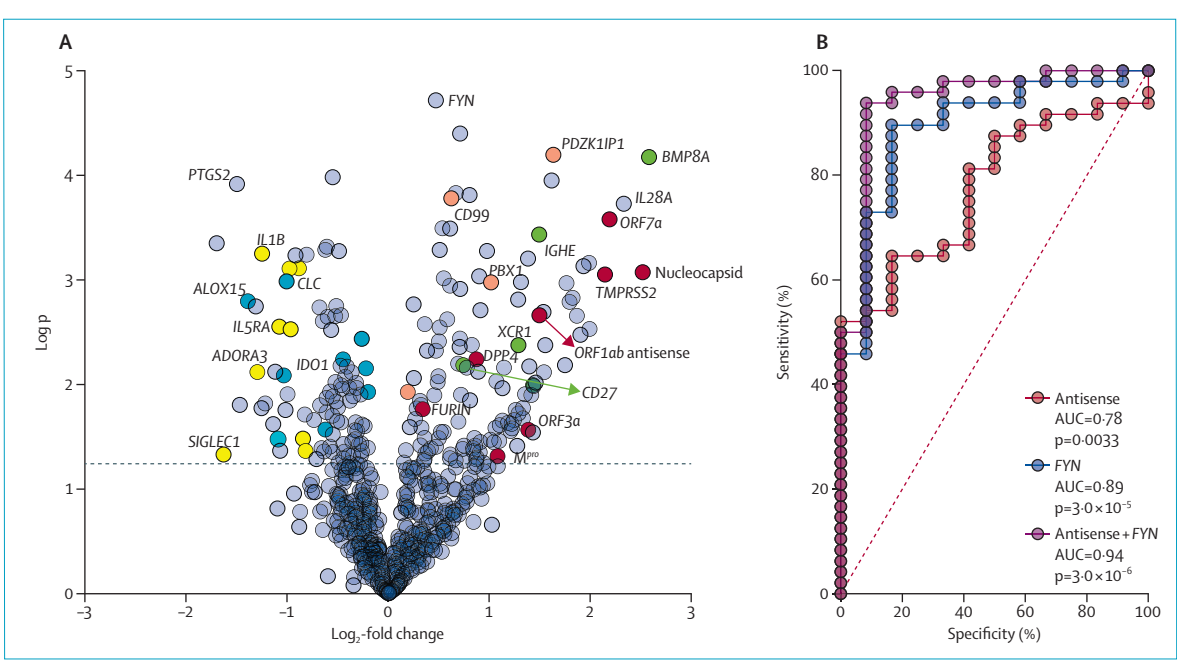
Blood transcriptomic analyses reveal persistent SARS-CoV-2 RNA and candidate biomarkers in post-COVID-19 condition

With an estimated 65 million individuals affected by post-COVID-19 condition (also known as long COVID),[1] non-invasive biomarkers are direly needed to guide clinical management. To address this pressing need, we used blood transcriptomics in a general practice-based case-control study. Individuals with long COVID were diagnosed according to WHO criteria, and validated clinical scales were used to quantify patient-reported outcomes.[2] Whole blood samples were collected from 48 individuals with long COVID and 12 control individuals matched for age, sex, time since acute COVID-19, severity, vaccination status, and comorbidities (appendix 1 p 2). Digital transcriptomic analysis was performed using the nCounter (Nanostring Technologies, Seattle, WA, USA) platform, as described for critical COVID-19.[3] Consequently, 212 genes were identified to be differentially expressed between individuals with long COVID and controls (figure A), of which 70 remained significant after adjustment for false discovery rate correction (appendix 1). Several viral RNAs were upregulated: nucleocapsid, *ORF7a, ORF3a, M*^{pro} (a nirmatrelvir plus ritonavir [Paxlovid] target), and antisense *ORF1ab* RNA. Specifically, the upregulation of antisense *ORF1ab* RNA suggests ongoing viral replication. SARS-CoV-2-related host RNAs (*ACE2/TMPRSS2* receptors, *DPP4/FURIN* proteases) and RNAs prototypical for memory B-cells and platelets[4] were also upregulated (figure A). Multivariable logistic regression identified antisense SARS-CoV-2 and *FYN* RNA concentrations as independent predictors of long COVID (corrected for age and sex; appendix 1 p 2). Receiver operating characteristic curve analysis showed significant discrimination (area under curve [AUC] 0·94, 95% CI 0·86–1·00) between individuals with long COVID (n=48) and controls (n=12), with 93·8% sensitivity and 91·7% specificity (figure B). Single biomarkers antisense SARS-CoV-2 (AUC 0·78, 0·65–0·90) and *FYN* RNA (AUC 0·89, 0·79–0·99) were significant predictors with lower sensitivity (52·1% and 72·9%, respectively) but similar specificity (91·7% and 100%, respectively; figure B). Upon summarising transcriptomic results into biological pathways, we found significantly decreased immunometabolism in individuals with long COVID, which was negatively correlated with the blood viral load (appendix 1 p 3). A qualitative analysis of individual SARS-CoV-2 transcript positivity revealed significant differences between individuals with long COVID and controls for antisense (65% *vs* 25%), *ORF7a* (60% *vs* 25%), and nucleocapsid (50% *vs* 8%) RNAs (figure C). Similarly, the SARS-CoV-2 transcript positivity with respect to the total blood viral load was also significantly different (60% *vs* 8%). By use of multivariable logistic regression, we found that age and sex were not associated with the distinction between a low and high viral RNA load status. Conversely, the number of comorbidities (odds ratio [OR] 1·61, 95% CI 1·14–2·49) and COVID vaccine doses (OR 0·36, 0·14–0·79) emerged

Lancet Microbe 2024
Published Online
https://doi.org/10.1016/S2666-5247(24)00055-7

See Online for appendix 1



(Figure continues on next page)

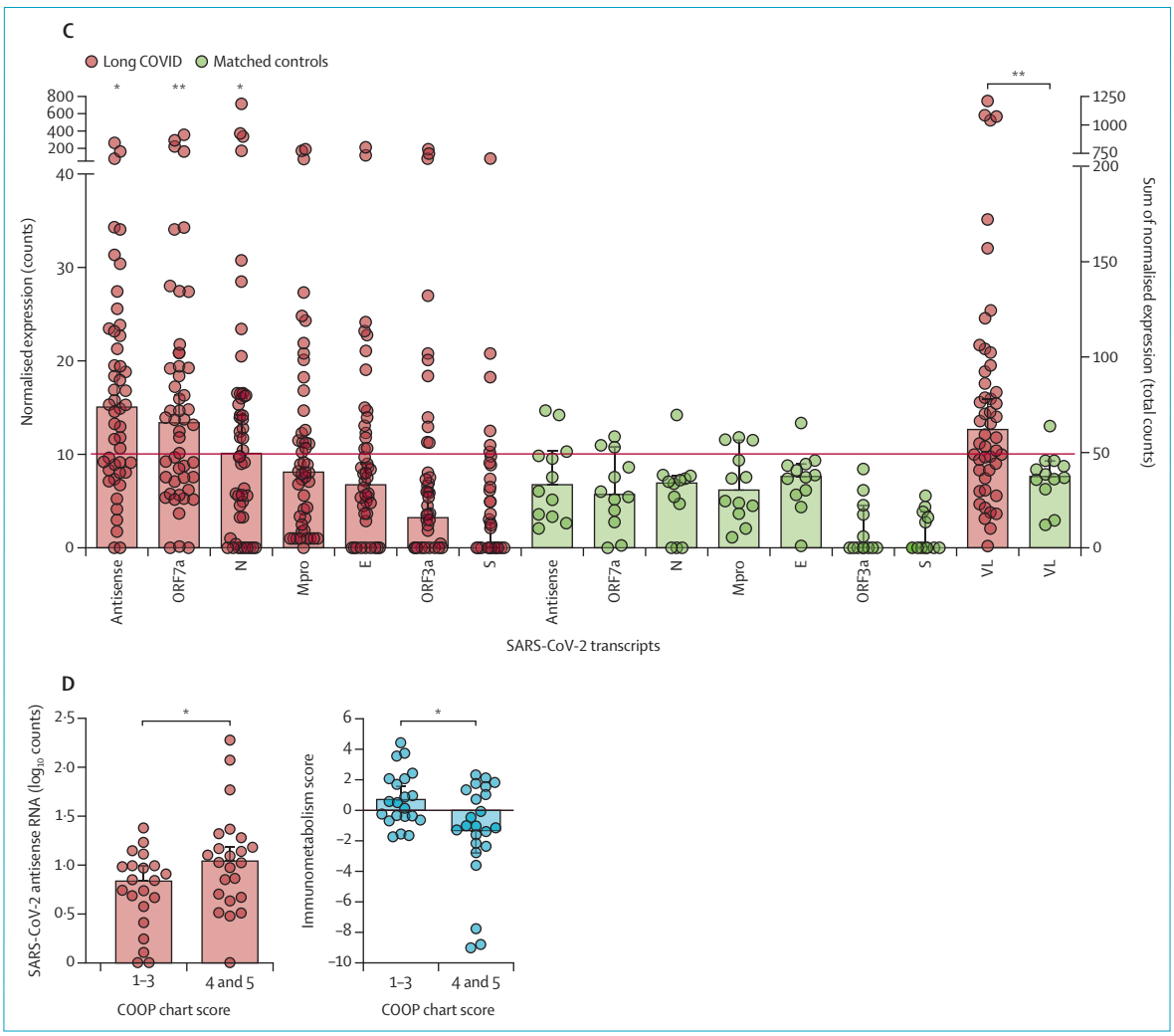


Figure: **Transcriptome analysis of immune response and SARS-CoV-2 transcripts and their association with patient-reported outcomes in long COVID**
(A) Volcano plot of differentially expressed genes in whole blood samples of individuals with long COVID (n=48) compared with those in control individuals (n=12) matched for age, sex, vaccine status, time since acute COVID-19, and number of comorbidities (appendix 1 p 2). Genes highlighted in red correspond to viral RNAs (nucleocapsid, *ORF3a*, *ORF7a*, M^{pro}, and antisense *ORF1ab*) and SARS-CoV-2-related host transcripts (*ACE2*/*TMPRSS2* (co)receptors, *DPP4* and *FURIN* proteases). Genes highlighted in green correspond to memory B-cell-expressed transcripts (*BMP8A*, *IGHE*, *CD27*, *XCR1*), and those highlighted in salmon correspond to platelet-expressed (*PDZK1IP1*, *PBX1*, *CD99*) transcripts. Genes highlighted in turquoise represent transcripts belonging to immunometabolism (eg, *PTGS2*, *ALOX15*, *IDO1*) and those in yellow represent lymphocyte activation (eg, *IL5RA*, *ADORA3A*, *SIGLEC1*, *IL1B*) biological pathways (appendix 1). (B) Receiver operator curve analysis shows significant discrimination (AUC 0·94; 95% CI 0·86–1·00, p=3·0 × 10^{-6}) between individuals with long COVID (n=48) and matched controls (n=12), as calculated by multivariable logistic regression with antisense SARS-CoV-2 and *FYN* transcript concentrations as independent predictors (corrected for age and sex; appendix 1). As single biomarkers, antisense SARS-CoV-2 (AUC 0·76, 0·86–1·00; p=3·0 × 10^{-6}) and *FYN* RNA (0·86–1·00; p=3·0 × 10^{-6}) were also significant predictors of long COVID disease status, albeit with lower sensitivity and specificity. (C) Overview of individual data for all SARS-CoV-2 transcripts (normalised expression in counts) and the blood viral load (sum of all SARS-CoV-2 normalised counts). Each circle represents a single individual with long COVID (red, n=48) or a matched control (green, n=12). The red horizontal line represents the cutoff for positivity (ten normalised counts for individual transcripts and 50 normalised counts for the viral load). Significant differences between individuals with long COVID and control individuals were determined using the Fisher's test (*p<0·05, **p<0·01). (D) Patient-reported outcome measures (COOP chart score in response to the question "During the last 2 weeks, how much have you been bothered by emotional problems such as feeling anxious, depressed, irritable or downhearted and sad?", on a visual scale ranging from 1 to 5, with 5 indicating the most severe condition; appendix 1 p 4). Compared with individuals with anxiety or depression categorised as mild (with scores 1–3, n=21), those with the condition categorised as severe (with scores of 4 and 5, n=23) were significantly associated with higher SARS-CoV-2 antisense RNA concentrations (*p<0·05, *t* test) and a lower immunometabolism score (*p<0·05, Mann-Whitney test). AUC=area under the curve.

as independent predictors of distinguishing between low and high viral RNA load status (appendix 2). We found that viral and immune parameters, such as the antisense *Orf1ab* RNA concentrations and immunometabolism score, were also linked to the patient-reported anxiety or depression score. Individuals classified as having severe anxiety or depression (with a score of 4 and 5) displayed significantly higher antisense RNA concentrations and lower immunometabolism scores ($p<0·05$) than those categorised as mild (with scores of 1–3; figure D). In conclusion, the associations among persistent viral RNA, immunometabolism, and patient-reported outcomes provide mechanistic insights for addressing the challenges posed by long COVID.

See **Online** for appendix 2

JVW received a speaker fee from Pfizer. The remaining authors declare no competing interests. This study was funded by Research Foundation Flanders (FWO) grants G0A0621N and G065421N (to JVW) and a grant from King Baudouin Foundation (2022-J51708200-F001) (to MJ).

*Soraya Maria Menezes, Marc Jamoulle, Maria Paula Carletto, Leen Moens, Isabelle Meyts, Piet Maes, *Johan Van Weyenbergh*
johan.vanweyenbergh@kuleuven.be

Laboratory of Clinical and Epidemiological Virology, Rega Institute for Medical Research, KU Leuven, 3000 Leuven, Belgium (SMM, MPC, PM, JVW); Immunogenetics Research Group, Inborn Errors of Immunity Unit; Department of Microbiology, Immunology & Transplantation, KU Leuven, Leuven, Belgium (LM, IM); HEC Information Sciences, University of Liège, Liège, Belgium (MJ)

1 Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. *Nat Rev Microbiol* 2023; **21:** 133–46.
2 Jamoulle M, Kazeneza-Mugisha G, Zayane A. Follow-up of a cohort of patients with post-acute COVID-19 syndrome in a Belgian family practice. *Viruses* 2022; **14:** 2000.
3 Menezes SM, Braz M, Llorens-Rico V, Wauters J, Van Weyenbergh J. Endogenous IFNβ expression predicts outcome in critical patients with COVID-19. *Lancet Microbe* 2021; **2:** e235–36.
4 Zhu A, Real F, Capron C, et al. Infection of lung megakaryocytes and platelets by SARS-CoV-2 anticipate fatal COVID-19. *Cell Mol Life Sci* 2022; **79:** 365.