

**Inhaled corticosteroids downregulate SARS-CoV-2-related gene expression in  
COPD: results from a RCT**

Stephen Milne\*, Xuan Li\*, Chen Xi Yang, Ana I Hernandez Cordero, Fernando Sergio Leitao Filho, Cheng Wei Tony Yang, Tawimas Shaipanich, Stephan F van Eeden, Janice M Leung, Stephen Lam, Don D Sin

\* authors made equal contributions to the work

---

**ONLINE DATA SUPPLEMENT**

**Detailed Methods**

**Supplementary Figures**

**References**

---

## **DETAILED METHODS**

### Design of the DISARM trial

To include: parallel allocation, location, co-primary outcomes (Change in total bacterial population in the bronchoalveolar lavage fluid at 12 weeks, Change in bacterial diversity using Shannon Index based on 16S sequencing at 12 weeks). Secondary outcomes are listed at [clinicaltrials.gov](https://clinicaltrials.gov). Binding and what was open. Dates.

The Differential Effects of Inhaled Symbicort and Advair on Lung Microbiota (DISARM) study aimed to determine the effects of ICS on the lung microbiome. The full trial protocol is registered at [clinicaltrials.gov](https://clinicaltrials.gov) (NCT02833480) and the study was approved by the Human Research Ethics Committee of the University of British Columbia and Providence Health Care (H14-02277). We recruited current and former smokers ( $\geq 10$  pack year history) aged 40-85 years with doctor-diagnosed COPD confirmed by spirometry (post-bronchodilator forced expiratory volume/forced vital capacity (FEV1/FVC)  $< 0.70$  and a post-bronchodilator FEV1  $> 20\%$  and  $\leq 80\%$  of predicted at the time of screening). Inclusion and exclusion criteria are summarized in Box E1, with details at <https://clinicaltrials.gov/ct2/show/NCT02833480>). The inclusion criteria were modified after trial commencement to include evidence of significant emphysema on thoracic computed tomography despite normal spirometry. Participants were recruited through St Paul's Hospital and the BC Cancer Agency, Vancouver, British Columbia, Canada, between October 2015 and June 2019.

**Box E1:** DISARM trial inclusion and exclusion criteria (see <https://clinicaltrials.gov/ct2/show/NCT02833480>).

<b>Inclusion criteria</b>
<ul style="list-style-type: none"> <li>• Provision of informed consent prior to any study specific procedures, AND</li> <li>• History of moderate to very severe COPD with a post-bronchodilator forced expiratory volume/forced vital capacity (FEV1/FVC) &lt;0.70 and a post-bronchodilator FEV1&gt;20% and ≤80% of predicted normal value at screening, OR</li> <li>• evidence of significant emphysema on thoracic computed tomography*, AND</li> <li>• Current smoker or ex-smoker with a tobacco history of ≥10 pack-years (1 pack year= 20 cigarettes smoked per day for 1 year)</li> </ul>
<b>Exclusion criteria</b>
<ul style="list-style-type: none"> <li>• Clinically important pulmonary disease other than COPD and/ or radiological findings suggestive of a respiratory disease other than COPD that is contributing to the subject's respiratory symptoms <ul style="list-style-type: none"> <li>○ Any disorder that is not stable in the opinion of the Investigator and could:</li> <li>○ Affect the safety of the subject throughout the study</li> <li>○ Influence the findings of the study or their interpretation</li> <li>○ Impede the subject's ability to complete the entire duration of study</li> </ul> </li> <li>• Unstable ischemic heart disease, or uncontrolled arrhythmia, cardiomyopathy, heart failure, and renal failure, or uncontrolled hypertension as defined by the Investigator, or any other relevant cardiovascular disorder as judged by the Investigator</li> <li>• Treatment with systemic corticosteroids and/or antibiotics, and/or hospitalization for a COPD exacerbation within 8 weeks prior to enrolment, based on last dose of steroids or last date of hospitalization whatever occurred later</li> <li>• Acute upper or lower respiratory infection requiring antibiotics or antiviral medication within 2 weeks prior to enrolment</li> <li>• Pneumonia within 8 weeks prior to enrolment</li> <li>• Pregnant, breastfeeding, or lactating women</li> <li>• Any clinically significant abnormal findings in physical examination, vital signs, haematology, clinical chemistry, or urinalysis during screening/run-in period, which, in the opinion of the Investigator, may put the subject at risk because of his/her participation in the study, or may influence the results of the study, or the subject's ability to complete entire duration of the study</li> <li>• Use of immunosuppressive medication within 28 days prior to randomization</li> <li>• Receipt of blood products within 30 days prior to enrollment</li> <li>• Receipt of any investigational non-biologic product within 30 days or 5 half-lives prior to Visit 1</li> <li>• History of alcohol or drug abuse within the past year, which may compromise the study data interpretation as judged by Investigator or Study Physician</li> <li>• Subjects who in the opinion of the investigator or qualified designee have evidence of active tuberculosis</li> <li>• Scheduled in-patient hospitalization or surgical procedure during the course of the study</li> <li>• Asthma as a primary or main diagnosis according to accepted guidelines. Subjects with a past medical history of asthma (e.g. childhood or adolescence) may be included</li> <li>• The male partner of someone who may become pregnant during the course of the study</li> </ul>

\*criterion added after trial commencement

After initial screening and providing informed written consent, enrolled participants were asked to replace their usual inhaled maintenance therapies (long-acting beta-agonists [LABAs] and/or long-acting muscarinic antagonists [LAMAs] and/or ICS) with formoterol (FOR) 12 microg twice daily via Turbuhaler for 4 weeks. Following this run-in period, participants proceeded to a bronchoscopy under conscious sedation. All bronchoscopies were conducted at a single centre (St Paul's Hospital) by experienced bronchoscopists who conducted the procedures according to a set protocol. They obtained cytological brush samples from 6-8<sup>th</sup> generation airways, preferentially from the right upper lobe, as well as bronchoalveolar lavage fluid samples, preferentially from the right middle lobe. Thoracic CT scans were used to avoid sampling from areas of pathology. Following this initial bronchoscopy, participants were block randomized (block size 6, stratified by history of ICS use prior to enrolment) in a 1:1:1 ratio to receive FOR monotherapy 12 µg twice daily via Turbuhaler, formoterol/budesonide (FOR/BUD) 12/400 µg twice daily via Turbuhaler, or salmeterol/fluticasone propionate (SAL/FLU) 25/250 µg twice daily via Diskus inhaler, for 12 weeks. At all times, participants were permitted to use inhaled short-acting beta-agonist as required. At the end of the 12 week treatment period, participants returned for a second bronchoscopy, with specimens collected from the same anatomical locations. Bronchoscopists and laboratory staff handling/processing the specimens were blinded to the participants' treatment allocation.

The co-primary outcomes for the DISARM study were: change in total bacterial population in the bronchoalveolar lavage fluid at 12 weeks, and change in bacterial diversity using Shannon Index based on 16S sequencing at 12 weeks. Secondary

outcomes are listed at <https://clinicaltrials.gov/ct2/show/NCT02833480>. The present study was an explorative analysis, and was not based on prespecified outcomes.

### Sample processing, RNA extraction and RNA sequencing

The details of our bronchoscopy specimen processing have been previously published (E1). RNA was extracted from cytological brush specimens stored in QIAzol RNA lysis buffer (QIAGEN, Stockach, Germany) using the RNeasy Plus kit (QIAGEN) according to the manufacturer's instructions, and specimens were submitted for sequencing at the University of British Columbia Biomedical Research Centre. Sample quality control was performed using the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). Qualifying samples were then prepared following the standard protocol for the NEBnext Ultra ii Stranded mRNA (New England Biolabs, Ipswich, MA, USA). Sequencing was performed on the Illumina NextSeq 500 (Illumina, San Diego, CA, USA) with paired end 42bp × 42bp reads. Samples were sequenced in batches of 24, with batches formed by random selection of samples.

### Bioinformatics and statistical comparisons

Quality control analysis of the raw sequencing reads was performed using FastQC (E2). Salmon (E3) was used for quasi-alignment and quantification of the reads to the GENCODE genome reference (assembly GRCh37, release 31) (E4). All subsequent analyses were performed in the R environment (version 3.6.0) (E5). Transcript level counts and transcript per million (TPM) were summarized to gene level using the "Tximport" package (E6), and gene level counts were filtered using the "filterByExpr"

function in the edgeR package (E7). The “voom” function in the Limma package (E8, 9) was used to normalize the count data to log<sub>2</sub> counts per million (CPM). Genes with low abundance (log<sub>2</sub> CPM <1 or TPM <2 in more than 80% of the samples) were filtered out, leaving a total of 15,263 genes. Principal component analysis on log<sub>2</sub>CPM showed no obvious treatment group clustering at baseline (Figure E1).

The effects of covariates on pre-treatment gene expression (log<sub>2</sub>CPM) were determined using linear regression (for age, BMI, and FEV1 percent predicted) and Wilcoxon rank-sum test (for sex, smoking status [current versus former], and hypertension). Changes in expression of individual genes were quantified as post-minus pre-treatment ( $\Delta\log_2\text{CPM}$ ). Pre- to post-treatment comparisons within each treatment group were performed using Wilcoxon signed rank test, and pairwise comparisons between treatment arms were performed using Wilcoxon rank-sum test.

We performed gene expression network analysis on the N=61 subjects with pre-treatment sequencing data available using the Weighted Gene Co-expression Network Analysis (WGCNA) package (E10). This analysis generates modules of highly-connected genes. Briefly, WGCNA generates a signed co-expression matrix based on the correlation between genes, which is then transformed into an adjacent matrix by raising the co-expression to a soft threshold power ( $\beta$ ). We used a  $\beta=6$  and a minimum module size of 50 genes. Since TPM were gene length normalised, they were first transformed to  $\log_2(\text{TPM}+1)$  and used as gene expression in this analysis. For each gene in the modules a ‘Module Membership’ (MM) was calculated by correlating the gene’s expression with the respective module’s expression (eigengene), i.e. the first principal component of each module gene expression

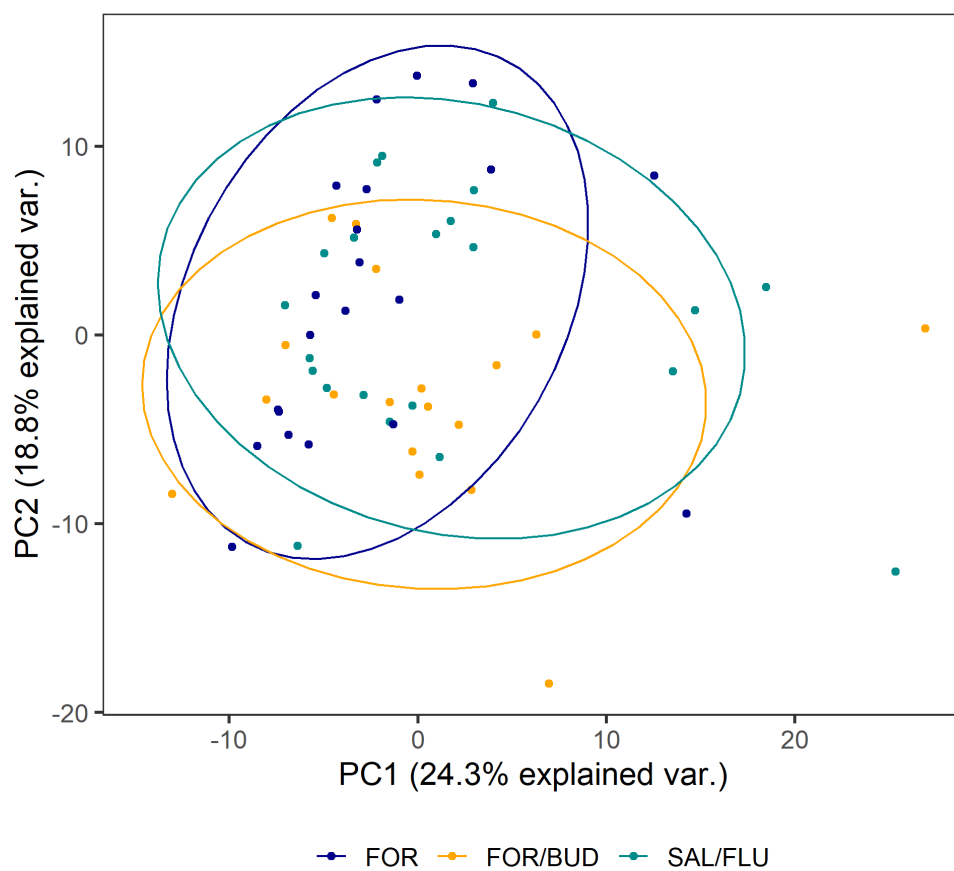
profile; the gene with the highest MM was termed the 'hub gene'. The modules of interest were those containing the SARS-CoV-2-related genes (ACE2, BSG, TMPRSS2, ADAM17 and FURIN).

Enrichment analyses for genes grouped by Gene Ontology (GO) (E11) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (E12) pathways (at FDR <0.05) were performed using the package WebGestaltR (E13).

Heat maps were generated using the "aheatmap" function in the NMF package (E14). For transcriptome-wide heat map (Figure 4), a false discovery rate of <0.05 was used to select genes based on pairwise Wilcoxon rank-sum test of pre- to post-treatment  $\Delta\log_2\text{CPM}$  between the treatment groups. For the innate immune response genes heat map (Figure 7A), pre- to post-treatment  $\Delta\log_2\text{CPM}$  was plotted for a panel of genes selected based on the literature.

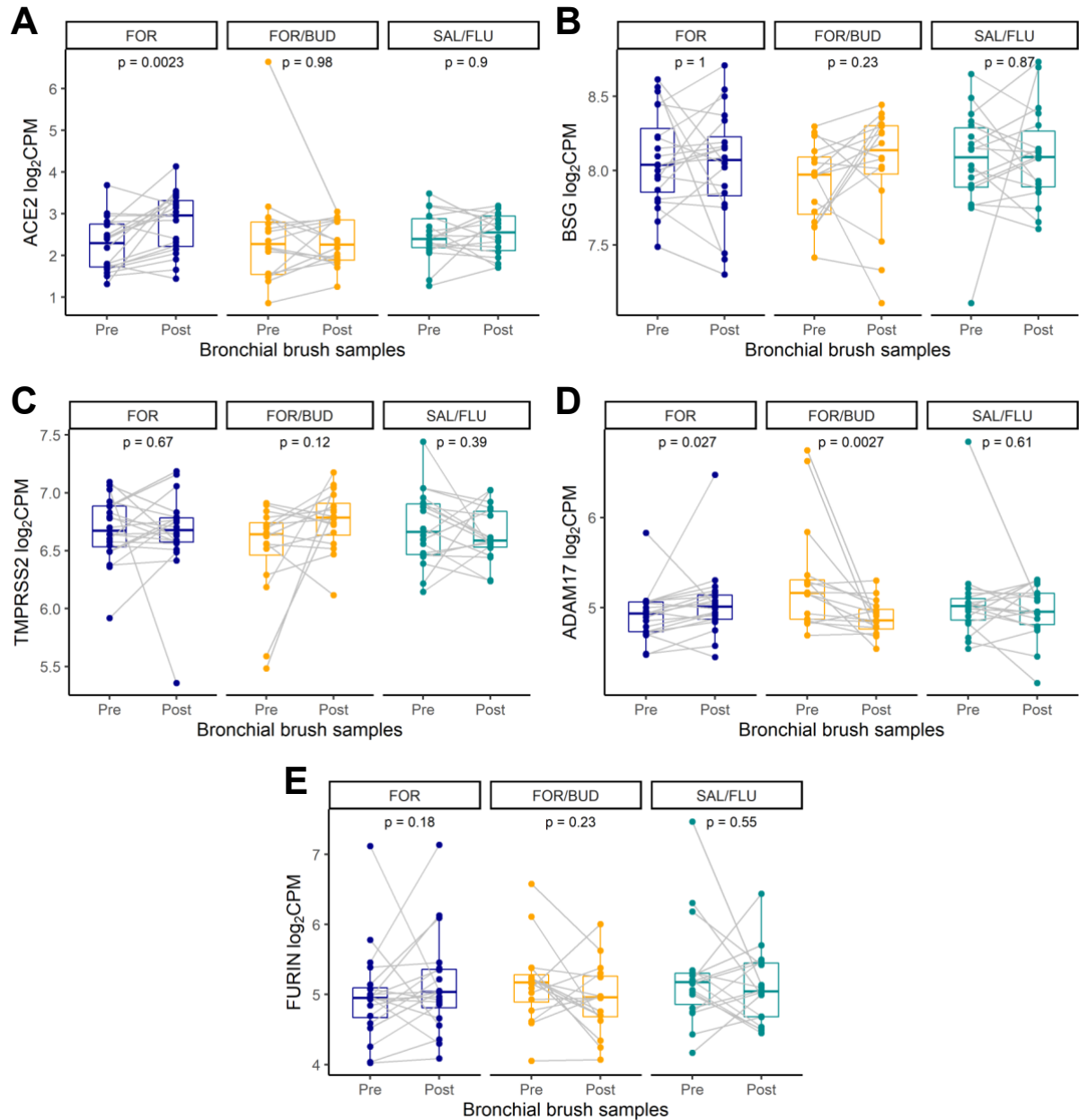
**SUPPLEMENTARY FIGURES:**

**Figure E1: Principal component analysis (PCA) of baseline whole-transcriptome gene expression data.** PCA plot grouped by treatment group (n=61 participants with pre-treatment sequencing data available) shows no obvious separation, suggesting that pre-treatment gene expression was similar between groups. Only genes passing quality control filtering were analyzed.

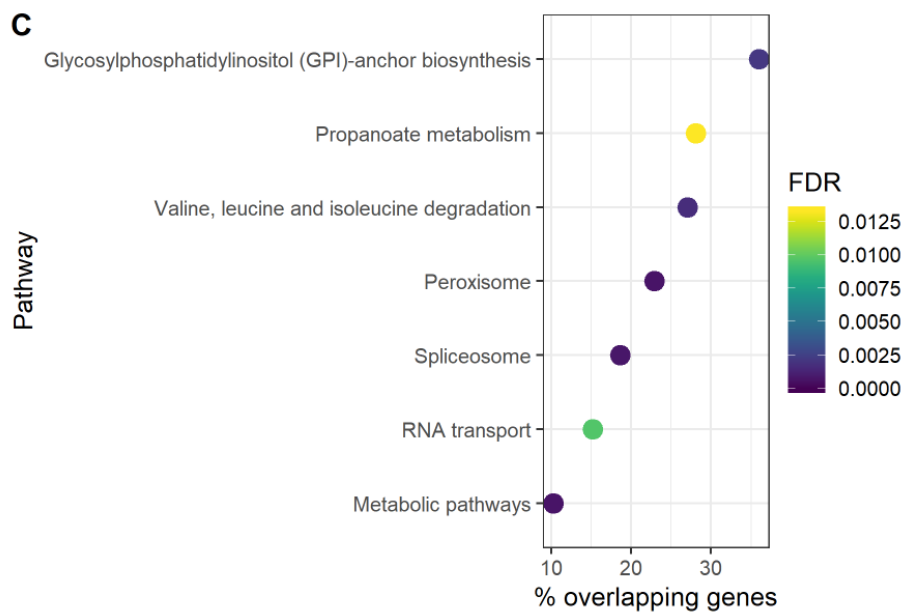
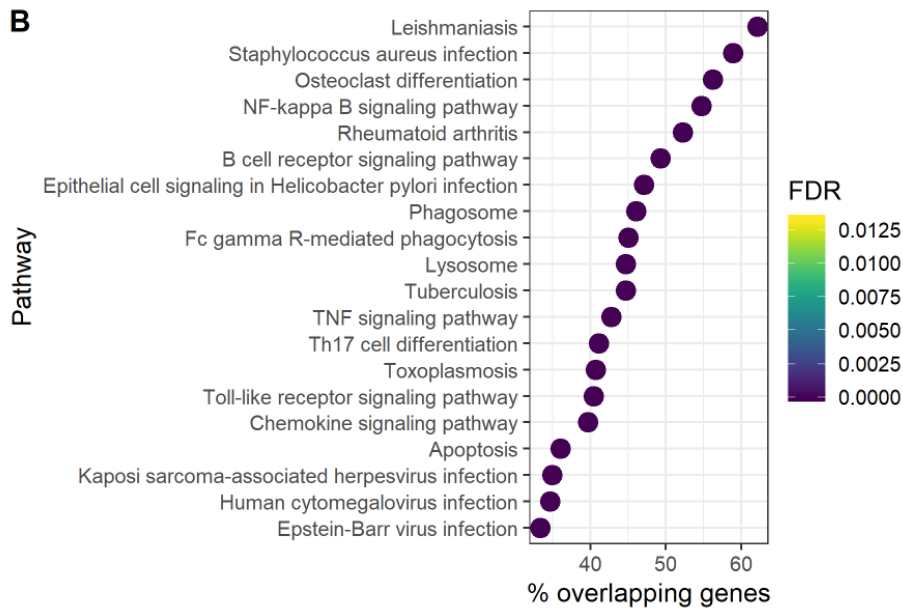
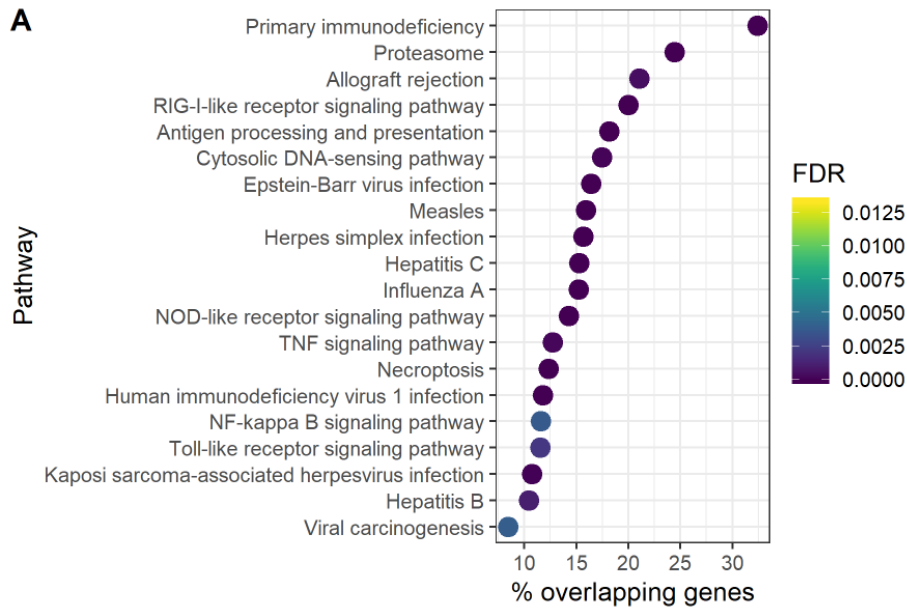




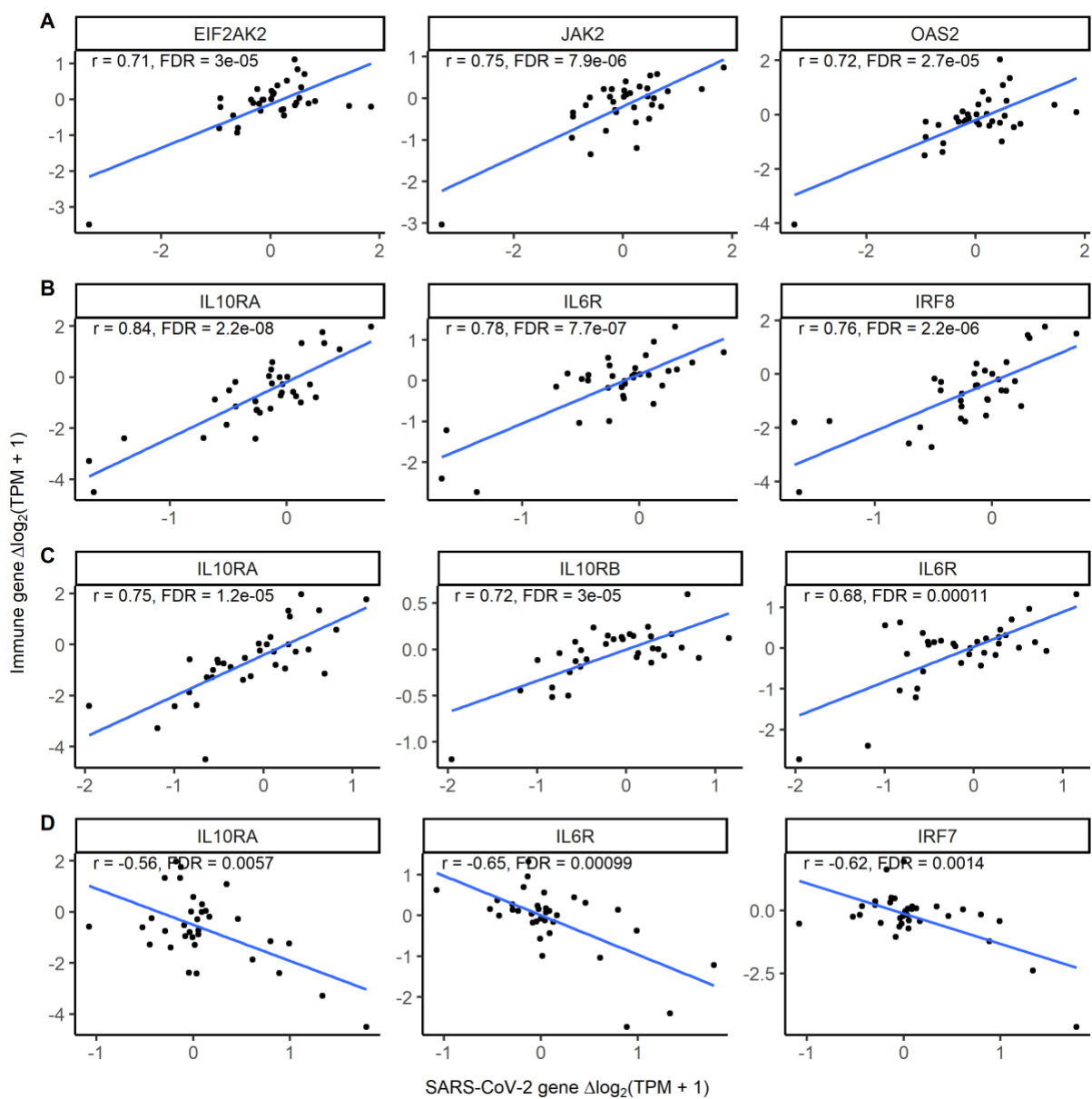
**Figure E2: Pre- to post-treatment change in expression ( $\Delta\log_2\text{CPM}$ ) of key SARS-CoV-2-related genes. (A) *ACE2*, (B) *BSG*, (C) *TMPRSS2*, (D) *ADAM17*, (E) *FURIN*. Boxplots show median and interquartile range. P value from Wilcoxon signed rank test. Abbreviations: CPM, counts per million.**



**Figure E3: Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment in SARS-CoV-2-related gene co-expression modules.** Top 20 KEGG pathways by FDR are shown, ordered by % overlapping genes (proportion of genes in the KEGG pathway that also appear in the gene module). The *ACE2* module (A) was enriched with genes related to viral diseases including influenza A and measles, as well as disorders of the immune system including primary immunodeficiency and allograft rejection. The *ADAM17/FURIN* module (B) was enriched with genes related to bacterial diseases including leishmaniasis and *Staphylococcus aureus* infection, as well as the autoimmune disease rheumatoid arthritis. The *BSG/TMPRSS2* module (C) was not enriched for any human disease pathways at  $FDR < 0.05$ . Abbreviations: FDR, false discovery rate.



**Figure E4: Scatter plots showing correlation of pre- to post-treatment change in selected immune response genes with change in key SARS-CoV-2-related genes for LABA/ICS treatment groups combined. Pre- to post-treatment change quantified as  $\Delta\log_2(\text{TPM}+1)$ .  $r$  from Pearson correlation. Top 3 correlated innate immune response genes for each SARS-CoV-2-related gene shown: (A) *ACE2*, (B) *ADAM17*, (C) *FURIN*, (D) *TMPRSS2*. There were no correlations with change in *BSG* at  $\text{FDR} < 0.05$ .**



## REFERENCES:

- E1. Xu S, Tsai A, Sze MA, Vucic EA, Shaipanich T, Harris M, Guillemi S, Yang J, Sinha S, Nislow C, Montaner J, Lam W, Lam S, Sin DD, Paul Man SF, Leung JM. Decreased microbiome diversity in the HIV small airway epithelium. *Respiratory Research* 2018; 19: 140.
- E2. Andrews S, Babraham Institute Bioinformatics. FastQC [computer software]. Available at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> [accessed 2020-04-01].
- E3. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 2017; 14: 417-419.
- E4. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, Barnes I, Berry A, Bignell A, Carbonell Sala S, Chrast J, Cunningham F, Di Domenico T, Donaldson S, Fiddes IT, García Girón C, Gonzalez JM, Grego T, Hardy M, Hourlier T, Hunt T, Izuogu OG, Lagarde J, Martin FJ, Martínez L, Mohanan S, Muir P, Navarro FCP, Parker A, Pei B, Pozo F, Ruffier M, Schmitt BM, Stapleton E, Suner MM, Sycheva I, Uszczyńska-Ratajczak B, Xu J, Yates A, Zerbino D, Zhang Y, Aken B, Choudhary JS, Gerstein M, Guigó R, Hubbard TJP, Kellis M, Paten B, Reymond A, Tress ML, Flicek P. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* 2019; 47: D766-d773.
- E5. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.
- E6. C S, MI L, MD R. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 1; peer review: 2 approved]. *F1000Research* 2015; 4: 1521.

- E7. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2009; 26: 139-140.
- E8. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 2015; 43: e47-e47.
- E9. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* 2014; 15: R29.
- E10. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 559.
- E11. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R, Gene Ontology C. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004; 32: D258-261.
- E12. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 1999; 27: 29-34.
- E13. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* 2019; 47: W199-W205.

E14. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* 2010; 11: 367.