BMC Medicine

RESEARCH

Open Access

VEGFA sex-specific signature is associated to long COVID symptom persistence

Xavier Farré[1,2], Natalia Blay[1,2], Susana Iraola-Guzmán[1,2], Francisco Fernández-Jiménez[1,2], Sayoa Alzate-Piñol[3], Laia Llucià-Carol[3], Ana Espinosa[4,5,6,7], Gemma Castaño-Vinyals[4,5,6], Carlota Dobaño[4,8], Gemma Moncunill[4,8], Marianna Karachaliou[4,9], Judith Garcia-Aymerich[4,5,6], Manolis Kogevinas[4,5,6,7], Carles Barceló[10,11], Israel Cadenas[3] and Rafael de Cid[1,2*]

Abstract

Background Long COVID involves persistent symptoms after COVID-19 recovery, affecting multiple organ systems for months or years. Risk factors include female sex, prior chronic conditions, severe SARS-CoV-2 infection, reinfections, and lack of vaccination. As a major public health concern, ongoing research continues to investigate its causes, mechanisms, and long-term effects.

Methods Proteomic expression analysis of 171 individuals, in two time points, with confirmed SARS-CoV-2 infection, including 133 long COVID patients from the deeply characterized COVICAT cohort, assessed 1395 protein biomarkers using Olink® technology. Statistical analyses with linear mixed models examined protein expression changes, long COVID status, and sex-specific differences. Functional analysis included gene set enrichment analysis and protein–protein interaction networks.

Results Findings revealed VEGFA overexpression in long COVID patients (effect size 0.322, SE = 0.098, $p = 0.0013$), along with sex-specific expression patterns and the influence of sex-hormonal status in females, with significant overexpression of circulating VEGFA levels specifically in postmenopausal women (Mann–Whitney *U* test *p* value = 8.55×10^{-3}). Network analysis identified 109 nodes and 274 edges, with VEGFA ranking highest in centrality. Dysregulated chemokine signaling, complement activation, and viral reactivation were also confirmed, consistent with prior studies.

Conclusions Using high-throughput proteomic profiling in a population-based cohort, we observed that vascular dysfunction, particularly involving VEGFA, is a key feature of long COVID, especially in milder cases, with significant overexpression of VEGFA in postmenopausal women. Sex-specific proteomic patterns suggest distinct recovery mechanisms, highlighting the need to consider sex, vascular health, and disease severity in the pathogenesis and management of long COVID.

Keywords COVID-19, Molecular epidemiology, Long COVID, Proteome, Sex differences

*Correspondence:
Rafael de Cid
rdecid@igtp.cat
Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Background

Long COVID, also known as post-acute sequelae of SARS-CoV-2 infection, refers to the persistence of symptoms long after the acute phase of COVID-19 [1]. It has become a major public health concern due to its widespread effects and significant impact on individuals' quality of life. Common symptoms of long COVID include fatigue or malaise, joint and muscle pain, persistent cough, sputum production, shortness of breath, chest discomfort, hair loss, memory problems, difficulty concentrating, headaches, depression, loss of smell, altered taste, heart palpitations, diarrhea, abdominal discomfort, sleep disturbances, and muscle weakness. Affecting multiple organ systems [2], long COVID persisted for up to 2 years and risk is influenced by multiple factors. The most common symptoms of long COVID are neurological symptoms, muscular problems, respiratory symptoms, and psychological and psychiatric symptoms. Sex dimorphism is already present here, being neurological and muscular symptoms more common in women, while respiratory symptoms were more common in men [3]. Research evidence links middle age, female sex, chronic conditions [3–5], severe initial COVID-19, multiple infections, and lack of vaccination to a higher risk of long COVID [3, 6, 7]. In addition to clinical and demographic factors, environmental and socioeconomic determinants such as low socioeconomic status [3] and exposure to air pollution [8] have also been associated with an increased risk of long COVID.

While individuals with mild or asymptomatic infections can still develop long COVID, those who experience severe acute infections, including hospitalization or severe respiratory distress, are at a higher risk of prolonged symptoms [9]. However, Mendelian randomization analysis has suggested specific host factors that contribute to susceptibility to long COVID independent from severity risk [10]. One particularly striking observation is the paradoxical sex disparity in COVID-19 severity and long COVID incidence. While women typically experience milder acute COVID-19, they are more prone to developing prolonged and severe long COVID symptoms [11], a trend also observed by our group in the COVICAT study [3]. In contrast, men tend to recover more quickly, exhibiting fewer persistent symptoms over time. While much of the research has focused on the immediate consequences of COVID-19, the mechanisms behind long COVID remain an ongoing area of study.

At the molecular level, various plausible pathophysiological mechanisms have been proposed [12, 13], but the complexity of long COVID, with symptoms spanning multiple organ systems and overlapping pathologies [14], poses significant challenges for developing robust diagnostic and therapeutic tools. Complement activation has been suggested as a significant mechanism in long COVID pathogenesis [15]. One of the most well-established hypotheses suggests that dysregulation of the complement system, an essential component of the immune response, plays a significant role in contributing to the pathogenesis of long COVID by driving inflammation and tissue damage. Dysregulation of the terminal complement system, including activation of the alternative and classical complement pathways, may contribute to vascular and endothelial dysfunction often observed in long COVID patients [15].

An alternative hypothesis gaining support is that endothelial dysfunction and altered angiogenesis contribute to the pathogenesis of long COVID. A 2024 study by Philippe et al. found elevated VEGF-A levels in patients with respiratory symptoms, correlating with reduced lung diffusion capacity and suggesting a vascular component to persistent pulmonary impairment [16]. Other angiogenic factors, such as ANG-1 and P-SEL, demonstrate high diagnostic accuracy for long COVID, reinforcing vascular dysfunction as a key mechanism [17, 18]. VEGF-A, which promotes vascular permeability and neuroinflammation, plays a significant role in the acute phase of COVID-19 and may also contribute to prolonged symptoms [19, 20]. Both infection and vaccination appear to interact with VEGF-A signaling, potentially influencing distinct stages of disease progression, although vaccine-related mechanisms remain to be fully elucidated [21].

Using a well-characterized general population cohort with detailed follow-up, this study applies blood proteomics to uncover molecular pathways linked to long COVID. By analyzing differential protein expression and sex-specific variations across 1395 proteins related to immune activation, inflammation, and vascular repair, we aim to uncover pathways linked to symptom persistence and their mechanistic roles. Ultimately, our findings could help identify potential targets for effective treatment strategies.

Methods

Design and study population

This study utilizes data from an adult COVID-19 population-based cohort in Catalonia (COVICAT) [3, 22], established in 2020 to assess the health impact of the COVID-19 pandemic on the population of Catalonia. COVICAT is a nested study survey in the GCAT cohort, a population-based cohort [23], which includes residents aged 40–65 years in Catalonia who had health data available prior to the pandemic. Participants from the GCAT cohort were invited to participate in COVID-19-related surveys and blood sampling during three distinct periods: June–December 2020 (baseline, $n = 9548$, with a sample

size of $n=3900$), June 2021–February 2022 (first follow-up, $n=7822$, with a sample size of $n=1086$), and February–September 2023 (second follow-up, $n=5215$, with a sample size of $n=1395$). Participants' data is linked to electronic heath records of the Catalan universal coverage health system.

The proteomic analysis included all long COVID cases available from the COVICAT study, who had been infected with SARS-CoV-2 before the 2023 follow-up and had consecutive samples from both the 2020 and 2023 follow-ups ($n=133$) and never long COVID participants at 2023 ($n=38$), totaling 342 paired samples (see flowchart of the study population (COVICAT study), Fig. 1). Samples were selected based on availability and although no formal matching, statistical models were adjusted for age and sex. Due to sample size constraints and the nature of the pandemic, 43.9% of participants in the 2020 follow-up who had already been infected before that time were included, and this was accounted for in the analysis. Online questionnaires were conducted during each follow-up (2020, 2021, and 2023) to collect self-reported data on COVID-19 symptoms, prior chronic conditions, vaccination status, and demographic factors such as age, sex, education, lifestyle, and BMI. Notably, 96.5% of the participants had received a COVID-19 vaccine before the 2023 follow-up.

Ethical approval for the COVICAT study was obtained from the Parc de Salut Mar Ethics Committee (CEIM-PS MAR, number 2020/9307/I) and the Hospital Universitari Germans Trias i Pujol Ethics Committee (CEI number PI-20–182). All participants provided informed consent.

Long COVID definition

Long COVID was defined according to the World Health Organization (WHO) definition and recent publications [14, 24–26] as the presence of a previous SARS-CoV-2 infection and a self-report of at least one symptom or sequel for a duration of at least 3 months. The complete list of symptoms considered in the follow-up questionnaires for long COVID is summarized in Additional file 1: Table S1. This definition aligns with the 2024 NASEM definition of long COVID [27] as it considers a larger duration of symptoms (three instead of 2 months), compared to that of WHO. Self-reported questionnaires were preferred because mild cases—most common in our cohort—are often missing from health records, risking major bias, and direct physician exams were unfeasible. The criteria for SARS-CoV-2 infection have been described in previous publications by our group [3, 22]. In summary, SARS-CoV-2 infection was defined through self-reported positive tests or illness, record linkage with EHRs, or seropositivity from 2020 to 2023. Seropositivity was assessed based on antibody presence. For a detailed seroanalysis description, see Karachaliou et al. [22, 28]. A detailed description of the questions from 2021 and 2023 surveys are available as Supplementary material (Additional file 2).

In concordance with our previous study [3], long COVID cases were classified by combining symptom reports from the first (2021) and second (2023) follow-ups. Since the 2023 interview covered a 2-year follow-up period, we ensured that symptoms in the active long COVID group were present in 2023 by including only those who reported symptoms during the week prior to the survey. Those who did not report symptoms during the last week were classified as recovered. In summary, the defined case groups and subgroups were as follows: never long COVID ($n=38$, 22.2%, no long COVID symptoms reported in 2021 and 2023); ever long COVID ($n=133$, 77.8%, long COVID symptoms reported in 2021 or 2023), further divided into two subgroups: active long COVID ($n=92$, 53.8%, long COVID symptoms reported in 2021 or 2023, and still experiencing the symptoms the week prior to the 2023 questionnaire) and recovered long COVID ($n=41$, 24%, long COVID symptoms reported in 2021 or 2023, but no symptoms reported in the week prior to the 2023 questionnaire) (see Fig. 1). This classification allowed for the analysis of the progression and resolution of long COVID symptoms over time. Sociodemographic characteristics and symptoms of the analyzed patients are presented in Additional file 1: Tables S2 and S3, respectively.

Blood samples and proteomic profiling

Blood samples

Samples were collected in EDTA tubes at the premises of the Banc de Sang i Teixits de Catalonia (BST, Barcelona), stored at 4 °C, and processed within 24 h at the GCAT lab (IGTP, Badalona). After collection, the samples were

(See figure on next page.)

Fig. 1 Overview of the samples with proteomics data in the COVICAT study and key findings related to long COVID and VEGFA. Long COVID status was assigned using questionnaire data, and proteomic profiles were obtained for 171 individuals, of whom 133 were classified as long COVID cases. Differential expression analysis revealed that 8% of proteins were significantly altered in long COVID. VEGFA emerged as the top-ranked protein across multiple centrality measures and was identified as the central hub of the long COVID protein network. A significant sex interaction was observed for VEGFA expression, and analysis in an independent sample from the COVICAT cohort showed that VEGFA was differentially expressed only among postmenopausal women

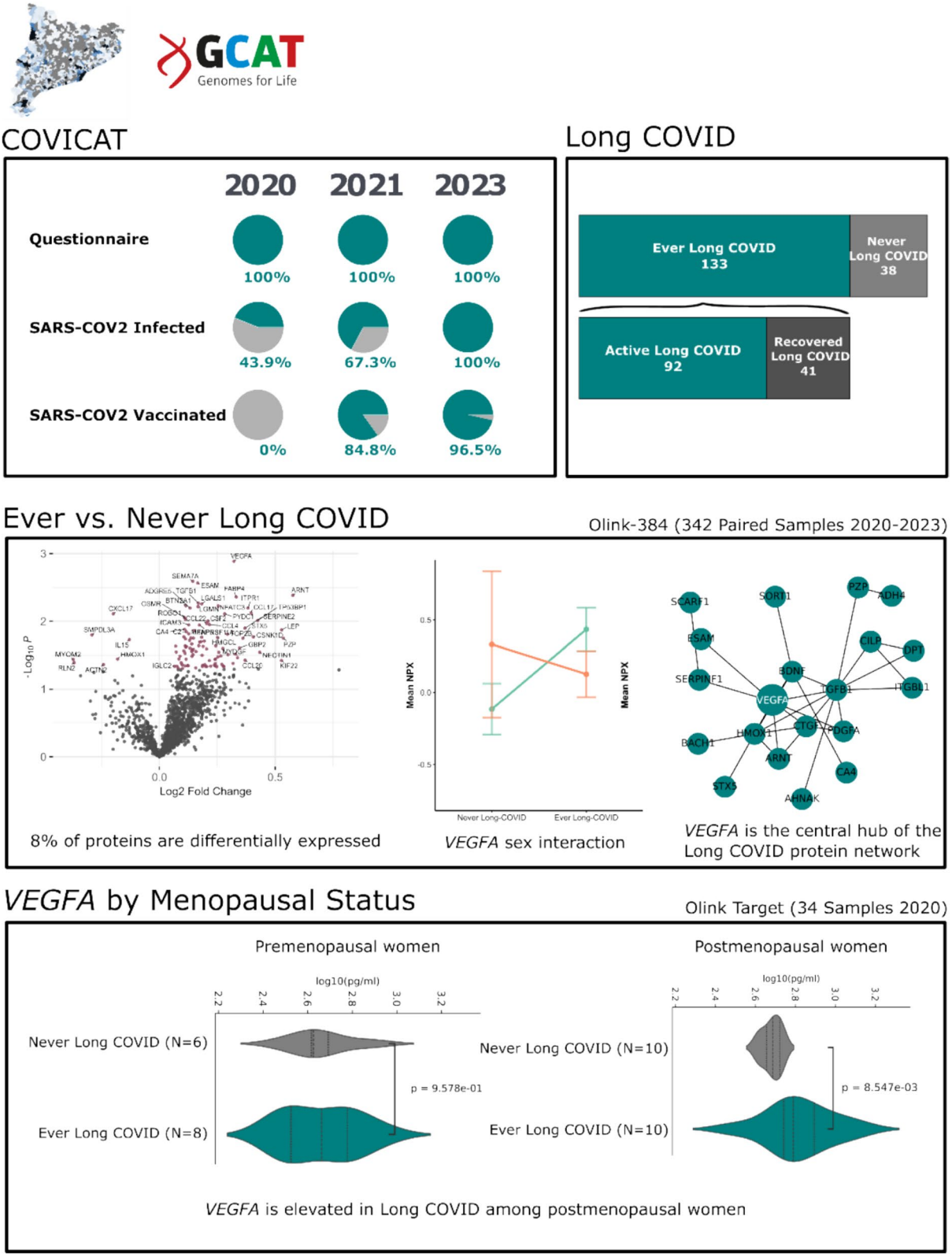


Fig. 1 (See legend on previous page.)

centrifuged, aliquoted into two fractions (plasma, buffy coat), and frozen at − 80 °C. Plasma samples were sent for protein analysis (in vials of 80 µl). To minimize bias, the 342 paired samples were analyzed in 171 pairs to include samples belonging to the same individual in the same plate, then randomization based on sex and age.

Proteomic profiling

A high-throughput proteomics analysis was performed to identify key biological pathways associated with inflammation and cardiometabolic health. This analysis included a total of 1467 assays using Olink® Technology at TATAA Biocenter (Uppsala, Sweden), which employs multiplexed protein panels and Proximity Extension Assay (PEA) technology for highly specific and sensitive measurement of protein levels. For this study, four high-multiplex biomarker discovery panels from Olink-Explore 384 were used: inflammation I, inflammation II, cardiometabolic I, and cardiometabolic II (complete list in Additional file 1: Table S4). Quality control procedures adhered to Olink's guidelines, using normalized protein expression (NPX) values to ensure data reliability and validity across all samples. Of the included assays, 27 were excluded due to low expression or technical errors (missing values in all samples), while the remaining assays had a 0% missing rate. Additionally, 45 assays were removed because they had values below the limit of detection in more than 75% of the samples. Six redundant proteins were present in two different panels, with correlations (r value from linear regression) ranging from 0.816 to 0.993; only one instance of each protein was retained. After quality control, 1395 proteins (95%) were included in the subsequent analyses. The list of proteins included in the analysis after quality control procedures is presented in Additional file 1: Table S5. For validation purposes, the protein of interest was analyzed in an independent sample from COVICAT ($n = 34$) (see later for group description) using the Olink Target 48 Cytokine panel, which quantifies key inflammatory cytokines with high-sensitivity PEA technology.

Statistical analysis

We analyzed plasma protein expression changes in individuals using consecutive samples collected during the 2020 and 2023 follow-ups. Linear mixed models were applied to examine protein changes and long COVID status. Additionally, functional and interaction analyses were conducted to explore biological pathways involved, the position of the proteins of interest within the protein–protein interaction (PPI) network, and the effect of sex in the association.

We first applied Unified Manifold Approximation and Projection (UMAP) to the entire proteomics dataset to uncover potential artefactual structures within the data. This technique utilized dimensionality reduction, transforming high-dimensional data into a visually interpretable space, which facilitated the identification of clusters and relationships among biomolecular entities.

Regression analyses

Linear mixed models (LMMs) were implemented using the lme4 package in R, to examine differential protein expression between the two time points in relation to long COVID status. The outcome variable was the normalized protein levels (NPX levels), and the variable of interest was long COVID status. Two separate analyses were performed: the first compared individuals classified as ever long COVID versus never long COVID, and the second involved pairwise comparisons among active long COVID, recovered long COVID, and never long COVID. As covariates, we included sex, age, follow-up year (sample batches from 2020 to 2023), and whether the subjects had been infected by SARS-CoV-2 in 2020. Individual identifier was included as a random effect to account within-subject variability. Post hoc pairwise comparisons were performed and marginal means differences were estimated using the emmeans package with Tukey adjustment for multiple comparisons. Degrees of freedom were estimated using Satterthwaite's approximation.

Sex-specific protein expression and interaction analyses

To explore sex-specific differences in protein expression related to long COVID, we applied the same linear mixed model as described earlier, but included an interaction term between long COVID status and sex, with sex excluded as a covariate.

Multivariate analysis of VEGFA expression

Given the relationship with disease severity, immune response, and the observed sex differences in long COVID, we aim to investigate how these factors affect *VEGFA* expression with three models: (I) first, the effect of COVID-19 severity on protein levels; (II) second, the strength of the immune humoral response by assessing IgG levels against SARS-CoV-2 as a proxy; and finally, (III) the potential impact of hormonal state by evaluating menopause status on protein expression in women.

Due to the high vaccination rates in our cohort, this analysis focused on serological SARS-CoV-2 data collected in 2020, prior to vaccine administration [22] (see Fig. 1). The severity of the acute infection was defined by self-reported data on hospitalization, ICU admission, and clinical assessment of disease severity. The immune response strength against the infection was

assessed through measurements of SARS-CoV-2 antibodies (IgG) in plasma samples, with IgG levels log10 transformed to normalize the data. Menopausal status was determined based on self-reported information from the pre-pandemic GCAT questionnaires in 2014 and 2018, combined with linked EHR data, using ICD-10 codes (N95: menopausal and female climacteric states, E894: postmenopausal states, E2831: ovarian failure, Z780: menopausal or female climacteric state, Z9079: other specified post procedural states, Z90722: postmenopausal bleeding) and ICD-9 codes (73,301: postmenopausal osteoporosis, V074: follow-up examination for postmenopausal status).

Additionally, *VEGFA* levels were measured using the Olink® Target 48 panel in an independent set of 34 female participants to validate the initial finding. All participants had completed the relevant questionnaires and were not pregnant at the time of the 2020 data collection. Blood samples were collected before any SARS-CoV-2 infection, as confirmed by negative serological status and the absence of self-reported or EHR evidence of infection. To minimize potential confounding, we included only women who reported a single SARS-CoV-2 infection by 2023 and selected individuals with comparable age distributions. The final sample included 14 premenopausal women (6 ever long COVID, 8 never long COVID) and 20 postmenopausal women (10 ever long COVID, 10 never long COVID). Group differences were tested using the Mann–Whitney *U* test.

Statistical analyses were performed using the R statistical software (version 4.3.0; R Foundation for Statistical Computing, Vienna, Austria) and Python 3.12.6. To account for multiple comparisons, p values were corrected using the Benjamini–Hochberg false discovery rate (FDR) method. In addition, nominal statistical significance was defined as p value < 0.05 without correction.

Functional analysis of expression patterns

All proteins were ranked according to their coefficients from the ever long COVID versus never long COVID comparison and analyzed using gene set enrichment analysis (GSEA) via the WebGestalt platform [29]. The analysis focused on Gene Ontology (GO) terms, diseases, biological pathways, and cellular function hallmarks. GSEA assesses whether predefined sets of genes associated with specific biological processes, pathways, or cellular functions are statistically enriched at the top or bottom of the ranked list. By applying this approach, we aimed to identify key biological functions and pathways associated with differential protein expression, providing insights into the molecular mechanisms underlying long COVID.

Protein–protein interaction network

*S*ignificant proteins (uncorrected p value > 0.05) identified in the previous analysis were explored for protein–protein interaction network using the STRING database (version 11.5; https://version-11-5.string-db.org/). We selected a medium-confidence interaction score threshold of 0.4 to identify relevant protein interactions. Reciprocal edges were downloaded in TSV format, and the Leiden algorithm was applied for community detection [30], utilizing the combined score as edge weights. Centrality metrics, including degree, betweenness, closeness, and eigenvector centrality, were computed for each protein using the NetworkX library. Z-scores for each centrality measure were calculated to evaluate the significance of the protein of interest relative to the entire network. Proteins were ranked based on their centrality metrics to determine the position of the protein of interest within the PPI network.

Results

Sociodemographic and clinical presentation of long COVID patients

The study included 171 SARS-CoV-2–infected individuals: 92 (54%) with active long COVID, 41 (24%) recovered long COVID, and 38 (22%) never experiencing long COVID. The sociodemographic and clinical characteristics of long COVID patients align with the risk factors identified in the COVICAT study; being female, having a lower education level, higher BMI, and greater SARS-CoV-2 severity were associated with increased risk, while physical activity was a protective factor [3]. However, no differences were observed among groups related to age and sleep habits (see detailed data on Additional file 1: Tables S2 and S3). Among the 171 participants, 117 (68.4%) were female, with a balanced sex distribution in the recovered group (49% female). The active long COVID patients had a higher baseline BMI (27.7 ± 5) compared to never long COVID group (25.4 ± 4.5), with the highest BMI in the recovered group (28.3 ± 5.7). Higher education was more common in recovered patients (63% vs. 45% overall). Sleep and smoking habits were similar across groups, but never long COVID individuals reported more physical activity, while recovered patients had lower alcohol consumption. Long COVID patients experienced more severe SARS-CoV-2 infections than the never long COVID group, though severity was similar between active and recovered cases. Vaccination rates were comparable, except for lower 3rd and 4th dose completion in the recovered group. Detailed data is in Additional file 1: Table S2.

In active long COVID patients, the most common symptoms were muscular (45%), neurological

symptoms (fatigue) (45%), and respiratory problems (29.3%). Recovered long COVID patients reported respiratory symptoms (26.8%) and loss of taste or smell (24.4%) most frequently, with lower rates of muscular (9.8%) and neurological (14.6%) symptoms. Prior chronic conditions were more common in the active long COVID group, including joint problems (12%), mental health disorders (6.5%), and digestive diseases (11%). Recovered long COVID patients had fewer prior conditions, with only one case of rheumatoid arthritis (2.4%) and two cases of hypertension (4.9%). Overall, recovered long COVID patients reported fewer symptoms and prior chronic conditions than active cases. Detailed information is in Additional file 1: Table S5.

Protein expression patterns in long COVID

The UMAP representation revealed two clusters that correspond to the year of the sample, indicating that samples from the same timepoint are more similar than samples from the same individual (Additional file 3: Fig. S1). Hence, in the subsequent analyses, we included the follow-up year as a covariate in the linear mixed model (LMM) to account for these differences.

The LMM analysis identified 110 proteins (~8% of all assessed proteins) with nominal associations (uncorrected p value < 0.05) differentially expressed between ever long COVID and never long COVID patients; 103 have higher levels in long COVID patients and 7 showed lower levels (Fig. 2). Our findings align with existing literature, confirming the overexpression of the *C2* protein in

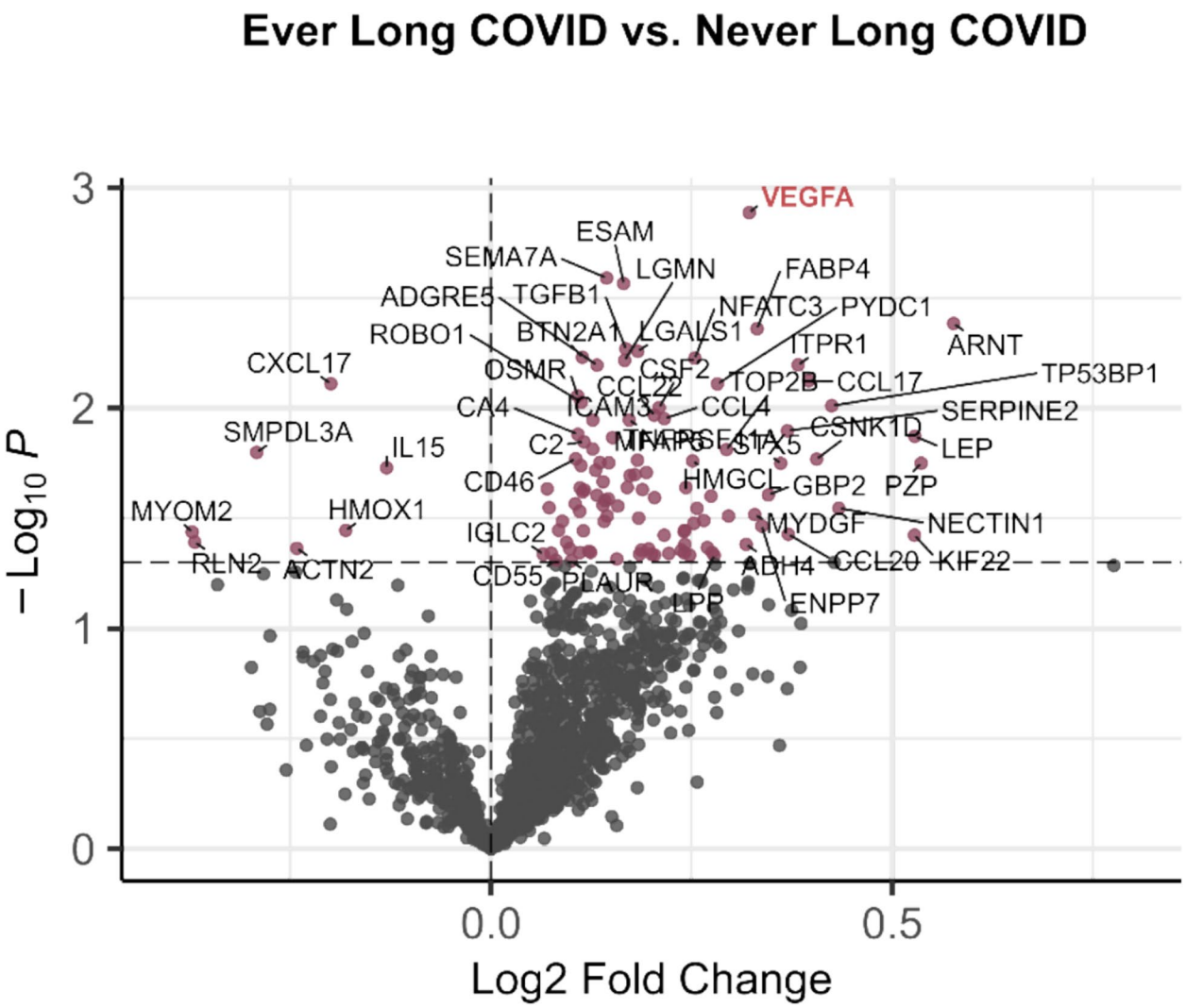


Fig. 2 Volcano plot showing the differential expression of proteins between ever long COVID cases versus never long COVID. Nominally significant proteins are highlighted, with notable examples such as *VEGFA*, colored in red, which exhibits differences associated with long COVID status

ever long COVID patients ($\beta = 0.116$, SE = 0.047, 95% CI [0.024, 0.208], $p = 0.014$), previously identified by Cervia-Hasler as a key pathomechanistic factor in long COVID, associated with complement activation [15]. Additionally, we observed elevated levels of certain chemokines, such as *CCL20*, *CCL22*, *CCL17*, and *CCL4*, with β estimates of 0.370, 0.204, 0.396, and 0.216, respectively; standard errors of 0.177, 0.079, 0.146, and 0.084; 95% confidence intervals of [0.022, 0.719], [0.048, 0.359], [0.107, 0.684], and [0.050, 0.382]; and nominal p values of 0.037, 0.011, 0.008, and 0.011, respectively. These findings are indicative of inflammatory states and have been similarly highlighted in other studies examining post-COVID conditions. Recent studies have shown that naturally occurring anti-chemokine antibodies are associated with more favorable outcomes in COVID-19 and may predict the absence of long COVID in individuals' post-infection. These findings suggest a sustained immune response and potential immune dysregulation following COVID-19, highlighting the role of chemokines in the ongoing inflammatory processes and pathophysiology of long COVID [31].

A differential key finding in our study was the overexpression of vascular endothelial growth factor A (VEGF-A) significantly overexpressed in individuals with ever long COVID compared to those without long COVID with an effect size estimate of 0.322 (SE = 0.098, 95% CI [0.128, 0.516], $p = 0.0013$). Moreover, we identified several other proteins not previously described in this context. These include *SEMA7A* (semaphorin 7A) ($\beta = 0.144$, SE = 0.047, 95% CI [0.051, 0.237], $p = 0.0026$), *ESAM* (endothelial cell adhesion molecule) ($\beta = 0.165$, SE = 0.054, 95% CI [0.058, 0.272], $p = 0.0027$), *ARNT* (aryl hydrocarbon receptor nuclear translocator) ($\beta = 0.576$, SE = 0.198, 95% CI [0.185, 0.968], $p = 0.0041$), and *FABP4* (fatty acid binding protein 4). The full details for all proteins are provided in Additional file 1: Table S6.

Given the constraints of the exploratory nature of the analysis, but grounded in the robustness of the observed results for C2 and chemokines (CCs) in long COVID [15, 31], we conducted a stratified analysis within long COVID pairwise comparisons based on symptom remission in 2023. This approach allowed us to further investigate the subgroups within long COVID and provide additional validation to the associations observed in the primary analysis. The pairwise analysis revealed significant differences in protein expression patterns. Specifically, 69 proteins distinguished active from never long COVID patients (Additional file 3: Fig. S2, Additional file 1: Table S7), 33 proteins differentiated the active and recovered long COVID patients (Additional file 3: Fig. S3, Additional file 1: Table S8), and 12 proteins distinguished the recovered patients from infected patients who had never experienced any symptoms (Additional file 3: Fig. S4, Additional file 1: Table S9).

The overlap analysis revealed distinct protein expression patterns among active long COVID, recovered long COVID, and never long COVID individuals. The most pronounced differences occurred between active and never long COVID, while recovered individuals showed a partial reversion toward baseline levels, suggesting functional involvement in long COVID pathophysiology. This suggests that long COVID is associated with sustained alterations in protein expression, with recovery linked to a gradual normalization of these changes. An Upset plot illustrating the overlap of differentially expressed proteins across multiple comparisons is shown in Fig. 3.

From identified proteins, *VEGFA*, with an effect size estimate of 0.375 (SE = 0.103, 95% CI [0.130, 0.620], $p = 0.001$) remained the best feature when compared active long COVID with never long COVID individuals. A volcano plot summarizing these findings is presented in Additional file 3: Fig. S2. Additionally, *SEMA7A* (semaphorin 7A) remained overexpressed in both active and recovered long COVID patients compared to never long COVID individuals ($\beta = 0.143$, SE = 0.050, 95% CI [0.025, 0.261], $p = 0.012$; and $\beta = 0.146$, SE = 0.058, 95% CI [0.009, 0.283], $p = 0.034$, respectively). In contrast, four proteins, *DLL4* (delta like canonical Notch ligand 4), *LGALS1* (galectin 1), *RNASET2* (ribonuclease T2), and *ST6GAL1* (ST6 beta-galactoside alpha-2,6-sialyltransferase 1), were significantly overexpressed exclusively in active long COVID patients, not only compared to never long COVID individuals (β estimates of 0.551, 0.236, 0.154, and 0.147; p values of 0.038, 0.0019, 0.0051, and 0.034, respectively) but also to recovered long COVID patients (β estimates of 0.553, 0.163, 0.151, and 0.140; p values of 0.036, 0.046, 0.006, and 0.047, respectively). Detailed results can be found in Additional file 1: Tables S7–S9.

Functional analysis of long COVID: immune, vascular, and muscular changes

Gene set enrichment analysis (GSEA) identified several significant pathways across multiple categories, suggesting a complex interplay between immune dysregulation, vascular dynamics, and muscular impairment in the pathophysiology of long COVID. Pathway analysis revealed strong enrichment for immune and inflammatory pathways, including human cytomegalovirus infection (normalized enrichment score [NES] = 1.905, FDR = 0.0097). EBV reactivation and underlying CMV infection have been linked to SARS-CoV-2 infection, contributing to the complex risk patterns observed in long COVID [32]. This is concordant with our observation in the COVICAT study that ever/active long COVID patients exhibited elevated immunoglobulin G levels to

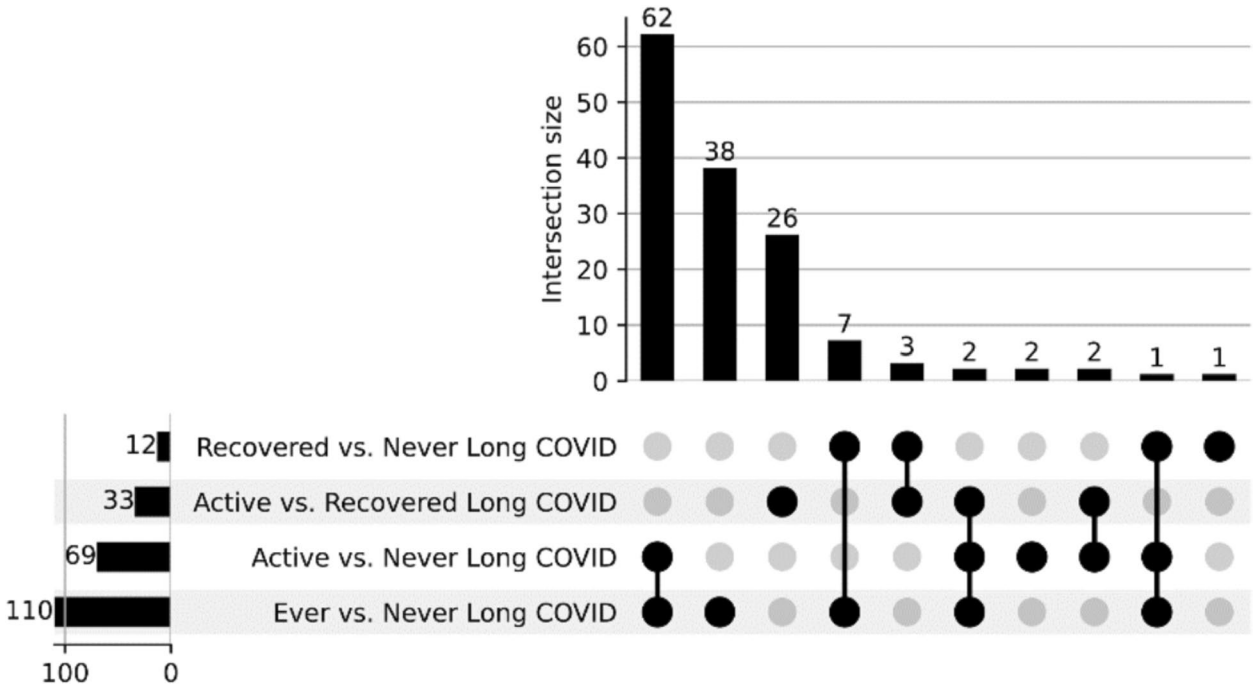


Fig. 3 Upset plot illustrating the overlap of differentially expressed proteins across comparisons between ever long COVID, active long COVID, recovered long COVID, and never long COVID groups. The plot highlights the shared and unique proteins identified in each comparison, emphasizing the proteins significantly altered in both ever long COVID versus never long COVID and active long COVID versus never long COVID

EBV early antigen-diffuse [28]. The chemokine signaling pathway (NES = 1.760, FDR = 0.029) is also enriched, in line with the observations of [31], who described that naturally occurred antibodies against specific chemokines were omnipresent post-COVID-19 and negatively correlated with the development of long COVID 1 year post-infection. This enrichment is driven by *CXCL8* [31], *CCL5* but also by *VEGFA*. Angiogenesis was significantly enriched across multiple databases, including *Panther* (NES = 1.873, FDR = 0.0023) and *Wikipathways* (NES = 1.917, FDR = 0.0056), emphasizing vascular involvement [18, 33] through genes like *VEGFA*, *HIF1A*, and *JUN*.

Only seven proteins showed reduced expression, which is a relatively small proportion compared to the 103 proteins that are overexpressed. Among the pathways, the cellular component *contractile fiber* (NES = −2.125, FDR = 0.0056) and molecular function *structural constituent of muscle* (NES = −2.374, FDR = 0.0015) were negatively enriched, implicating proteins such as *ACTN2*, *MYL3*, and *DMD* in potential muscle dysfunction. Additionally, the pathway *striated muscle contraction* showed negative enrichment in both Reactome (NES = −2.118, FDR = 0.0359) and Wikipathways (NES = −2.141, FDR = 0.023). Complete list of results is presented in Additional file 1: Table S10. These findings in muscle pathways align with further evidence linking long COVID to disrupted muscle structure and function, potentially associated with muscular issues—one of the most prevalent symptoms in active long COVID patients [13, 34].

Network analysis

Analysis of 110 ever long COVID differentially expressed proteins (ever long COVID vs. never long COVID) using the STRING database revealed a PPI network with 109 nodes and 274 edges. Within this network, *VEGFA* was identified as the most central protein, ranking highest across all centrality metrics. This suggests a crucial role for *VEGFA* in the molecular interactions associated with long COVID (Additional file 1: Table S11, Additional file 3: Fig. S5). The calculated centrality values for *VEGFA* were degree centrality (0.385), betweenness centrality (0.189), closeness centrality (0.571), and eigenvector centrality (0.297). Z-scores for *VEGFA* further underscore its prominence, with a degree z-score of 3.29, a betweenness z-score of 4.73, a closeness z-score of 1.99, and an eigenvector z-score of 2.63. These scores indicate that the connectivity and influence of *VEGFA* are significantly higher than the average of the proteins in the network, reinforcing its critical role in mediating interactions relevant to long COVID.

Community detection using the Leiden algorithm identified nine distinct clusters, or communities, within

the protein–protein interaction network. These communities represent groups of proteins that are more interconnected with each other than with the rest of the network, suggesting potential shared biological functions or regulatory pathways. The overall modularity score of 0.321 indicates a well-structured but moderately modular organization (Additional file 3: Fig. S6). *VEGFA* was positioned within the largest community, which consisted of 20 proteins, including *TGFB1*, *CTGF*, *HMOX1*, *BDNF*, *SERPINF1*, *PDGFA*, *CILP*, *DPT*, *ARNT*, *PZP*, *ITGBL1*, *AHNAK*, *STX5*, *ESAM*, *ADH4*, *CA4*, *SORT1*, *BACH1*, and *SCARF1* (Fig. 4). Other communities detected in this analysis are represented in Additional file 3: Fig. S7.

Sex-specific interactions

We identified 35 proteins with sex-specific differences in expression when comparing ever long COVID vs. never long COVID, showing a nominally significant interaction with sex. Notably, *C2* (p value of 3.63×10^{-3}), *VEGFA* (p value $=7.78\times10^{-4}$), *TSPAN1* (p value $=1.63\times10^{-3}$), *M6PR* (p value $=3.94\times10^{-4}$), and *DENND2B* (1.82×10^{-3}) were among the most significant (see detailed results in Additional file 1: Table S12). Interestingly, relative *VEGFA* expression levels between males and females are reversed in patients with ever long COVID, with higher levels observed in females than in males. For the other proteins, a sex interaction is observed, although their levels converge in both sexes at the ever long COVID stage (Fig. 5 and Additional file 3: Figs. S8, S9, S10, S11, and S12). Moreover, the expression levels of *VEGFA* were slightly higher in women exposed to the virus compared to those not exposed, suggesting a gender-specific difference in response (Additional file 3: Fig. S13).

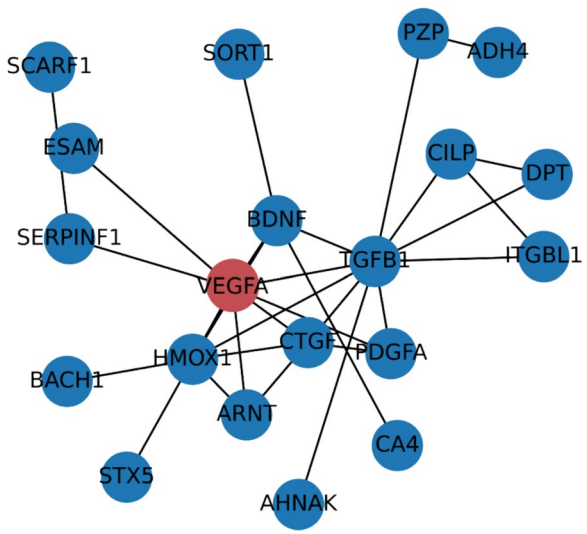


Fig. 4 Community structure of the long COVID-associated protein–protein interaction (PPI) network, identified using the Leiden algorithm. The network was divided into nine distinct communities. The figure depicts the largest community (community 1: $n=20$ proteins), where *VEGFA*, the most central protein in the network, is located and highlighted in red to emphasize its central role

Multivariate modeling on VEGFA expression

Regression modeling was conducted in long COVID patients to examine the influence of COVID-19 severity, immune humoral response (IgG levels), and hormonal status (menopausal state) on *VEGFA* expression. The results indicate that while *VEGFA* was significantly overexpressed in long COVID individuals (ever and active) versus never long COVID, none of the models identified a significant impact of menopausal status, IgG levels, or infection severity on its expression (Additional file 1: Table S13).

Replication and validation

We replicated the *VEGFA* findings and further analyzed the impact of sex and hormonal status by stratifying participants based on menopausal status. Using an independent assay with the same technology, *VEGFA* results were validated in a small dataset of 34 female participants sampled in 2020, confirming its overexpression in individuals with ever long COVID. Interestingly, when participants were stratified by menopausal status, a significant difference in circulating *VEGFA* levels was observed, specifically among postmenopausal women (Mann–Whitney U test p value $=8.55\times10^{-3}$) (Fig. 6). This suggests that menopausal status may influence *VEGFA* expression in the context of long COVID.

Discussion

This study provides novel insights into the pathophysiology of long COVID, uncovering significant immune dysregulation, vascular remodeling, and sex-specific differences. By integrating comprehensive proteomic data from a deeply characterized, general population-based cohort, we identified distinct patterns in protein expression that correlate with (i) the persistence of long COVID symptoms and (ii) exhibit notable sex interactions. These findings highlight critical mechanisms that may contribute to the long-term effects of the disease and suggest a need for sex-specific approaches in both diagnostics and treatment. Our findings uncover new pathophysiological pathways, including those associated with angiogenesis, such as *VEGFA* (vascular endothelial growth factor), while also corroborating prior observations, such as the involvement of chemokine signaling pathways, complement activation, and viral reactivation in post-COVID-19 syndromes.

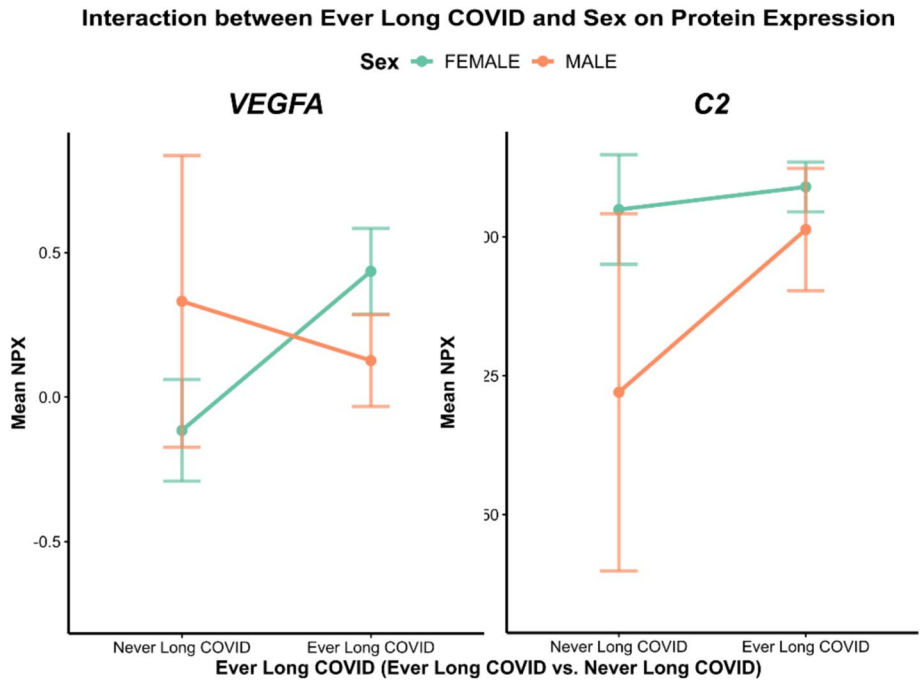


Fig. 5 Interaction plot depicting the expression levels of VEGFA and C2 in 2020 stratified by sex and long COVID status. The results indicate that the expression levels of these genes exhibited differences prior to the onset of long COVID. For C2, the expression levels in long COVID cases are more similar between males and females, while a greater difference is observed in the control group. In contrast, VEGFA levels show a different pattern: in males, expression decreases in long COVID cases compared to controls, whereas in females, expression increases in cases compared to controls, resulting in an "X" shaped interaction

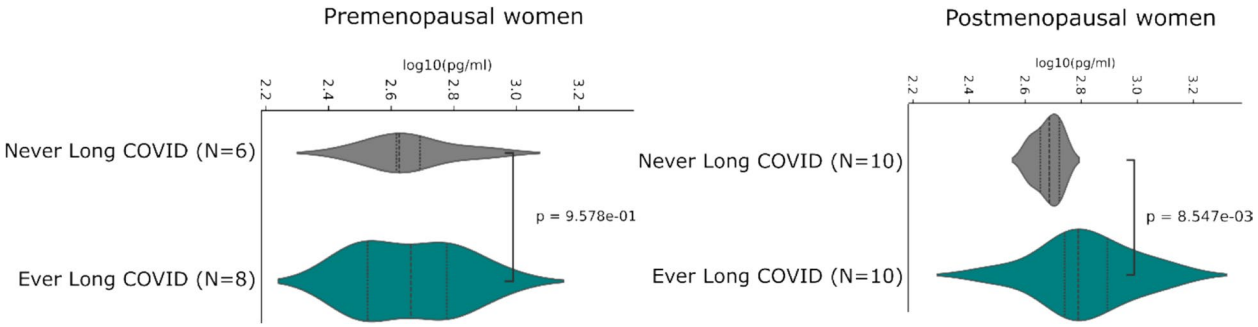


Fig. 6 Validation of circulating VEGFA levels (pg/mL) in an independent cohort of women, stratified by menopausal status and long COVID history. Among postmenopausal participants, VEGFA levels were significantly higher in ever long COVID cases compared to never long COVID controls, suggesting a potential interaction between menopausal status and long COVID in modulating VEGFA expression

We confirmed the enrichment of pathways such as chemokine signaling correlate with long COVID aligning with reported observations that natural occurring antibodies against chemokines are protective against the persistence of long COVID symptoms [31]. Chemokines have been associated with disease severity and were among the first widespread genetic signals to be identified in COVID-19 research [35]. Interestingly, Cheong et al. described that alterations in innate immune phenotypes and epigenetic programs persisted for months following severe COVID-19 [36]. These findings underscore a potential interplay between viral persistence, immune reprogramming, and dysregulated inflammatory responses that may drive the pathophysiology of long COVID.

Our findings highlight sex-based differences in protein expression in ever long COVID patients. For instance, while *C2* levels were comparable in both sexes

for ever long COVID patients, there was a higher relative increase in males compared to females never long COVID. Conversely, proteins like *VEGFA*, *TSPAN1*, *DENND2B*, and *M6PR* showed an opposite trend in each sex: in females, these proteins were lower in never long COVID but increased in ever long COVID, while these levels decreased in ever long COVID males compared to females. These results suggest a complex relationship between sex, immune response, recovery changes, and susceptibility in long COVID. *TSPAN1*, *M6PR*, and *DENND2B* are involved in immune cell communication, trafficking, and viral responses. Their sex-specific expression patterns, where males show higher expression in never infected states but converge in long COVID status, suggest that while there are inherent differences in immune baseline states between males and females, these differences may diminish during active immune responses. Additionally, *C2*, a protein involved in activating the classical complement pathway, showed notable sex differences. The complement system is a key component of innate immunity, and its role in immune regulation is known to exhibit sexual dimorphism. For example, previous studies have shown that complement system components, including *C4*, are elevated in women and contribute to their increased susceptibility to autoimmune diseases such as systemic lupus erythematosus (SLE) and Sjögren's syndrome [37]. Our findings suggest that such sex-specific complement activity may also play a role in modulating long COVID, but despite sex variations in expression, there are no significant differences in the response to long COVID across sex, indicating a more universal pattern of pathophysiological processes.

VEGFA behaves differently. This protein is a powerful growth factor that promotes endothelial cell migration, proliferation, tube formation, and survival. It also acts as a key regulator of vascular permeability, serving as a critical link between inflammation, permeability, and angiogenesis. Network analysis positioned VEGFA within the largest community of the network, which included 20 proteins with key roles in tissue remodeling and vascular function. For example, some of the genes in the *VEGFA*-associated community, such as *ARNT*, *SORT1*, *SERPINF1*, and *HMOX1*, play key roles in regulating *VEGFA*-driven processes. Hypoxia signaling, mediated by *ARNT*, regulates *VEGFA* expression in response to oxygen levels. *SORT1* controls endothelial function and lipid metabolism, which are vital for vascular integrity. *SERPINF1* modulates angiogenesis by counteracting *VEGFA*, while *HMOX1* manages oxidative stress and inflammation, both of which influence *VEGFA*-driven vascular remodeling. These proteins collectively regulate *VEGFA*-driven vascular remodeling and inflammation, processes that are disrupted by SARS-CoV-2 infection suggesting that *VEGFA*'s role in long COVID pathophysiology may extend beyond angiogenesis, involving complex regulatory mechanisms related to inflammation, neurovascular remodeling, and extracellular matrix dynamics.

Our discovery of *VEGFA* altered expression with long COVID aligns with the hypothesis of endothelial dysfunction as a contributing factor to long COVID. Supporting this hypothesis, several studies have reported vascular abnormalities across various organ systems in individuals affected by COVID-19 [38], and vascular biomarkers in long COVID patients, 1 to 6 months post-infection [18]. The virus is known to target the endothelium via the angiotensin-converting enzyme 2 (*ACE2*), resulting in direct damage to vascular endothelial cells by spike proteins [39] affecting different organs [40]. A wide range of vascular complications have been associated with SARS-CoV-2 infection, leading some to propose that long COVID may primarily be a vascular disease, with vascular pathologies observed post-infection, which have been identified as potential manifestations of long COVID [38]. Reduced retinal vessel density (RVD), a surrogate marker for endothelial dysfunction, has been associated with long COVID [41], and a 10% of long COVID patients within cohort studies have retinal microvascular changes [42]. Recent research by Philippe et al. [16] emphasizes the role of vascular endothelial growth factor A (VEGF-A) in reduced DLCO and radiological abnormalities associated with respiratory symptoms in long COVID, particularly among hospitalized and ICU-treated patients. However, their study included a limited number of outpatients and underrepresented women, without exploring sex-based differences. Our study, which primarily involves mild cases, confirms that VEGF-A–related vascular alterations persist beyond the acute phase and occur independently of disease severity.

Regarding the mechanisms involved, VEGF-A's pro-inflammatory actions—such as increasing vascular permeability and promoting immune cell infiltration—may drive a neuroinflammatory feedback loop that contributes to long COVID symptoms [20, 21]. Angiogenesis biomarkers, including ANG-1 and P-SEL, have shown high diagnostic accuracy (96%) for identifying long COVID patients, underscoring vascular dysregulation as a central pathogenic factor [17, 18]. Although the hypothesis that VEGF-A–induced disruption of neurovascular communication could be triggered by vaccination is intriguing [21], our study excludes vaccination as

a confounding variable because samples were collected prior to widespread vaccine availability.

Observed sex dimorphisms for VEGF-A align with findings from animal models underscoring the critical role of VEGF-A in ovarian function, particularly in follicle development, where its expression appears to follow dynamic patterns tightly regulated by hormonal signals [43]. VEGF-A expression is hormonally regulated with well-established sex-specific differences; estrogen directly enhances VEGF transcription [43, 44]. Menopausal declines in estrogen are linked to altered vascular signaling and increasing vascular vulnerability in women [45]. Higher VEGF-A expression was observed in long COVID patients who were postmenopausal, with a bimodal distribution among premenopausal individuals. This pattern suggests a plausible influence of female hormonal status on VEGF-A regulation. It is known that the uterus regulates circulating VEGF-A levels [46], which could be compromised in postmenopausal women, potentially increasing their susceptibility to developing long COVID symptomatology, suggesting the co-existence of distinct mechanistic pathways in long COVID females. This provides an additional clue to the sex-specific interactions observed in long COVID, providing a mechanistic hypothesis between sex differences in long COVID susceptibility and vascular health [47].

While this study provides valuable insights into the proteomic alterations associated with long COVID, with a replication in a different subsample, several limitations should be noted. First, the cohort, though deeply characterized, was relatively small and geographically constrained, which may not fully capture the heterogeneity of populations affected by long COVID. This limitation could influence the generalizability of the findings to broader, more diverse demographics. Moreover, although the cohort includes adults aged 40–65—limiting generalization to younger and older age groups—this range captures the population most commonly affected by long COVID and therefore does not limit the relevance of the findings. Similarly, while the study is geographically based in Catalonia, it builds on prior epidemiological research in this population that aligns with international evidence [3]. Second, the absence of a clear clinical phenotype for long COVID reflects the current uncertainty in defining the condition. This complicates symptom classification and documentation in epidemiological studies. Our reliance on self-reported symptoms helps address this challenge by capturing cases often underreported in health records, especially mild ones. Third, while the study spanned data collection from 2020 to 2023, this temporal window may not encompass the entire spectrum of disease progression or resolution patterns, particularly for individuals with prolonged recovery trajectories. Longer-term follow-up studies are essential to understand the persistence or evolution of the observed proteomic changes. Fourth, the integration of proteomic data offers a comprehensive view of molecular alterations. Although the enrichment results are consistent and align with VEGFA observations in a small subsample, they should be interpreted cautiously and require validation in larger, independent cohorts. Additionally, the analysis is limited by potential confounding factors—such as comorbidities, medication use, and environmental influences—that were not fully accounted for in this study. Finally, while the identification of sexually dimorphic protein expression patterns offers novel insights and has been replicated in an independent sample, these findings require further validation in larger, independent populations.

These findings suggest sex-specific mechanisms in long COVID, potentially driven by hormonal regulation of *VEGFA* and altered susceptibility in postmenopausal women. While exploratory, our findings suggest that VEGFA may be a relevant therapeutic target in long COVID, warranting further research into vascular or endothelial-directed interventions aimed at mitigating persistent symptoms. Notably, anti-VEGF therapies have been successfully applied in diseases characterized by pathological angiogenesis and vascular dysfunction such as cancer and retinal disorders [48, 49], and recent studies propose vascular stabilization as a promising strategy for post-viral syndromes involving endothelial dysregulation [40].

Taken together, these findings contribute to a growing body of evidence that underscores the need for stratified approaches in understanding and treating long COVID. Such approaches must address both biological heterogeneity and sex-based differences.

Conclusions

This study identifies key molecular signatures underlying long COVID, including immune dysregulation, vascular remodeling, and sex-specific proteomic alterations. *VEGFA* emerged as a central node associated with symptom persistence and showed distinct expression patterns across sexes. These findings suggest hormonal regulation may influence long COVID, but the role of hormonal modulation in VEGF-A expression remains preliminary and based on indirect evidence. Further studies with direct hormonal measurements are needed for confirmation. Our results underscore the importance of incorporating sex-stratified approaches in both research and clinical management of long COVID. Future studies should validate these findings in larger, more diverse populations and explore the long-term trajectories of proteomic changes.

Abbreviations

NPX	Normalized protein expression
LMM	Linear mixed model
PPI	Protein-protein interaction
GSEA	Gene set enrichment analysis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-025-04402-6.

Additional file 1: Tables S1 to S13. Table S1 Long COVID self-reported symptoms included in 2021 and 2023 COVICAT questionnaires. Table S2 Sociodemographic and clinical variables among three study groups ($n=171$) at 2023 follow-up: active long COVID, recovered long COVID, and never long COVID. Table S3 Self-reported symptoms and comorbidities among the two long COVID groups ($n=133$) and 2021 and 2023 surveys. Table S4 List of proteins used in the analysis, belonging to the cardiometabolic (I and II) and inflammation (I and II) panels. Table S5 Proteomics quality control and the number of proteins removed in each step. Table S6 Results from the linear mixed model for the ever long COVID vs. never long COVID comparison. Table S7 Results from the linear mixed model for the active long COVID vs. never long COVID comparison. Table S8 Results from the linear mixed model for the active long COVID vs. recovered long COVID comparison. Table S9 Results from the linear mixed model for the recovered long COVID vs. never long COVID comparison. Table S10 Results from the GSEA analysis performed with WebGestalt. Table S11 Centrality measures and community assignments for proteins in the long COVID-associated protein–protein interaction (PPI) network. Table S12 Results from the linear mixed model with ever vs. never long COVID and sex interaction. Table S13 Results from the linear regressions using 2020 data to test the association between VEGFA levels and menopausal status, IgG levels, and susceptibility to COVID-19 severity.

Additional file 2: Questions in 2021 and 2023 surveys to evaluate long COVID symptoms in our study population.

Additional file 3: Figures S1–S13. Fig. S1 Unified Manifold Approximation and Projection (UMAP) visualization of the proteomics dataset. Fig. S2 Volcano plot showing differential protein expression between active long COVID and never long COVID. Fig. S3 Volcano plot showing differential protein expression between active long COVID and recovered long COVID. Fig. S4 Volcano plot showing differential protein expression between recovered long COVID and never long COVID. Fig. S5 Distribution of centrality measures in the long COVID-associated protein–protein interaction (PPI) network. Fig. S6 Protein–protein interaction (PPI) network visualized with Leiden community detection. Fig. S7 Community structure of the long COVID-associated protein–protein interaction (PPI) network, identified using the Leiden algorithm. Fig. S8 Interaction plot depicting the expression levels of VEGFA stratified by sex and long COVID status, grouped by sample year. Fig. S9 Interaction plot depicting the expression levels of C2 stratified by sex and long COVID status, grouped by sample year. Fig. S10 Interaction plot depicting the expression levels of DENND2B stratified by sex and long COVID status, grouped by sample year. Fig. S11 Interaction plot depicting the expression levels of TSPAN1 stratified by sex and long COVID status, grouped by sample year. Fig. S12 Interaction plot depicting the expression levels of M6PR stratified by sex and long COVID status, grouped by sample year. Fig. S13 Expression levels of VEGFA in females in 2020 comparing females that were exposed to the virus before sampling to women that had not yet been exposed to the virus.

Acknowledgements

We express our sincere gratitude to the cohort study volunteers and the Blood and Tissue Bank team, coordinated by Dr. Alonso Nogues and Dr. Grifols, for their invaluable assistance with sample collection. We also wish to extend our special appreciation to the GCAT project investigators, particularly Anna Carreras, Beatriz Cortés, and May Myatt. Anonymized data were provided by the Catalan Agency for Quality and Health Assessment (PADRIS Program). For a complete list of investigators, please visit www.genomesforlife.com.

Authors' contributions

All authors satisfy the criteria for authorship. RdC and XF designed the study, XF did the statistical analysis, SI, FF, NB, JG, GC, MK, SA, LL contributed to data acquisition, CD, GM contributed to antibody and protein data acquisition, AE contributed to statistical interpretation. RdC, CB, IC contributed to discussion and funding, RdC drafted the manuscript and with XF contributed to interpretation of the work. MKa contributed to data interpretation. All authors reviewed the manuscript critically for important intellectual content, gave final approval of the version to be published, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and recovered. All authors read and approved the final manuscript.

Authors' social media handles

Bluesky handles: @xavifar.bsky.social (Xavier Farré).

Funding

La Caixa Foundation (SR20-01024), La Marató TV3 (167/C/2021), the Spanish Ministry of Science and Innovation (CEX2018-000806-S, CEX2023-0001290-S, TED2021-130626B-I00, RYC2020-029886-I), the Spanish Ministry of Health (PI18/01512, PI20/01267, MS19/00100), and Horizon Europe END-VOC (GA:101046314).

Data availability

The data generated and analysed during this study are available from the GCAT biobank and can be accessed by other researchers upon reasonable request. Access to the data is subject to approval by the ethics committee to ensure confidentiality and compliance with ethical standards. Data will be shared in accordance with institutional policies and any applicable legal or ethical restrictions.

Declarations

Ethics approval and consent to participate

All participants contacted had consented in the past to be re-contacted and had provided informed consent. Ethical approval was obtained from the Parc de Salut Mar Ethics Committee (CEIM-PS MAR, no. 2020/9307/I) and Hospital Universitari Germans Trias i Pujol Ethics Committee (CEI no. PI-20–182).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Genomes for Life-GCAT Lab, CORE Program, Germans Trias i Pujol Research Institute (IGTP), 08916 Badalona, Spain. [2]Grup de Recerca en Impacte de Les Malalties Cròniques i Les Seves Trajectòries (GRIMTra), IGTP, Badalona, Spain. [3]Stroke Pharmacogenomics and Genetics Group, Institut de Recerca Sant Pau (IR SANT PAU), Barcelona, Spain. [4]ISGlobal, Barcelona, Spain. [5]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. [6]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [7]IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain. [8]CIBER Enfermedades Infecciosas (CIBERINFEC), Barcelona, Spain. [9]Erasmus MC, University Medical Center, Rotterdam, the Netherlands. [10]Faculty of Health Sciences at Manresa, Universitat de Vic—Universitat Central de Catalunya (UVic-UCC), Barcelona, Spain. [11]Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain.

Received: 7 May 2025 Accepted: 17 September 2025
Published online: 10 October 2025

References

1. Thaweethai T, Jolley SE, Karlson EW, Levitan EB, Levy B, McComsey GA, et al. Development of a definition of postacute sequelae of SARS-CoV-2 infection. JAMA. 2023;329(22):1934–46.

2. Global Burden of Disease Long CC, Wulf Hanson S, Abbafati C, Aerts JG, Al-Aly Z, Ashbaugh C, et al. Estimated global proportions of individuals with persistent fatigue, cognitive, and respiratory symptom clusters following symptomatic COVID-19 in 2020 and 2021. JAMA. 2022;328(16):1604–15.
3. Kogevinas M, Karachaliou M, Espinosa A, Iraola-Guzman S, Castano-Vinyals G, Delgado-Ortiz L, et al. Risk, determinants, and persistence of long-COVID in a population-based cohort study in Catalonia. BMC Med. 2025;23(1):140.
4. Sha'ari NI, Ismail A, Abdul Aziz AF, Suddin LS, Azzeri A, Sk Abd Razak R, et al. Cardiovascular diseases as risk factors of post-COVID syndrome: a systematic review. BMC Public Health. 2024;24(1):1846.
5. Thompson EJ, Williams DM, Walker AJ, Mitchell RE, Niedzwiedz CL, Yang TC, et al. Long COVID burden and risk factors in 10 UK longitudinal studies and electronic health records. Nat Commun. 2022;13(1):3528.
6. Reme BA, Gjesvik J, Magnusson K. Predictors of the post-COVID condition following mild SARS-CoV-2 infection. Nat Commun. 2023;14(1):5839.
7. Subramanian A, Nirantharakumar K, Hughes S, Myles P, Williams T, Gokhale KM, et al. Symptoms and risk factors for long COVID in non-hospitalized adults. Nat Med. 2022;28(8):1706–14.
8. Yu Z, Ekstrom S, Bellander T, Ljungman P, Pershagen G, Eneroth K, et al. Ambient air pollution exposure linked to long COVID among young adults: a nested survey in a population-based cohort in Sweden. Lancet Reg Health. 2023;28:100608.
9. Fischer A, Zhang L, Elbeji A, Wilmes P, Oustric P, Staub T, et al. Long COVID symptomatology after 12 months and its impact on quality of life according to initial coronavirus disease 2019 disease severity. Open Forum Infect Dis. 2022;9(8):ofac397.
10. Lammi V, Nakanishi T, Jones SE, Andrews SJ, Karjalainen J, Cortés B, O'Brien HE, Fulton-Howard BE, Haapaniemi HH, Schmidt A et al. Genome-wide association study of long COVID. medRxiv 2023:2023.2006.2029.23292056.
11. Bai F, Tomasoni D, Falcinella C, Barbanotti D, Castoldi R, Mule G, Augello M, Mondatore D, Allegrini M, Cona A, et al. Female gender is associated with long COVID syndrome: a prospective cohort study. Clin Microbiol Infect. 2022;28(4):611 e619-611 e616.
12. Gheorghita R, Soldanescu I, Lobiuc A, Caliman Sturdza OA, Filip R, Constantinescu-Bercu A, et al. The knowns and unknowns of long COVID-19: from mechanisms to therapeutical approaches. Front Immunol. 2024;15:1344086.
13. Peluso MJ, Deeks SG. Mechanisms of long COVID and the path toward therapeutics. Cell. 2024;187(20):5500–29.
14. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA, Cuapio A, et al. More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. Sci Rep. 2021;11(1):16144.
15. Cervia-Hasler C, Bruningk SC, Hoch T, Fan B, Muzio G, Thompson RC, et al. Persistent complement dysregulation with signs of thromboinflammation in active long COVID. Science. 2024;383(6680):eadg7942.
16. Philippe A, Gunther S, Rancic J, Cavagna P, Renaud B, Gendron N, et al. Vegf-A plasma levels are associated with impaired DLCO and radiological sequelae in long COVID patients. Angiogenesis. 2024;27(1):51–66.
17. Iosef C, Knauer MJ, Nicholson M, Van Nynatten LR, Cepinskas G, Draghici S, et al. Plasma proteome of long-COVID patients indicates HIF-mediated vasculo-proliferative disease with impact on brain and heart function. J Transl Med. 2023;21(1):377.
18. Patel MA, Knauer MJ, Nicholson M, Daley M, Van Nynatten LR, Martin C, et al. Elevated vascular transformation blood biomarkers in long-COVID indicate angiogenesis as a key pathophysiological mechanism. Mol Med. 2022;28(1):122.
19. Bahreiny SS, Bastani MN, Keyvani H, Mohammadpour Fard R, Aghaei M, Mansouri Z, et al. VEGF-A in COVID-19: a systematic review and meta-analytical approach to its prognostic value. Clin Exp Med. 2025;25(1):81.
20. Reinders ME, Sho M, Izawa A, Wang P, Mukhopadhyay D, Koss KE, et al. Proinflammatory functions of vascular endothelial growth factor in alloimmunity. J Clin Invest. 2003;112(11):1655–65.
21. Talotta R. Impaired VEGF-A-mediated neurovascular crosstalk induced by SARS-CoV-2 spike protein: a potential hypothesis explaining long COVID-19 symptoms and COVID-19 vaccine side effects? Microorganisms. 2022:(10)2.
22. Karachaliou M, Moncunill G, Espinosa A, Castano-Vinyals G, Jimenez A, Vidal M, et al. Infection induced SARS-CoV-2 seroprevalence and heterogeneity of antibody responses in a general population cohort study in Catalonia Spain. Sci Rep. 2021;11(1):21571.
23. Obon-Santacana M, Vilardell M, Carreras A, Duran X, Velasco J, Galvan-Femenia I, et al. Gcat|genomes for life: a prospective cohort study of the genomes of Catalonia. BMJ Open. 2018;8(3):e018324.
24. Carfi A, Bernabei R, Landi F, Gemelli Against C-P-ACSG. Persistent symptoms in patients after acute COVID-19. JAMA. 2020;324(6):603–5.
25. Havervall S, Rosell A, Phillipson M, Mangsbo SM, Nilsson P, Hober S, et al. Symptoms and functional impairment assessed 8 months after mild COVID-19 among health care workers. JAMA. 2021;325(19):2015–6.
26. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV. Condition WHOCCDW-GoP-C-: a clinical case definition of post-COVID-19 condition by a Delphi consensus. Lancet Infect Dis. 2022;22(4):e102–7.
27. NASEM. In: A long COVID definition: a chronic, systemic disease state with profound consequences. edn. Edited by Goldowitz I, Worku T, Brown L, Fineberg HV. Washington (DC): National Academies Press (US). 2024.
28. Karachaliou M, Ranzani O, Espinosa A, Iraola-Guzmán S, Castaño-Vinyals G, Vidal M, et al. Antibody responses to common viruses according to COVID-19 severity and postacute sequelae of COVID-19. J Med Virol. 2024;96(9):e29862.
29. Elizarraras JM, Liao Y, Shi Z, Zhu Q, Pico AR, Zhang B. Webgestalt 2024: faster gene set analysis and new support for metabolomics and multi-omics. Nucleic Acids Res. 2024;52(W1):W415–21.
30. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. 2019;9(1):5233.
31. Muri J, Cecchinato V, Cavalli A, Shanbhag AA, Matkovic M, Biggiogero M, et al. Autoantibodies against chemokines post-SARS-CoV-2 infection correlate with disease course. Nat Immunol. 2023;24(4):604–11.
32. Peluso MJ, Deveau TM, Munter SE, Ryder D, Buck A, Beck-Engeser G, et al. Chronic viral coinfections differentially affect the likelihood of developing long COVID. J Clin Invest. 2023:133(3).
33. Wu X, Xiang M, Jing H, Wang C, Novakovic VA, Shi J. Damage to endothelial barriers and its contribution to long COVID. Angiogenesis. 2024;27(1):5–22.
34. Antar AAR, Yu T, Demko ZO, Hu C, Tornheim JA, Blair PW, et al. Long COVID brain fog and muscle pain are associated with longer time to clearance of SARS-CoV-2 RNA from the upper respiratory tract during acute infection. Front Immunol. 2023;14:1147549.
35. Severe Covid GG, Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, et al. Genomewide association study of severe COVID-19 with respiratory failure. N Engl J Med. 2020;383(16):1522–34.
36. Cheong JG, Ravishankar A, Sharma S, Parkhurst CN, Grassmann SA, Wingert CK, Laurent P, Ma S, Paddock L, Miranda IC, et al. Epigenetic memory of coronavirus infection in innate immune cells and their progenitors. Cell. 2023;186(18):3882-3902 e3824.
37. Kamitaki N, Sekar A, Handsaker RE, de Rivera H, Tooley K, Morris DL, et al. Complement genes contribute sex-biased vulnerability in diverse disorders. Nature. 2020;582(7813):577–81.
38. Zanini G, Selleri V, Roncati L, Coppi F, Nasi M, Farinetti A, et al. Vascular "long COVID": a new vessel disease? Angiology. 2024;75(1):8–14.
39. Lei Y, Zhang J, Schiavon CR, He M, Chen L, Shen H, et al. SARS-CoV-2 spike protein impairs endothelial function via downregulation of ACE 2. Circ Res. 2021;128(9):1323–6.
40. Varga Z, Flammer AJ, Steiger P, Haberecker M, Andermatt R, Zinkernagel AS, et al. Endothelial cell infection and endotheliitis in COVID-19. Lancet. 2020;395(10234):1417–8.
41. Kuchler T, Gunthner R, Ribeiro A, Hausinger R, Streese L, Wohnl A, et al. Persistent endothelial dysfunction in post-COVID-19 syndrome and its associations with symptom severity and chronic inflammation. Angiogenesis. 2023;26(4):547–63.
42. Carmona-Cervello M, Leon-Gomez BB, Dacosta-Aguayo R, Lamonja-Vicente N, Montero-Alia P, Molist G, et al. Long COVID: cognitive, balance, and retina manifestations. Front Med. 2024;11:1399145.
43. Shweiki D, Itin A, Neufeld G, Gitay-Goren H, Keshet E. Patterns of expression of vascular endothelial growth factor (VEGF) and VEGF receptors in mice suggest a role in hormonally regulated angiogenesis. J Clin Invest. 1993;91(5):2235–43.
44. Mueller MD, Vigne JL, Minchenko A, Lebovic DI, Leitman DC, Taylor RN. Regulation of vascular endothelial growth factor (VEGF) gene transcription by estrogen receptors alpha and beta. Proc Natl Acad Sci U S A. 2000;97(20):10972–7.

45. Li T, Thoen ZE, Applebaum JM, Khalil RA. Menopause-related changes in vascular signaling by sex hormones. J Pharmacol Exp Ther. 2025;392(4):103526.
46. Lowery AJ, Sweeney KJ, Molloy AP, Hennessy E, Curran C, Kerin MJ. The effect of menopause and hysterectomy on systemic vascular endothelial growth factor in women undergoing surgery for breast cancer. BMC Cancer. 2008;8:279.
47. Sargent KM, Lu N, Clopton DT, Pohlmeier WE, Brauer VM, Ferrara N, et al. Loss of vascular endothelial growth factor A (VEGFA) isoforms in granulosa cells using pDmrt-1-Cre or Amhr2-Cre reduces fertility by arresting follicular development and by reducing litter size in female mice. PLoS ONE. 2015;10(2):e0116332.
48. Carmeliet P, Jain RK. Molecular mechanisms and clinical applications of angiogenesis. Nature. 2011;473(7347):298–307.
49. Ferrara N. VEGF as a therapeutic target in cancer. Oncology. 2005;69(Suppl 3):11–6.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.