (dominant model), cirrhosis status, age and gender on *ACE2* expression. The black circles indicate the point estimate 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 CC and nonCC genotypes respectively) and the gray area indicates their 95% confidence interval. (**e**) Expression of four representative interferon stimulated genes (ISG) and their observed negative correlation with *ACE2* expression. The Pearson's correlation coefficient is shown for each gene. (**f**) Gene Ontology gene set enrichment analysis among genes with significant negative correlation with *ACE2* expression. Only the top ten enriched gene sets are shown, which are all ISG related pathways.

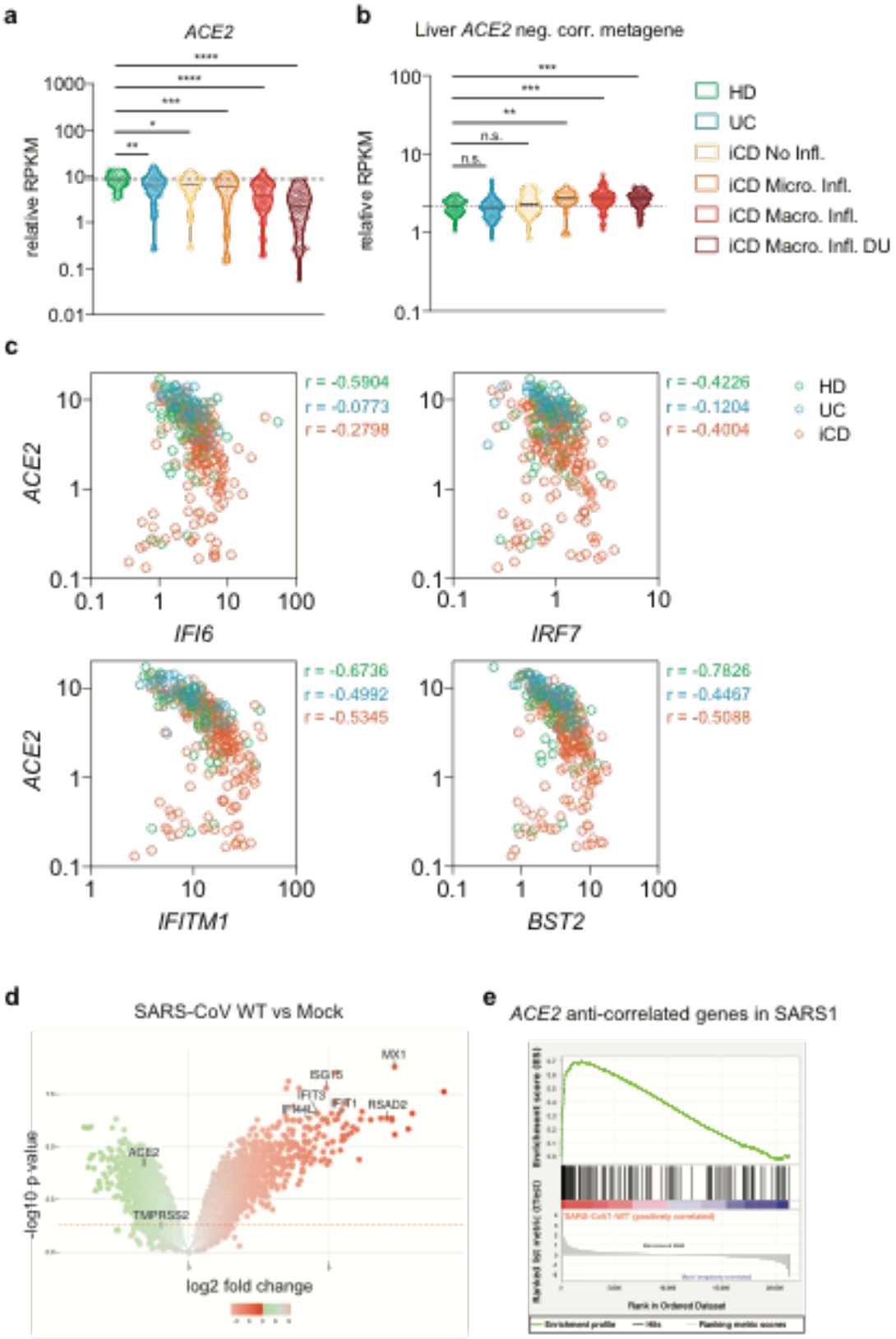

Figure 2: Conservation of interferon signalling-*ACE2* anti-correlation across tissues, chronic inflammation and species. (**a**) *ACE2* expression in terminal ileum biopsies transcriptomes of the RISK cohort grouped based on health state and histologic assessment of inflammation. (HD = Healthy Donor; UC = 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 RPKM relative to the epithelial cell identity metagene (see methods) (**b**) Expression of liver *ACE2* expression anti-correlated genes in the RISK cohort. Kruskal-Wallis test with multiple comparison correction controlling the FDR was used for association testing. (**c**) Expression and correlation of representative interferon stimulated genes (ISG) and *ACE2* in the RISK cohort. Pearson's correlation coefficients are shown. (**d**) Volcano plot of differential gene expression pattern induced by SARS-CoV1 infection in mouse model vs mock. Representative ISG genes and ACE2 are indicated. (**e**) GSEA plot of genes negatively correlated with ACE2 expression (enriched in ISGs) in the SARS-CoV-1 infected mouse model vs mock.

References:

1. Zhang, Y.-Z. & Holmes, E. C. A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. *Cell* **181**, 223–227 (2020).
2. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).
3. Team, T. N. C. P. E. R. E. The epidemiological characteristics of an outbreak of 2019 novel coronavirus disease (COVD-19). *China CDC Wkly.* **2**, 113–122 (2020).
4. Wenham, C., Smith, J. & Morgan, R. COVID-19: the gendered impacts of the outbreak. *The Lancet* **395**, 846–848 (2020).
5. Verity, R. *et al.* Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **0**, (2020).
6. Mozzi, A., Pontremoli, C. & Sironi, M. Genetic susceptibility to infectious diseases: Current status and future perspectives from genome-wide approaches. *Infect. Genet. Evol.* **66**, 286–307 (2018).
7. Mesev, E. V., LeDesma, R. A. & Ploss, A. Decoding type I and III interferon signalling during viral infection. *Nature Microbiology* **4**, 914–924 (2019).
8. Smale, S. T. Selective Transcription in Response to an Inflammatory Stimulus. *Cell* **140**, 833–844 (2010).
9. Nédélec, Y. *et al.* Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* **167**, 657-669.e21 (2016).
10. Ge, D. *et al.* Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399–401 (2009).
11. Fischer, J. *et al.* Combined effects of different interleukin-28B gene variants on the outcome of dual combination therapy in chronic hepatitis C virus type 1 infection. *Hepatology* **55**, 1700–1710 (2012).
12. Ansari, M. A. *et al.* Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. *Nat. Genet.* **49**, 666–673 (2017).
13. Ansari, M. A. *et al.* Interferon lambda 4 impacts the genetic diversity of hepatitis C virus. *Elife* **8**, (2019).
14. Prokunina-Olsson, L. *et al.* A variant upstream of IFNL3 (IL28B)

creating a new interferon gene IFNL4 is associated with impaired clearance of hepatitis C virus. *Nat. Genet.* **45**, 164–71 (2013).
15. Key, F. M. *et al.* Selection on a Variant Associated with Improved Viral Clearance Drives Local, Adaptive Pseudogenization of Interferon Lambda 4 (IFNL4). *PLoS Genet.* **10**, (2014).
16. Terczyńska-Dyla, E. *et al.* Reduced IFNλ4 activity is associated with improved HCV clearance and reduced expression of interferon-stimulated genes. *Nat. Commun.* **5**, 5699 (2014).
17. Bamford, C. G. G. *et al.* A polymorphic residue that attenuates the antiviral potential of interferon lambda 4 in hominid lineages. *PLOS Pathog.* **14**, e1007307 (2018).
18. Picoraro, J. A. *et al.* Pediatric Inflammatory Bowel Disease Clinical Innovations Meeting of the Crohn's & Colitis Foundation: Charting the Future of Pediatric IBD. *Inflamm. Bowel Dis.* **25**, 27–32 (2019).
19. Aran, D., Hu, Z. & Butte, A. J. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
20. Mo, A. *et al.* African Ancestry Proportion Influences Ileal Gene Expression in Inflammatory Bowel Disease. *Cell. Mol. Gastroenterol. Hepatol.* **0**, (2020).
21. Regla-Nava, J. A. *et al.* Severe Acute Respiratory Syndrome Coronaviruses with Mutations in the E Protein Are Attenuated and Promising Vaccine Candidates. *J. Virol.* **89**, 3870–3887 (2015).
22. Kotenko, S. V. *et al.* IFN-λs mediate antiviral protection through a distinct class II cytokine receptor complex. *Nature Immunology* **4**, 69–77 (2003).
23. Sheppard, P. *et al.* IL-28, IL-29 and their class II cytokine receptor IL-28R. *Nature Immunology* **4**, 63–68 (2003).
24. Marcello, T. *et al.* Interferons α and λ Inhibit Hepatitis C Virus Replication With Distinct Signal Transduction and Gene Regulation Kinetics. *Gastroenterology* **131**, 1887–1898 (2006).
25. Hamming, O. J. *et al.* Interferon lambda 4 signals via the IFNλ receptor to regulate antiviral activity against HCV and coronaviruses. *EMBO J.* **32**, 3055–3065 (2013).
26. Hong, M. A. *et al.* Interferon lambda 4 expression is suppressed by the host during viral infection. *J. Exp. Med.* **213**, 2539–2552 (2016).
27. Sheahan, T. *et al.* Interferon lambda alleles predict innate antiviral immune responses and hepatitis C virus permissiveness. *Cell Host Microbe* **15**, 190–202 (2014).
28. Ramamurthy, N. *et al.* Impact of Interferon Lambda 4 Genotype on Interferon-Stimulated Gene Expression During Direct-Acting Antiviral Therapy for Hepatitis C. *Hepatology* **68**, 859–871 (2018).
29. Price, A. A. *et al.* Prolonged activation of innate antiviral gene signature after childbirth is determined by IFNL3 genotype. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10678–10683 (2016).
30. Wack, A., Terczyńska-Dyla, E. & Hartmann, R. Guarding the frontiers: The biology of type III interferons. *Nature Immunology* **16**, 802–809 (2015).
31. Ye, L., Schnepf, D. & Staeheli, P. Interferon-λ orchestrates innate and adaptive mucosal immune responses. *Nature Reviews Immunology* **19**, 614–625 (2019).

32. Rugwizangoga, B. *et al.* IFNL4 Genotypes Predict Clearance of RNA Viruses in Rwandan Children With Upper Respiratory Tract Infections. *Front. Cell. Infect. Microbiol.* **9**, (2019).
33. Stanifer, M. L. *et al.* Critical role of type III interferon in controlling SARS-CoV-2 infection, replication and spread in primary human intestinal epithelial cells. *bioRxiv* 2020.04.24.059667 (2020). doi:10.1101/2020.04.24.059667
34. Perlot, T. & Penninger, J. M. ACE2 - From the renin-angiotensin system to gut microbiota and malnutrition. *Microbes and Infection* **15**, 866–873 (2013).
35. Yang, P. *et al.* Angiotensin-converting enzyme 2 (ACE2) mediates influenza H7N9 virus-induced acute lung injury. *Sci. Rep.* **4**, 7027 (2014).
36. Ziegler, C. *et al.* SARS-CoV-2 Receptor ACE2 is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Enriched in Specific Cell Subsets Across Tissues. *SSRN Electron. J.* (2020). doi:10.2139/ssrn.3555145
37. Prokunina-Olsson, L. *et al.* COVID-19 and emerging viral infections: The case for interferon lambda. *J. Exp. Med.* **217**, (2020).
38. Foster, G. R. *et al.* Efficacy of Sofosbuvir Plus Ribavirin with or Without Peginterferon-Alfa in Patients with Hepatitis C Virus Genotype 3 Infection and Treatment-Experienced Patients with Cirrhosis and Hepatitis C Virus Genotype 2 Infection. *Gastroenterology* **149**, 1462–1470 (2015).
39. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1100 (2013).
40. Lamble, S. *et al.* Improved workflows for high throughput library preparation using the transposome-based nextera system. *BMC Biotechnol.* **13**, 104 (2013).
41. Jiang, H., Lei, R., Ding, S. W. & Zhu, S. Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182 (2014).
42. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
43. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* (2014). doi:10.1093/bioinformatics/btt656
44. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
45. Smyth, G. K., Law, C. W., Alhamdoosh, M., Su, S. & Ritchie, M. E. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research* (2016). doi:10.12688/f1000research.9005.2
46. Delaneau, O., Howie, B., Cox, A. J., Zagury, J. F. & Marchini, J. Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* (2013). doi:10.1016/j.ajhg.2013.09.002
47. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* (2012). doi:10.1038/ng.2354

48. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
49. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
50. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* (2012). doi:10.1089/omi.2011.0118
51. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* (2014). doi:10.1186/gb-2014-15-2-r29
52. Kugathasan, S. *et al.* Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *Lancet* **389**, 1710–1718 (2017).
53. Bindea, G. *et al.* ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* (2009). doi:10.1093/bioinformatics/btp101
54. Brant, S. R. *et al.* Genome-Wide Association Study Identifies African-Specific Susceptibility Loci in African Americans With Inflammatory Bowel Disease. *Gastroenterology* **152**, 206-217.e2 (2017).
55. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (2013). doi:10.1093/bioinformatics/bts635
56. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
57. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (2009). doi:10.1093/bioinformatics/btp616
58. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).