International Journal of
Molecular Sciences

MDPI

Article

A Unique Patient Stratification Method Combined with a Machine Learning Approach Identifies Novel Genetic Susceptibility and Protective Factors for Severe COVID-19 in a Hungarian Population

Alexandra Neller [1], Mátyás Bukva [2], Bence Gálik [3], József Kun [3,4], Nikoletta Nagy [1], Ferenc Somogyvári [5], Valéria Endrész [5], Margit Pál [1,*], Barbara Anna Bokor [1], Zsófia Blazovich [1], Ádám Visnyovszky [6], Balázs Bende [6], Péter Urbán [3], Szilvia Kovácsné Levang [7], Zoltán Péterfi [7], Gábor L. Kovács [3,8,9], Katalin Gombos [3,8,9], Attila Gyenesei [3,†] and Márta Széll [1,†]

1 Department of Medical Genetics, University of Szeged, 6720 Szeged, Hungary; neller.alexandra@o365.u-szeged.hu (A.N.); nagy.nikoletta@med.u-szeged.hu (N.N.)
2 Biological Research Centre Szeged, 6726 Szeged, Hungary
3 Hungarian Centre for Genomics and Bioinformatics, Szentágothai Research Centre, University of Pécs, 7624 Pécs, Hungary
4 Department of Pharmacology and Pharmacotherapy, Medical School, University of Pécs, 7624 Pécs, Hungary
5 Department of Medical Microbiology, University of Szeged, 6725 Szeged, Hungary; endresz.valeria@med.u-szeged.hu (V.E.)
6 Department of Dermatology and Allergology, University of Szeged, 6720 Szeged, Hungary
7 Clinical Centre, First Department of Internal Medicine, University of Pécs, 6724 Pécs, Hungary
8 Department of Laboratory Medicine, Medical School, University of Pécs, 7624 Pécs, Hungary
9 Hungarian National Laboratory on Reproduction, University of Pécs, 7624 Pécs, Hungary
* Correspondence: pal.margit@med.u-szeged.hu
† These authors contributed equally to this work.

Abstract

Intensive research has shown that severe COVID-19 outcomes are influenced by antiviral pathways and immune responses, both shaped by genetic predisposition. In this study, we aimed to identify genetic variants associated with disease severity in a cohort of Hungarian patients. We applied a novel stratification method based on age, disease severity, and clinical background to classify patients by susceptibility to severe COVID-19. Whole-exome sequencing (WES) was performed on 168 individuals, and gene mutation loads were assessed. Using a Random Forest machine learning approach, we identified variants of 877 genes that distinguished between severe and non-severe cases. We further categorized these genes as either susceptibility or protective factors. Gene-set enrichment analysis highlighted the most affected biological pathways. Our findings support the development of personalized diagnostic tools to assess the risk of severe COVID-19 and guide targeted treatment strategies. Our findings further extend the results of previous studies, providing novel insights into the genetic determinants of COVID-19 severity.

Keywords: COVID-19; immunogenetics; machine learning; genomics

Check for updates

Academic Editor: Mihaela Chivu-Economescu

Received: 10 January 2026
Revised: 7 February 2026
Accepted: 11 February 2026
Published: 3 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Coronavirus 2019 (COVID-19) disease develops as a consequence of infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. Since its emergence in 2019, SARS-CoV-2 has infected hundreds of millions of people worldwide and caused more

than seven million deaths as of 10 September 2025 (https://covid19.who.int/, accessed on 10 September 2025). The mortality rate for vulnerable individuals with advanced age and/or medical comorbidities has been notably high. Infected patients exhibit a wide spectrum of COVID-19 severity, ranging from asymptomatic to critically affected. The frequency of asymptomatic infections is estimated to range between 10% and 70%, depending on study design and population characteristics. The disease outcome is highly influenced by multiple factors, such as age, gender, social status, and geographic location [2]. Among symptomatic cases, ~40% present with mild upper respiratory tract infection, ~20% with pneumonia, and ~3% progress to severe disease [3]. Patients of advanced age and those with major comorbidities are more frequently affected by severe and critical COVID-19 development, necessitating admission to intensive care units [4]. The most relevant comorbidities include pulmonary diseases, hypertension, diabetes, obesity, and secondary immunosuppression caused by malignant diseases or medications [5]. Despite these general observations, the social and economic differences between countries make it difficult to create a coherent system for the assessment of the risk factors for the disease. In addition to the typical scenarios described above, COVID-19 can result in other, anomalous outcomes [5]. Anomalies in mortality and morbidity have been observed both between and within populations [6].

SARS-CoV-2 strains that emerged during the pandemic exhibited variability in virulence and COVID-19 severity [7,8]. Since the beginning of the pandemic, intensive research on the host genetic makeup has identified several genes associated with COVID-19 susceptibility and severity. Genetic susceptibility to SARS-CoV-2 infection has been associated with interferon-related genes (TLR3, IFNAR1/2, and IRF9), viral entry factors (ACE2, TMPRSS2), and specific HLA haplotypes (HLA-DRB115, HLA-A30:02) [9–11]. Some researchers and groups consider the major genetic determinants of severe COVID-19 to be monogenic traits [12]; however, studies performed in broad, global collaborations found several genetic traits associated with severe COVID-19 in genome-wide association studies (GWAS) [13,14]. Intriguingly, some regions of chromosome 3 that are of Neanderthal origin are associated with severe COVID-19 [15].

Genetic and genomic studies have the potential to identify strong genetic determinants that may help healthcare providers predict the course of the disease and potentially identify therapeutic targets. Research in the last 2–3 years of the COVID-19 pandemic resulted in the identification and publication of such human host genetic determinants. There are already algorithms that can predict the outcome of the disease based on comorbidities, disease symptoms, and specific genetic haplotypes [16]; however, a clear and exact gene set and polymorphism-based prediction of COVID-19 are yet to be established.

However, the strong effect of classical clinical risk factors, particularly age and comorbidities, may obscure the contribution of host genetic variation when heterogeneous patient populations are analyzed. As a result, genetic signals relevant to disease severity can remain undetected if traditional stratification strategies are applied.

The clinical course of COVID-19 is shaped by a complex interplay of demographic, clinical, and genetic factors. Established evidence indicates that age and comorbidities, such as cardiovascular disease, diabetes, or chronic pulmonary disorders, are among the strongest predictors of severe disease outcomes. However, these well-recognized determinants may mask the potential contribution of host genetic variation when analyzed in heterogeneous patient populations [4,5].

Rather than only stratifying by age and comorbidities—which are established predictors of COVID-19 severity—our approach focused on creating two contrasting groups that reflect the paradoxical nature of the disease. While younger, otherwise low-risk patients occasionally developed severe or even fatal COVID-19 [17,18], some elderly and

multimorbid patients remained asymptomatic or experienced only mild disease [19–21]. These polar opposites formed the basis of our focus cohorts, where we hypothesized that genetic determinants underlie these unexpected outcomes. Specifically, we expected genetic factors to predispose younger patients to severe disease, while in elderly patients with unexpectedly mild disease, genetic factors were assumed to contribute to protection. To test this hypothesis, we compared the focus cohorts, in which a genetic contribution was presumed, with control cohorts in which disease outcome could be attributed primarily to age and comorbidities, consistent with classical epidemiological expectations. Although similar paradoxical patterns had been described in previous influenza outbreaks, where severe disease occasionally occurred in otherwise healthy young adults [22–24], and some elderly remained asymptomatic [25], these extremes became particularly prominent in COVID-19. We therefore propose that this stratification framework, originally highlighted by COVID-19, could also serve as a useful methodological approach for studying host determinants of severity in other viral infections.

From a clinical utility perspective, genetic testing is most effective when based on a limited number of carefully selected variants that can be assessed rapidly and provide meaningful predictive value. While COVID-19 is a complex, polygenic disease, our aim was not to capture the entirety of its genetic architecture but rather to establish a focused set of determinants that may be practical for diagnostic application. To this end, we combined a stringent patient stratification method with a machine learning approach to identify a concise and clinically useful gene panel.

Here, we report a newly identified set of genetic factors that are associated with the severity of COVID-19. During the study, we developed and applied a COVID-19-specific patient stratification method for Hungarian patients. These factors were identified with whole-exome sequencing (WES) and calculation of mutation load for each gene. A machine learning approach with a Random Forest algorithm was used to define the gene set that clearly distinguishes patients with respect to COVID-19 severity.

2. Results

2.1. Identification of Susceptibility and Protective Genetic Factors Responsible for COVID-19 Severity in Stratified Patient Cohorts

We conducted a WES analysis of the stratified patient cohorts using standard bioinformatics techniques, as described in the Materials and Methods section. Data were filtered, and single-nucleotide polymorphisms (SNPs) were summarized, resulting in a variant dataset of 20,048 genes (Supplementary Table S1), prior to applying machine learning.

For machine learning, we followed the workflow shown in Figure 1 to identify genes distinguishing the two focus groups, YFC (Young Focus Cohort, n = 38) and OFC (Old Focus Cohort, n = 34), thereby revealing host genetic factors underlying COVID-19 severity (Figure 1A).

We ran an information gain algorithm to score the genes on 80% of the data (Figure 1A, Steps 1–2). The average information gain was found to be 0.03. Genes above the mean were selected, resulting in a panel of 877 genes (Supplementary Table S2).

Subsequently, we applied the Random Forest algorithm with 5-fold cross-validation to 80% of the data, using the 877 selected genes. The model generated by the Random Forest algorithm was able to distinguish the remaining YFC and OFC data, with an average classification accuracy of 89.20% (sensitivity = 88.80%, specificity = 89.60%) (Figure 1A, Step 4).

Next, the datasets from the focus groups (YFC, OFC) were compared with their corresponding control groups, YCC (Young Control Cohort, n = 31) and OCC (Old Control Cohort, n = 49). Classification results show that, based on the mutation patterns of the

877 genes, the YFC and YCC could be distinguished with an average classification accuracy of 84.10%, corresponding to a sensitivity of 83.80% and a specificity of 84.40% (Figure 1A).

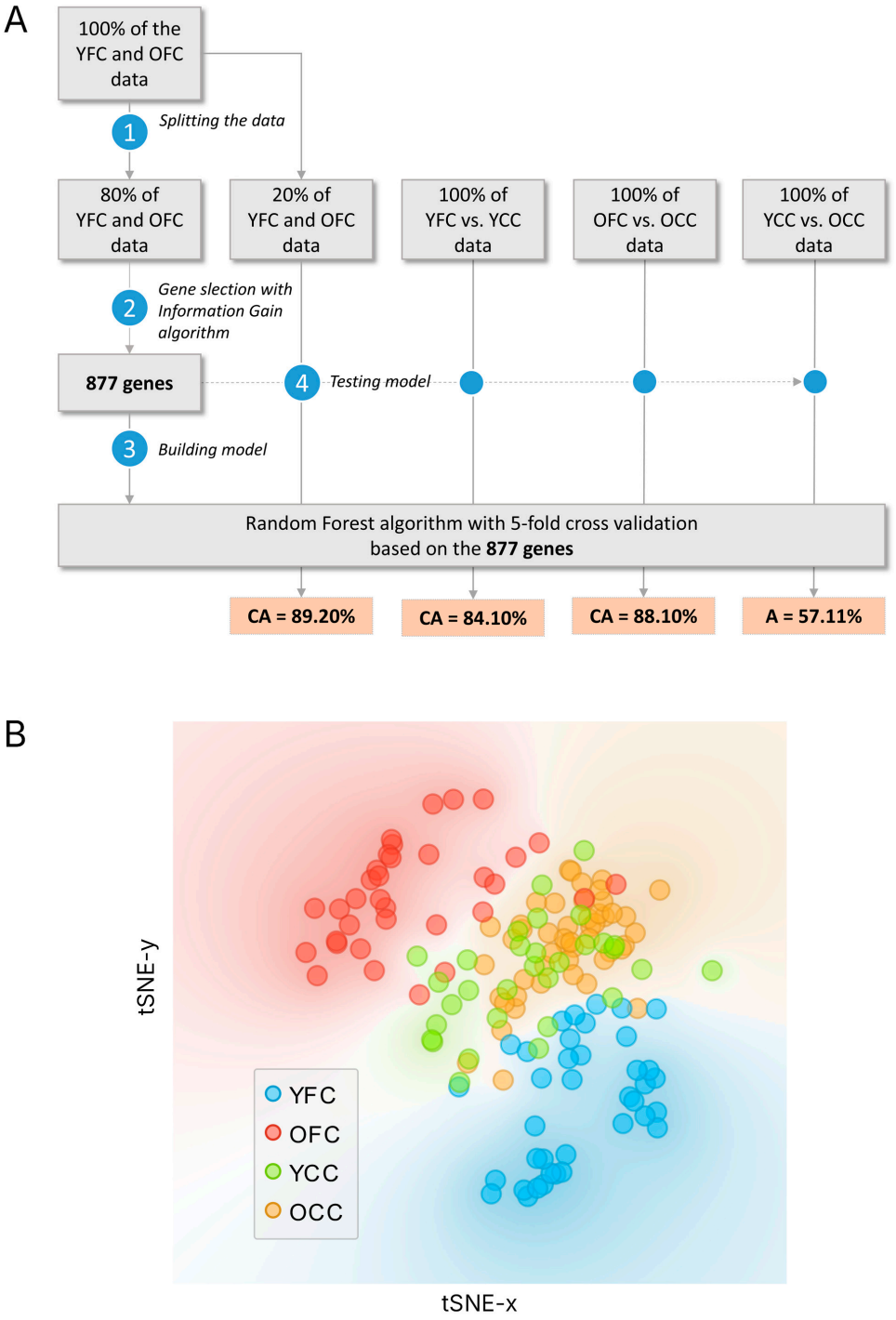


Figure 1. Candidate gene selection and differentiation among patient groups: (**A**) Workflow of the classification analysis. The YFC and OFC datasets were split into 80% training and 20% testing subsets. Gene selection using the information gain algorithm identified 877 genes, which were used to build a Random Forest model with 5-fold cross-validation. The model achieved a classification accuracy (CA) of 89.20% for YFC vs. OFC data. The same gene set was subsequently applied to classify YFC vs. YCC (CA = 84.10%, sensitivity = 83.80%, specificity = 84.40%), OFC vs. OCC (CA = 88.10%, sensitivity = 87.90%, specificity = 88.30%), and YCC vs. OCC (A = 57.11%, sensitivity = 56.80%, specificity = 57.40%). (**B**) t-distributed Stochastic Neighbor Embedding (t-SNE) visualization of the four patient cohorts. Note how YFC and OFC samples form distinct clusters, while control groups (YCC and OCC) overlap partially but remain separated from their corresponding focus cohorts.

Similarly, the OFC and OCC could be separated with an accuracy of 88.10%, with a sensitivity of 87.90% and a specificity of 88.30% (Figure 1A)

In contrast, when comparing the YCC and OCC directly, the classification performance dropped to 57.11%, with a sensitivity of 56.80% and a specificity of 57.40%, indicating a substantial similarity between the two control datasets (Figure 1A).

As the final step of the analysis workflow, the dataset for the two control groups—YCC and OCC—was analyzed using the Random Forest algorithm, which employed the genetic patterns of the 877 identified genes. The method was unable to distinguish between the two control groups. This is consistent with our a priori hypothesis: differences in COVID-19 severity between the two control groups are caused mainly by classical determinants, such as age and the presence of comorbidities, not by underlying genetic factors (Figure 1A).

The final Random Forest results are shown in Figure 1B, illustrating the separation of the stratified patient cohorts (highlighted in different colors) based on the variant pattern of the 877 genes. For clarity, in the subsequent sections, we refer to genes whose variant patterns are associated with an increased likelihood of severe COVID-19 as "susceptibility genes." In contrast, those associated with a reduced likelihood are termed "protective genes." These designations describe statistical associations with disease outcomes rather than implying an intrinsically harmful or beneficial function of the genes themselves.

Given that YFC and OFC represent polarized severity strata, we evaluated potential confounding to ensure that the gene-based signal predicting differences between these groups reflects disease course rather than demographic or clinical covariates. Specifically, we tested whether age, sex, and major comorbidities are related to the genetic summaries and model behavior.

SNP burden was summarized per subject from the 877-gene panel and correlated with clinical variables (age, sex, major comorbidities). Across 13,155 covariate–SNP-burden pairs, only 4/13,155 associations remained significant after Benjamini–Hochberg correction (FDR $q < 0.05$); all corresponding effect sizes were small ($|\rho| \leq 0.50$), and directions were inconsistent. The remaining 13,151/13,155 comparisons were non-significant at $q \geq 0.05$, indicating that SNP burden is not meaningfully explained by demographic or comorbidity structure (Supplementary Table S3).

To assess potential effects on model behavior, we also correlated these clinical variables with per-sample classification error. Associations were again negligible ($|\rho| = -15$–0.14; all $p = 0.279$–0.765; Figure 2), supporting that the classification performance reflects genetic variation rather than demographic or comorbidity profiles.

2.2. Genetic-Variant Landscape of the Susceptibility and Protective Genes

Within the 877 genes considered, we classified loci as susceptibility or protective by comparing mean variant counts between the two focus cohorts (YFC vs. OFC). Using this criterion, 431 genes were labeled as susceptibility genes (YFC > OFC), and 446 genes were labeled as protective (OFC $\geq$ YFC).

When contrasting focus groups to their age-matched controls, 246/431 susceptibility genes (57.08%) showed a higher average SNP count in OCC than in OFC, whereas 419/446 protective genes (93.95%) had a higher average SNP frequency in OFC than in OCC. In the younger cohorts, 369/431 susceptibility genes (85.61%) had higher average SNP counts in YFC than in YCC, while 302/446 protective genes (67.71%) had lower average SNP counts in YFC than in YCC.

For each individual, we computed the mean SNP count across susceptibility genes and across protective genes and formed their ratio. The resulting group means were 0.65 (OFC) and 1.33 (YFC) in the focus cohorts, and 0.97 (YCC) and 0.93 (OCC) in the controls. One-way ANOVA on these ratios revealed a strong group effect ($p < 0.0001$).

Tukey *post hoc* tests indicated no significant difference between the two control groups (YCC vs. OCC: $\Delta = -0.04$, $p = 0.121$), while both focus groups differed markedly from their corresponding controls (OFC vs. OCC: $\Delta = 0.27$, $p < 0.0001$; YFC vs. YCC: $\Delta = -0.36$, $p < 0.0001$) (Figure 3).

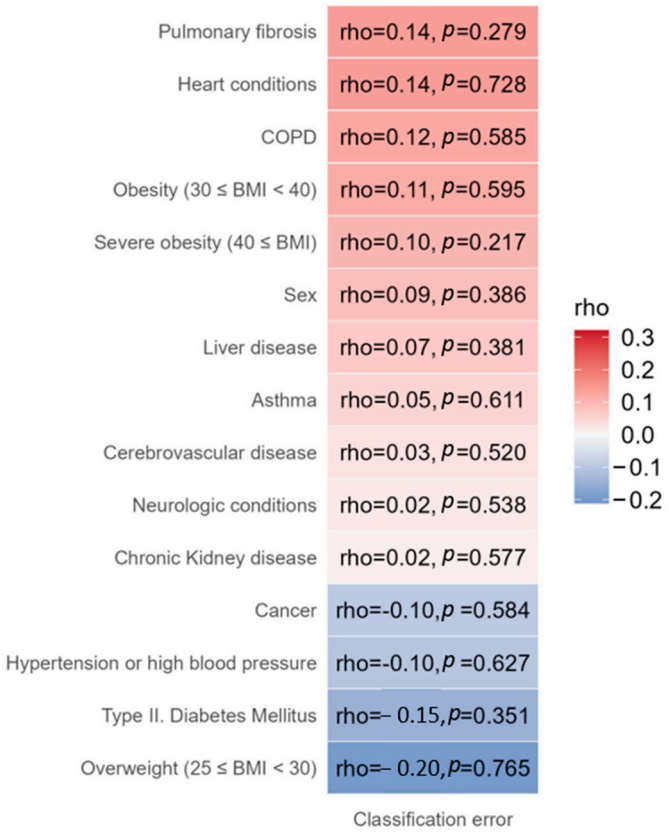


Figure 2. Correlations between clinical variables and per-sample classification error. Spearman's rank correlation coefficients (rho) and corresponding *p*-values are shown for each clinical variable. All associations were weak and statistically non-significant ($|\rho| = 0.02$–0.20; all $p = 0.217$–0.765), indicating that model performance was not influenced by demographic or comorbidity-related factors. The color scale represents the direction and strength of correlations (red: positive; blue: negative).

These patterns confirm that the directionality imposed by the gene categorization (YFC > OFC ⇒ susceptibility; otherwise protective) is preserved at the individual level across groups, i.e., the subject-level susceptibility-to-protective SNP ratio follows the same directionality.

2.3. Identification of Biological Processes Affected by Susceptibility and Protective Genes

Gene Ontology (GO) enrichment analysis of the 877 genes highlighted several significant categories (reported here as *p* values). In molecular function (MF), enrichment was observed for protein binding (GO:0005515, $p = 0.007$) and cation binding (GO:0043169, $p = 0.02$). Within the biological process (BP) category, biological regulation (GO:0065007, $p = 0.005$) was enriched. For cellular component (CC), significant terms included cytoplasm (GO:0005737, $p = 0.007$) and lateral element (GO:0000800, $p = 0.04$). Detailed results: https://biit.cs.ut.ee/gplink/l/aio83Uc8kSI (accessed on 25 September 2025).

Using the Metascape signature module (Immunological Signatures subset), the enrichment analysis of 877 genes showed a significant overrepresentation of gene sets related to macrophage activation, interferon-stimulated conditions, Toll-like receptor ligand stimulation, and T cell activation and differentiation experiments. The enriched signatures included datasets on STAT6-knockout macrophages, IFNG- and IL6-perturbed

systems, and $CD4^+/CD8^+$ T cell polarization states. Complete results are provided in Supplementary Document S1.

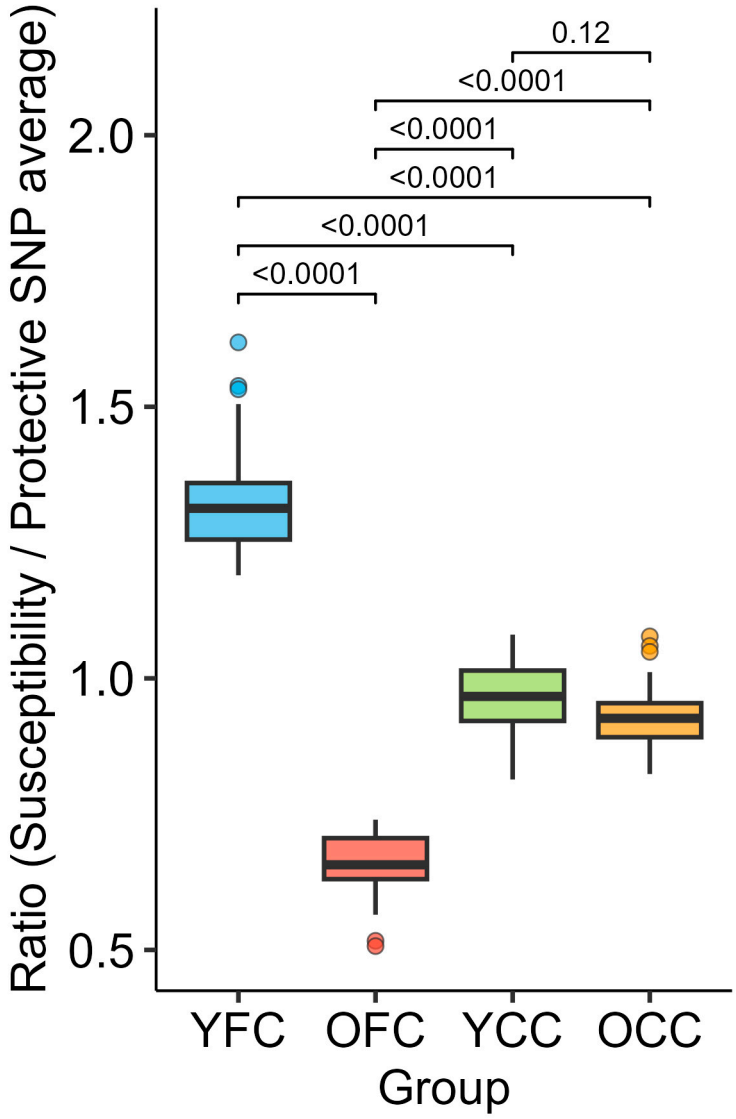


Figure 3. Group-wise susceptibility-to-protective SNP ratio (boxplots; one-way ANOVA $p < 0.001$). Across 877 genes, 431 were classified as susceptibility (YFC > OFC) and 446 as protective (OFC $\geq$ YFC). Among susceptibility genes, 246/431 (57.08%) show higher average SNP counts in OCC than OFC, and 369/431 (85.61%) are higher in YFC than YCC. Among protective genes, 419/446 (93.95%) are higher in OFC than OCC, and 302/446 (67.71%) are lower in YFC than YCC. Mean ratios: OFC 0.65, YFC 1.33, YCC 0.97, OCC 0.93—confirming that the gene-categorization directionality is preserved at the individual level.

2.4. Determining the Minimal Discriminating Gene Set

To enhance the practical applicability of our findings, we next sought to reduce the set of 877 genes to achieve an optimal balance between panel size and predictive efficiency.

The genes were ranked in descending order based on their information gain values, reflecting their relative contribution to the classification of YFC and OFC.

Subsequently, we incrementally expanded the gene panel by sequentially adding genes according to their ranked importance and evaluated the resulting models to identify the smallest subset that retained maximal classification accuracy (Figure 4).

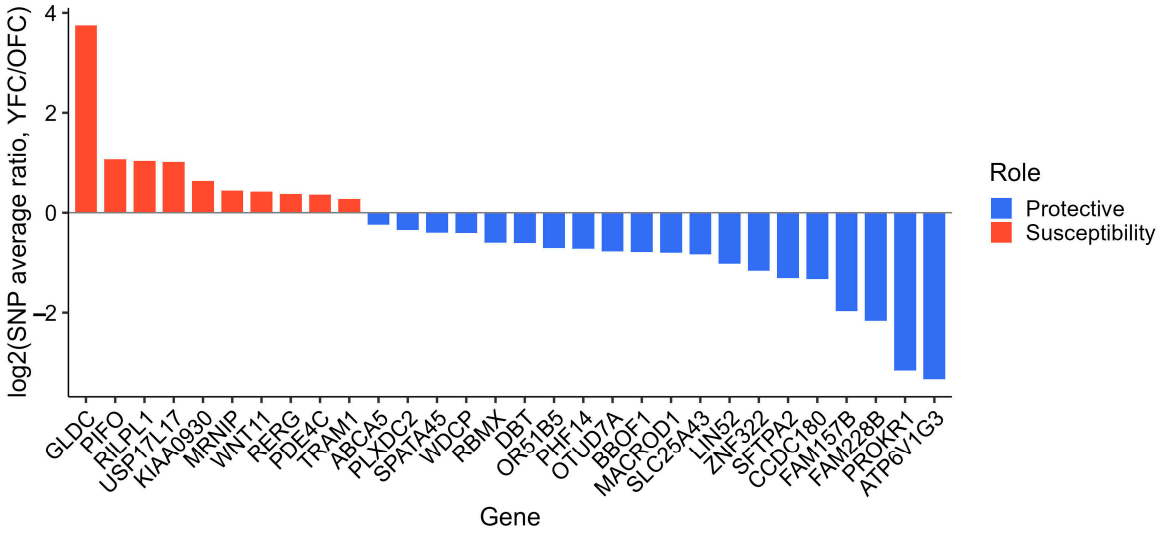


Figure 4. The 30 genes exerting the strongest association with disease severity. Log_2 (severe/mild) SNP ratios for the top genes, showing susceptibility (red) and protective (blue) associations.

As the number of included genes increased, classification performance initially improved, reaching a peak accuracy of 80.60% with the top 30 most significant genes (Supplementary Table S2).

Beyond this point, the inclusion of additional genes resulted in only minimal further improvement until the final 877-gene panel achieved the maximum accuracy observed in the model (Figure 5).

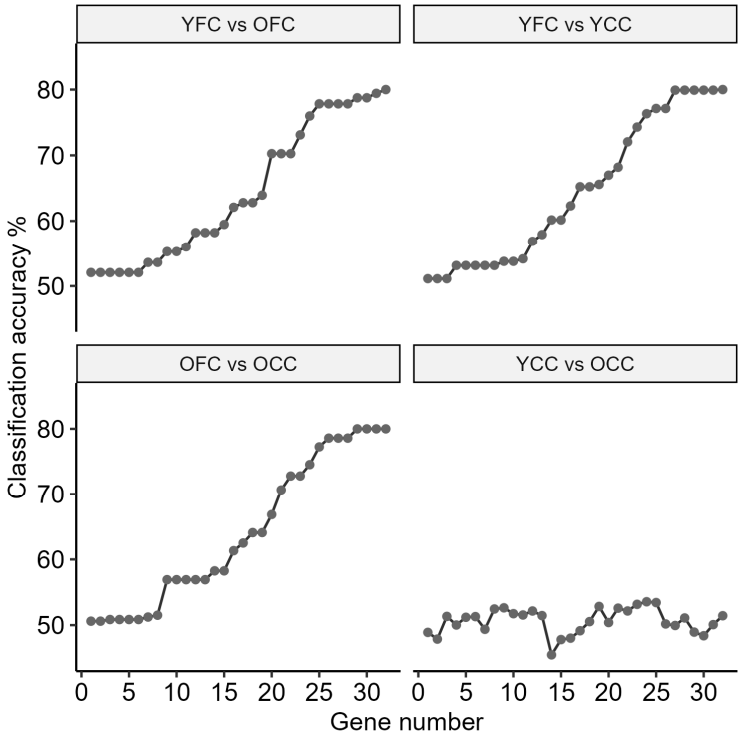


Figure 5. Classification accuracy as a function of the number of top-ranked genes. Each panel shows one pairwise comparison (YFC vs. OFC, YFC vs. YCC, OFC vs. OCC, and YCC vs. OCC). For each comparison, genes were ranked by information gain on the training split, and a Random Forest classifier was trained using the top N genes (N = 1–30). Points denote the mean cross-validated accuracy from 5-fold CV; lines connect points as a visual guide. Curves for YFC–OFC, YFC–YCC, and OFC–OCC rise toward ~80% as N increases, whereas YCC–OCC fluctuates around ~50% (chance level), indicating negligible separability of the two control cohorts.

Notably, this 30-gene panel achieved an average efficiency of 90% across all comparisons. To test the predictive value of this minimal set of genes, we analyzed the control groups (YCC and OCC) and found that they did not differ from one another.

3. Discussion

In this study, we applied whole-exome sequencing, combined with machine learning-based feature selection, to identify genetic variants associated with COVID-19 severity in a clinically stratified Hungarian cohort. By leveraging a polarized cohort design that contrasts individuals with discordant clinical risk profiles and disease outcomes, we derived a gene panel that reflects both susceptibility- and protection-associated signals in this population. This strategy enabled the detection of variant patterns that may remain undetected in conventional severity comparisons relying solely on clinical stratification.

Several genes within the identified panel are functionally linked to biological processes previously implicated in host responses to viral infection, including immune regulation, cellular signaling, and metabolic pathways [26–63]. While these associations support the biological plausibility of the model output, it is important to emphasize that the present findings reflect statistical prioritization derived from enrichment and machine learning analyses and should not be interpreted as direct evidence of causal molecular mechanisms.

Our results show that combining a cohort design based on clinical information with exome-level variant analysis can identify candidate genetic signatures linked to disease severity. This creates a basis for future functional and translational studies. Notably, comparison with ancestry-specific severe-versus-mild analyses from the COVID-19 Host Genetics Initiative revealed no direct gene overlap with the core 30-gene panel [14,26]. This lack of concordance is not unexpected, as the final panel represents a highly constrained subset selected to optimize classification performance within our cohort, and therefore reflects a parsimonious predictive signature rather than an exhaustive catalog of disease-associated loci. In contrast, comparison at the broader feature-selection level (n = 877 discriminating genes) demonstrated direct overlap with several HGI-reported genes, including THBS3, FBRSL1, KANSL1, DPP9, TYK2, IFNAR2, ELF5, and SLC22A31. This pattern indicates that the analytical framework captures established host genetic signals at the broader feature-space level. In contrast, the final reduced panel prioritizes a compact subset that maximizes predictive performance within the studied cohort. Several overlapping genes converge on interferon-driven antiviral signaling and cytokine regulation, including TYK2 and IFNAR2, which directly participate in JAK–STAT-mediated immune responses [27], as well as regulators of transcriptional and inflammasome-associated signaling such as KANSL1 and DPP9 [28–31]. These links suggest a convergence in host antiviral response-modulation pathways. A second functional theme concerns epithelial–extracellular interface biology, represented by THBS3 and ELF5, which are linked to tissue organization and epithelial regulation [32–35], together with transport and epigenetic regulators such as SLC22A31 and FBRSL1 [36,37]. These associations may reflect contributions of host-tissue context to viral susceptibility and the inflammatory response.

Gene set enrichment analysis using immune signature collections revealed significant associations with transcriptional programs related to macrophage activation, interferon signaling, Toll-like receptor-mediated innate sensing, and T cell functional polarization. These processes are consistent with established host-response pathways during viral infection, where pattern-recognition receptor activation induces cytokine production and immune modulation, and interferon signaling orchestrates antiviral defense and adaptive immune coordination [38,39]. Importantly, interferon-driven signaling has been shown to regulate macrophage inflammatory responses [40] and T cell functionality during viral disease [41]. At the same time, excessive activation of innate sensing pathways may contribute

to inflammatory dysregulation observed in severe COVID-19 [42,43]. Accordingly, these enrichment patterns support the biological plausibility of the identified gene set while remaining hypothesis-generating rather than indicative of causal mechanisms.

A general gene set enrichment analysis of the full discriminating gene set identified functional themes related to molecular interaction networks, metal-dependent processes, intracellular signaling environments, and epithelial organization. Instead of pointing out specific mechanisms, these enrichments offer context for biological processes that have been previously connected to host–virus interaction dynamics [44–63]. Enrichment of the protein and small-molecule binding categories highlights the role of molecular interaction networks in viral infections. Host protein–protein interfaces and cofactor-dependent processes play a part in immune signaling and the response to pathogens [44–48]. The representation of cation-binding functions is consistent with the involvement of metal-dependent enzymatic and regulatory pathways in cellular redox balance and antiviral defense, although these associations remain interpretive [49–51]. Intracellular localization signatures, especially those of cytoplasmic components, show the cellular areas where antiviral sensing, mitochondrial stress responses, and viral replication occur [52–56]. In parallel, enrichment for epithelial structural elements supports the relevance of tissue interface biology, including mucociliary and barrier-associated functions that are frequently implicated in respiratory disease severity [57–59]. Finally, broad regulatory process categories indicate that prioritized genes span diverse roles in transcriptional, metabolic, and cell cycle regulation [60–63]. Taken together, these pathway-level observations suggest functional convergence across interaction networks, cellular regulation, and tissue interface processes; however, they should be interpreted as hypothesis-generating annotations rather than evidence of causal mechanistic involvement.

Monogenic [64,65] and polygenic-focused [13,66–68] genetic studies provide complementary approaches to understanding the role of host genetics in COVID-19 susceptibility and severity. While the two approaches differ in methodology and scope, they can provide valuable insights when combined. Broad analyses of genetic data can help identify common genetic variants influencing COVID-19 outcomes across populations. In contrast, research on rare genetic mutations provides specific insights into how individual immune responses affect disease severity.

Lai et al. (2022) and Zheng et al. (2024) both used transcriptomic data and bioinformatics approaches to study COVID-19, with Lai et al. constructing a three-gene diagnostic signature and Zheng et al. developing an immune cell infiltration (ICI) score to classify immune subtypes and predict severity [69,70]. While such transcriptomic-based tools capture the dynamic immune landscape and reflect disease severity during infection, our study takes a fundamentally different approach by focusing on underlying genetic predisposition. Using WES in a well-defined Hungarian cohort and a stratification method based on age, background, and disease severity, we combined mutational load analysis with a Random Forest model to identify genetic variants associated with protection against or susceptibility to severe COVID-19. In contrast to expression-based methods, our SNP-based severity panel uncovers host-level genomic determinants and can identify individuals at risk before or early in infection. Together, these complementary approaches could provide a multi-layered framework for risk prediction and patient stratification.

The present findings provide a foundation for several ongoing and planned research directions aimed at refining biological interpretation and translational relevance. Future work will focus on the functional characterization of prioritized genes through gene-expression profiling, in vitro assays, and targeted mechanistic studies, including pathway-focused perturbation approaches.

Additional validation efforts will assess the strength and applicability of the predictive framework in independent cohorts. This will help determine how the proposed gene panel can be used for clinical classification. We also plan to conduct comparative analyses across different respiratory diseases, such as influenza. This will allow us to explore whether the identified signals are specific to COVID-19 or reflect broader patterns of host susceptibility. We are currently conducting complementary transcriptomic profiling of samples from the same cohort. This enables combined genomic and expression analyses, which may reveal downstream regulatory patterns and guide subsequent experimental validation. Together, these efforts aim to advance the interpretation of statistically prioritized signals toward biologically grounded and clinically informative applications.

Strengths and Limitations

This pilot study employs WES and machine learning to examine genetic variants associated with COVID-19 severity in a Central-Eastern European (Hungarian) cohort, which is largely underrepresented in current genomic research. A major strength of our approach is the clinical–genetic stratification of patients. This offers a useful framework for future precision risk assessments. Using machine learning helped us identify candidate genes that may predict disease severity. This shows the potential of data-driven methods in infectious disease genomics. Our methodology could serve as a proof of concept for including genomic data in clinical decision-making processes. Nonetheless, several limitations should be acknowledged. The modest sample size and the single-population design may limit the generalizability of our findings, although the inclusion of an underrepresented ancestry also enhances the diversity of existing datasets. Functional validation of the implicated genes was beyond the scope of this study; therefore, the reported associations should not be interpreted as direct evidence of causal biological mechanisms. Further work will be required to confirm these findings through gene expression analyses and mechanistic studies.

Furthermore, while certain well-known immune-related genes, such as those in the interferon signaling pathway, were not prioritized by the model, their biological relevance cannot be excluded. It will be further evaluated in future panel iterations.

Despite cross-validation, using a single dataset for both feature selection and model training introduces the risk of overfitting and may overestimate predictive performance. This highlights the importance of validating the proposed gene panel and machine learning framework in independent external cohorts. Future validation of the proposed 30-gene panel should involve its application to independent external cohorts using the same stratification criteria and machine learning framework, followed by an evaluation of predictive performance and reproducibility across different populations.

Overall, the study provides an important first step toward establishing a genetically informed framework for COVID-19 severity stratification and lays the groundwork for expanded, multi-ethnic validation and functional follow-up.

4. Materials and Methods

4.1. Patient Recruitment

This study was conducted at the Albert Szent-Györgyi Clinical Centre of the University of Szeged and the University of Pécs, First Department of Internal Medicine. From March 2021 to December 2022, after applying strict and unique enrollment criteria (described in detail in Materials and Methods, Section 4.3) a total of 700 patients were enrolled in the study. For the prospective part of the study, patients over 18 years of age with positive COVID-19 PCR tests, which were not older than 48 h, were recruited. COVID-19 positivity was confirmed using the SARS-CoV-2 real-time PCR assay on nasopharyngeal

samples. After the patient signed the informed consent sheet, relevant patient data were collected, including medical history and demographic data. Biological samples (peripheral blood and nasopharyngeal swabs) were collected according to the ethical approval (19697-6/2020/EÜIG) issued by the National Center for Public Health, Hungary.

4.2. Development of a Unique Patient Stratification Method

The disease severity of the patients was determined based on the instructions in the Hungarian COVID-19 manual [71]. Patients were classified into five groups (asymptomatic, mild, moderate, severe, and critical COVID-19) according to the most severe stage of their disease progression.

For the basis of our genetic study, we developed a unique and stringent stratification method. Age was one of the bases of our distinction between the cohorts. According to the calculations by the Centers for Disease Control and Prevention (CDC, cdc.gov) and as reported in the international literature, the course of COVID-19 was more severe in patients over 65 years of age than in younger individuals [72]. Thus, we chose this age as a cut-off for specifying patient groups. Other distinguishing aspects were clinical risk factors for severe COVID-19 identified in epidemiological studies [72–94]. We defined four patient groups based on age, COVID-19 severity, and the presence of major and minor risk factors. Patients were assigned to one of four pre-defined cohorts immediately after their enrollment. The definition of these patient groups ensured homogeneity within groups and a clear separation between the focus and control cohorts (genetic determinants of severity vs. environmental determinants of severity).

For this separation, clinical data were collected from all enrolled patients based on the risk stratification criteria issued by the CDC at the beginning of the COVID-19 pandemic. Data collection initially encompassed major and minor clinical risk factors for severe COVID-19, including chronic cardiovascular, pulmonary, renal, metabolic, oncological, and neurological conditions, obesity-related categories, smoking status, pregnancy, and immunocompromised states.

As evidence accumulated, the stratification framework was iteratively refined to reflect emerging, new epidemiological data. Patients with thalassemia and sickle cell disease were excluded, as their independent associations with COVID-19 severity remained inconclusive. Individuals with immunocompromised states were also excluded to avoid confounding effects on disease severity and immune response. In addition, risk factors with no representation in the cohort—namely Down syndrome, pregnancy, cystic fibrosis, and type 1 diabetes mellitus—were removed from the scoring system. These refinements resulted in a streamlined and unique stratification approach suitable for genetic analyses. The complete list of risk factors and a detailed table of the reasoning behind patient cohorts can be found in Supplementary Document S2 [95].

4.3. Cohort Definitions

Based on the predefined stratification framework, we specified four distinct patient cohorts, which allowed us to map the contribution of genetic versus clinical factors to COVID-19 outcomes. The young focus cohort (YFC, $n = 38$) comprised patients under 65 years of age who developed severe or critical COVID-19 (severity categories 4–5) despite having few or no clinical risk factors. In contrast, the old focus cohort (OFC, $n = 34$) included patients over 65 years presenting with only mild-to-moderate symptoms (severity categories 1–2), in the presence of multiple comorbidities. These two groups were designed to highlight outlier cases where disease severity could not be explained by age or comorbidities alone.

For comparison, we also defined two control cohorts, representing more typical clinical outcomes. The young control cohort (YCC, $n = 31$) consisted of patients under 65 years with mild COVID-19 (severity categories 1–2) and minimal comorbidities, while the old control cohort (OCC, $n = 49$) included older patients (≥65 years) with severe disease (severity categories 4–5) and a high burden of risk factors. Together, these four cohorts provided a structured framework to contrast genetic susceptibility with severity, which contributes to well-known clinical determinants. In Table 1, we provide a general description of our patient cohorts.

Table 1. Descriptive summary of the patient cohorts.

Cohort	N	Mean Age	Median Age	SD Age	Male (N)	Female (N)	Mean Severity	Median Severity
OCC	49	75.184	75	7.126	28	19	4.326	4
OFC	34	75.735	74	7.805	12	21	1.853	2
YCC	31	49.774	53	10.698	14	17	1.806	2
YFC	38	54.131	56	8.302	25	12	4.131	4

The prevalence of each risk factor in every cohort can be seen in Figure 6.

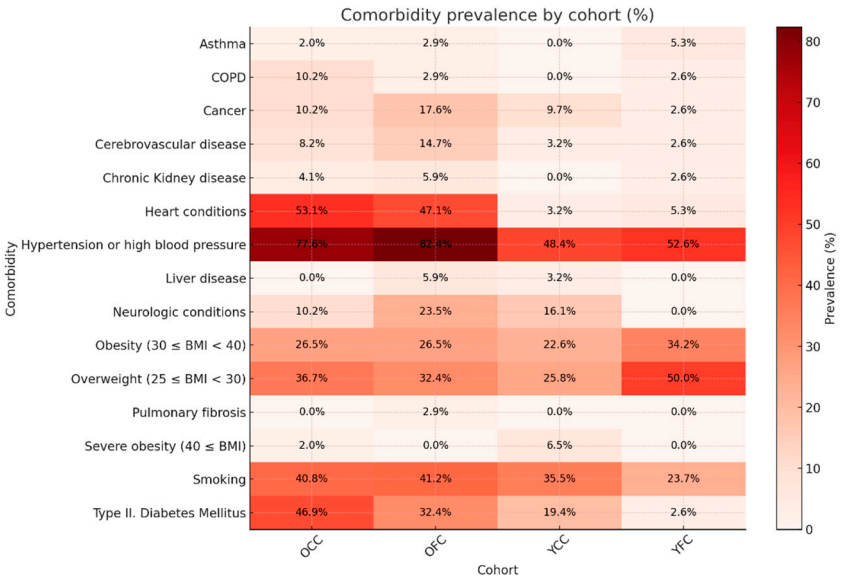


Figure 6. Heatmap representation of risk factor prevalence per study cohort. The heatmap shows the percentage of patients affected by each comorbidity within the four cohorts. Warmer colors indicate a higher prevalence. As expected, comorbidities were more frequent in the older cohorts (OFC, OCC), while the young cohorts (YFC, YCC) displayed markedly lower rates.

We tested whether our a priori patient stratification strategy resulted in comparable patient cohorts. The distribution of risk factors per cohort can be found in Figure 7.

To assess whether the patient cohorts differed in overall comorbidity burden, we compared the number of comorbidities per patient across groups. The Kruskal–Wallis test indicated a significant difference between cohorts ($\chi^2(3) = 36.42$, $p < 0.001$). Post hoc pairwise Mann–Whitney U tests with Bonferroni correction revealed that the young cohorts (YFC, YCC) did not differ from each other, and likewise, no difference was detected between the two older cohorts (OFC, OCC). In contrast, all comparisons between young versus older cohorts were significant ($p < 0.001$), confirming that the stratification effectively separated patients with low versus high comorbidity burden.

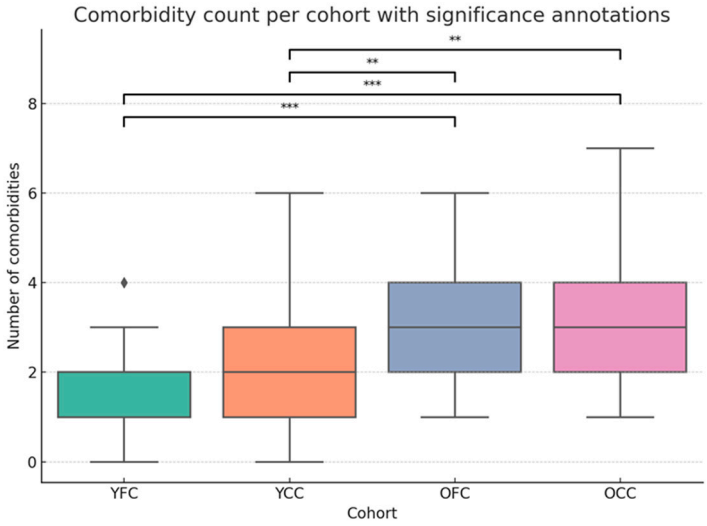


Figure 7. Comorbidity counts across patient cohorts. Boxplots show the distribution of comorbidity counts in the young focus cohort (YFC), young control cohort (YCC), old focus cohort (OFC), and old control cohort (OCC). Older cohorts displayed significantly higher comorbidity counts than younger ones, while no difference was found between OFC and OCC or between YFC and YCC. Significance was determined by pairwise Mann–Whitney U tests with Bonferroni correction (*** $p < 0.001$, ** $p < 0.01$).

4.4. DNA Extraction

Blood samples were taken from the patients, and DNA was extracted from the frozen, EDTA-treated samples using the QIAsymphony DSP DNA Kit (QIAGEN, Hilden, Germany) with the help of the QIAsymphony SP nucleic acid purification instrument, according to the manufacturer's instructions.

4.5. Whole-Exome Sequencing

Whole-exome libraries were prepared using the DNA Prep with Enrichment workflow (Illumina, Inc., San Diego, CA, USA). Briefly, 100 ng of genomic DNA was fragmented, amplified, and purified. Exome capture was performed using Illumina Exome Panel probes (Illumina). The captured libraries were quality-checked using TapeStation 4200 (Agilent Technologies, Santa Clara, CA, USA) and Qubit 3.0 (Thermo Fisher Scientific, Waltham, MA, USA). The libraries were sequenced on NovaSeq 6000 (Illumina) for 2 × 150 paired-end reads.

4.6. Bioinformatics Analysis

Raw sequence generation, including base calling and demultiplexing, was performed using bcl2fastq (v2.20.0.422) (bcl2fastq (RRID: SCR_015058)) on a local high-performance computing (HPC) cluster. First, the quality of the raw FASTQ files was checked by FastQC (v0.11.9) [96]. Based on the quality control results, the dataset was filtered, and adapters and low-quality bases/reads were removed with fastp (v0.21.0) [97]. Clean, high-quality reads were aligned to the human reference genome (GRCh37) by applying the bwa mem algorithm with default parameters (v0.7.17) [98]. BAM file modifications (e.g., sorting, adding read group information, indexing, and deduplication) were performed using Picard Tools (v2.23.3) subcommands (http://broadinstitute.github.io/picard/ accessed on 26 September 2025). Mapping statistics and coverage metrics were calculated and collected by Picard Tools (v2.23.3) and Qualimap (v2.2.1) bamqc [99], respectively. Before variant calling, base quality scores were recalibrated by the GATK BSQR module (v4.5.0.0). Single-nucleotide variants (SNVs) and short insertions and deletions (INDELs) were identified with the GATK HaplotypeCaller algorithm by applying +/− 100 bp padding to the

target region. Individual samples were merged into a joint VCF to perform the VQSR step for SNPs and INDELs by applying GATK CombineGVCFs, VariantRecalibrator, and ApplyVQSR modules (truth sensitivity filter level 99.5 for SNPs and 99.0 for INDELs) using dbsnp156, hg38-v0-1000G_omni2.5, and hg38-v0-1000G_phase1.snps, hapmap_3.3.hg38 and Mills_and_1000G_gold_standard.indels databases (accessed on 26 September 2025). Next, raw variants were pre-filtered by the GATK VariantFiltration module [100]. For SNPs and INDELs, the "QD < 2.0 || FS > 60.0 || MQ < 40.0" and "QD < 2.0 || FS > || 200.0 || GQ < 20 || DP < 19" parameters were used, respectively. Finally, variants with the "PASS" flag were selected and were functionally annotated using SnpEff v5.1b [101] to match variants with genes, and ANNOVAR (Version: 2018-04-16) was applied for further annotation using ClinVar (2025.07.29) [102]. For downstream analyses, variants with LOW annotation impact (from SnpEff), protective, conflicting, benign, and uncertain significance (from ClinVar) tags were filtered out as well. Since the average number of INDELs was significantly decreased after the filtering steps, only high-confidence SNPs were used in the downstream analyses. SNPs were summed up per gene and per sample in a matrix using R v4.5.1 (https://www.R-project.org/, accessed on 26 September 2025). The mutation matrix was used as input to compare mutation load among the predefined groups (YFC, OFC, YCC, and OCC) for each gene using the previously described machine learning data analysis.

4.7. Machine Learning and Data Analysis

The analyses were performed with Orange 3.35.0 software. For feature selection, we used an information gain algorithm with the Rank package. The visualization based on the selected features was performed using t-distributed Stochastic Neighbor Embedding (t-SNE), applying Global Structure Preservation, Data Scaling, and Principal Component Analysis preprocessing, which are available within the package. The classification was performed using the Random Forest algorithm based on the selected features. This involved creating 500 classification trees using the square root of the number of features at each split. The efficiency of the classification was expressed in terms of sensitivity, specificity, and classification accuracy. Gene set enrichment analysis was performed using the Reactome Pathway Database (v.84). In some cases, the continuous variables between groups were compared using one-way ANOVA with Tukey's post hoc test. In all instances, $p < 0.05$ was considered significant.

4.8. Gene Set Enrichment Analysis

The gene set enrichment analysis was performed simultaneously on the 494 genes, with the statistical domain scope set to all annotated genes. Multiple testing correction was carried out using the g:SCS algorithm, which was specifically developed for GSEA analysis. We used the following online tool: https://biit.cs.ut.ee/gplink/l/aio83Uc8kSI (accessed on 25 September 2025).

5. Conclusions

Here, we report 30 genes that were the minimum set of identified susceptibility and protective genes for the precise and accurate prediction of disease severity in the investigated Hungarian patient cohort. Assessing the risk of whether an individual will develop severe COVID-19 is important for the patient and the clinicians involved in the treatment. Genetic screening with the identified 30 genes could potentially provide a cost-effective approach, especially in light of the decreasing costs of WES and screening based on gene panels. Further work is necessary to determine if the identified genes can aid in precise prediction in other populations. The use of this minimal gene set could aid the prediction of COVID-19 disease severity, which could have an important impact on

treatment decisions. Our results could also contribute to a better understanding of the pathomechanism of the disease, elucidating molecular pathways involved in COVID-19 disease development, possibly revealing novel therapeutic targets and modalities.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms27052358/s1.

Author Contributions: Conceptualization, A.N., G.L.K., A.G. and M.S.; patient selection and analysis of clinical data, A.N., N.N., M.P., B.A.B., Z.B., Á.V. and B.B.; patient recruitment and clinical sample collection, B.A.B., Z.B., Á.V., B.B., S.K.L. and Z.P.; processing of blood samples and DNA isolation, F.S., V.E. and K.G.; whole-exome sequencing, A.N. and P.U.; bioinformatics analysis, B.G. and J.K.; machine learning analysis, M.B.; development of the patient stratification method, A.N., G.L.K., A.G. and M.S.; writing—original draft preparation, A.N.; writing—review and editing, A.N., A.G. and M.S.; visualization, M.B.; supervision, M.S., A.G. and G.L.K.; project administration, A.N.; funding acquisition, A.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was made possible through generous funding from the National Research, Development and Innovation Office's 2020-2.1.1-ED-2020-00009 tender. The grant provided essential financial support for the design, execution, and analysis of the experiments conducted in this study.

Institutional Review Board Statement: The COVID-19 HUNGEN Project received approval from the national ethics committee on 10 August 2020. All procedures were performed in accordance with the ethical authorization number 19697-6/2020/EÜIG issued by the National Center for Public Health, Hungary, and the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: We would like to express our sincere gratitude to the members of the COVID-19 HUNGEN project for their contributions, valuable input, and assistance during various stages of the project. The list of all participants is provided in Supplementary Document S3.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

COVID-19	Coronavirus disease 2019
GWAS	Genome-Wide Association Studies
WES	Whole-exome sequencing
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
YFC	Young focus cohort
OFC	Old focus cohort
YCC	Young control cohort
OCC	Old control cohort
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
ANOVA	Analysis of variance
HPC	High-performance computing
BAM	Binary alignment and map
GATK	Genome Analysis Toolkit
BSQR	Base quality score recalibration
SNV	Single-nucleotide variant
INDEL	Insertion/deletion polymorphism

GSEA	Gene set enrichment analysis
t-SNE	t-distributed stochastic neighbor embedding
CDK	Cyclin-dependent Kinase
PBMCs	Peripheral blood mononuclear cells

References

1. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733. [CrossRef]
2. Sah, P.; Fitzpatrick, M.C.; Zimmer, C.F.; Abdollahi, E.; Juden-Kelly, L.; Moghadas, S.M.; Singer, B.H.; Galvani, A.P. Asymptomatic SARS-CoV-2 Infection: A Systematic Review and Meta-Analysis. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2109229118. [CrossRef]
3. Zhang, Q.; Bastard, P.; COVID Human Genetic Effort; Karbuz, A.; Gervais, A.; Tayoun, A.A.; Aiuti, A.; Belot, A.; Bolze, A.; Gaudet, A.; et al. Human Genetic and Immunological Determinants of Critical COVID-19 Pneumonia. *Nature* **2022**, *603*, 587–598. [CrossRef]
4. Peixoto, V.R.; Vieira, A.; Aguiar, P.; Sousa, P.; Carvalho, C.; Thomas, D.; Abrantes, A.; Nunes, C. Determinants for hospitalisations, intensive care unit admission and death among 20,293 reported COVID-19 cases in Portugal, March to April 2020. *Eurosurveillance* **2021**, *26*, 2001059. [CrossRef] [PubMed]
5. Richardson, S.; Hirsch, J.S.; Narasimhan, M.; Crawford, J.M.; McGinn, T.; Davidson, K.W.; the Northwell COVID-19 Research Consortium. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* **2020**, *323*, 2052. [CrossRef]
6. Mbarek, H.; Cocca, M.; Al-Sarraj, Y.; Saad, C.; Mezzavilla, M.; AlMuftah, W.; Cocciadiferro, D.; Novelli, A.; Quinti, I.; AlTawashi, A.; et al. Poking COVID-19: Insights on Genomic Constraints among Immune-Related Genes between Qatari and Italian Populations. *Genes* **2021**, *12*, 1842. [CrossRef]
7. Quintero, A.M.; Eisner, M.; Sayegh, R.; Wright, T.; Ramilo, O.; Leber, A.L.; Wang, H.; Mejias, A. Differences in SARS-CoV-2 Clinical Manifestations and Disease Severity in Children and Adolescents by Infecting Variant. *Emerg. Infect. Dis.* **2022**, *28*, 2278–2288. [CrossRef] [PubMed]
8. Wei, Y.-Y.; Wang, R.-R.; Zhang, D.-W.; Chen, S.-H.; Tan, Y.-Y.; Zhang, W.-T.; Han, M.-F.; Fei, G.-H. Differential Characteristics of Patients for Hospitalized Severe COVID-19 Infected by the Omicron Variants and Wild Type of SARS-CoV-2 in China. *J. Inflamm. Res.* **2023**, *16*, 3063–3078. [CrossRef] [PubMed]
9. Langton, D.J.; Bourke, S.C.; Lie, B.A.; Reiff, G.; Natu, S.; Darlay, R.; Burn, J.; Echevarria, C. The Influence of HLA Genotype on the Severity of COVID-19 Infection. *HLA* **2021**, *98*, 14–22. [CrossRef]
10. Bucciol, G.; Meyts, I.; Abel, L.; Al-Muhsen, S.; Aiuti, A.; Al-Mulla, F.; Andreakos, E.; Antonio, N.; Arias, A.A.; Trouillet-Assant, S.; et al. Inherited and Acquired Errors of Type I Interferon Immunity Govern Susceptibility to COVID-19 and Multisystem Inflammatory Syndrome in Children. *J. Allergy Clin. Immunol.* **2023**, *151*, 832–840. [CrossRef]
11. Cobat, A.; Zhang, Q.; COVID Human Genetic Effort; Abel, L.; Casanova, J.-L.; Fellay, J. Human Genomics of COVID-19 Pneumonia: Contributions of Rare and Common Variants. *Annu. Rev. Biomed. Data Sci.* **2023**, *6*, 465–486. [CrossRef]
12. Casanova, J.-L.; Anderson, M.S. Unlocking Life-Threatening COVID-19 through Two Types of Inborn Errors of Type I IFNs. *J. Clin. Investig.* **2023**, *133*, e166283. [CrossRef]
13. Kousathanas, A.; Pairo-Castineira, E.; Rawlik, K.; Stuckey, A.; Odhams, C.A.; Walker, S.; Russell, C.D.; Malinauskas, T.; Wu, Y.; Millar, J.; et al. Whole-Genome Sequencing Reveals Host Factors Underlying Critical COVID-19. *Nature* **2022**, *607*, 97–103. [CrossRef] [PubMed]
14. COVID-19 Host Genetics Initiative. Mapping the Human Genetic Architecture of COVID-19. *Nature* **2021**, *600*, 472–477. [CrossRef] [PubMed]
15. Zeberg, H.; Pääbo, S. The Major Genetic Risk Factor for Severe COVID-19 Is Inherited from Neanderthals. *Nature* **2020**, *587*, 610–612. [CrossRef]
16. Wang, R.Y.; Guo, T.Q.; Li, L.G.; Jiao, J.Y.; Wang, L.Y. Predictions of COVID-19 Infection Severity Based on Co-Associations between the SNPs of Co-Morbid Diseases and COVID-19 through Machine Learning of Genetic Data. In *Proceedings of the 2020 IEEE 8th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 20 November 2020*; IEEE: New York, NY, USA, 2020; pp. 92–96.
17. Cunningham, J.W.; Vaduganathan, M.; Claggett, B.L.; Jering, K.S.; Bhatt, A.S.; Rosenthal, N.; Solomon, S.D. Clinical Outcomes in Young US Adults Hospitalized With COVID-19. *JAMA Intern. Med.* **2021**, *181*, 379–381. [CrossRef]
18. Rocha, G.D.; Oliveira, P.R.S.; Sá, M.V.B.O.; Campos, T.L.; Galisa, S.L.G.; Silva, A.S.; Moura, P.; São Pedro, R.B.; Tavares, N.M.; Boaventura, V.S.; et al. Rare genetic variants and severe COVID-19 in previously healthy admixed Latin American adults. *Sci. Rep.* **2025**, *15*, 23074. [CrossRef] [PubMed]

19. Arons, M.M.; Hatfield, K.M.; Reddy, S.C.; Kimball, A.; James, A.; Jacobs, J.R.; Taylor, J.; Spicer, K.; Bardossy, A.C.; Oakley, L.P.; et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. *N. Engl. J. Med.* **2020**, *382*, 2081–2090. [CrossRef] [PubMed]
20. Leung, S.; Hossain, N. Recurrence and Recovery of COVID-19 in an Older Adult Patient with Multiple Comorbidities: A Case Report. *Gerontology* **2021**, *67*, 445–448. [CrossRef] [PubMed]
21. Borras-Bermejo, B.; Martínez-Gómez, X.; Gutierrez San Miguel, M.; Esperalba, J.; Antón, A.; Martin, E.; Selvi, M.; Abadías, M.J.; Román, A.; Pumarola, T.; et al. Asymptomatic SARS-CoV-2 Infection in Nursing Homes, Barcelona, Spain, April 2020. *Emerg. Infect. Dis.* **2020**, *26*, 2281–2283. [CrossRef] [PubMed]
22. Bautista, E.; Chotpitayasunondh, T.; Gao, Z.; Harper, S.A.; Shaw, M.; Uyeki, T.M.; Zaki, S.R.; Hayden, F.G.; Hui, D.S.; Kettner, J.D.; et al. Clinical aspects of pandemic 2009 influenza A (H1N1) virus infection. *N. Engl. J. Med.* **2010**, *362*, 1708–1719.
23. Jain, S.; Kamimoto, L.; Bramley, A.M.; Schmitz, A.M.; Benoit, S.R.; Louie, J.; Sugerman, D.E.; Druckenmiller, J.K.; Ritger, K.A.; Chugh, R.; et al. Hospitalized Patients with 2009 H1N1 Influenza in the United States, April–June 2009. *N. Engl. J. Med.* **2009**, *361*, 1935–1944. [CrossRef] [PubMed]
24. Van Kerkhove, M.D.; Vandemaele, K.A.H.; Shinde, V.; Jaramillo-Gutierrez, G.; Koukounari, A.; Donnelly, C.A.; Carlino, L.O.; Owen, R.; Paterson, B.; Pelletier, L.; et al. Risk Factors for Severe Outcomes following 2009 Influenza A (H1N1) Infection: A Global Pooled Analysis. *PLoS Med.* **2011**, *8*, e1001053. [CrossRef] [PubMed]
25. Biddle, J.E.; Nguyen, H.Q.; Talbot, H.K.; Rolfes, M.A.; Biggerstaff, M.; Johnson, S.; Reed, C.; Belongia, E.A.; Grijalva, C.G.; Mellis, A.M.; et al. Asymptomatic and Mildly Symptomatic Influenza Virus Infections by Season: Case-Ascertained Household Transmission Studies, United States, 2017–2023. *J. Infect. Dis.* **2025**, *232*, e637–e641. [CrossRef] [PubMed]
26. COVID-19 Host Genetics Initiative. A first update on mapping the human genetic architecture of COVID-19. *Nature* **2022**, *608*, E1–E10. [CrossRef]
27. Shemesh, M.; Lochte, S.; Piehler, J.; Schreiber, G. IFNAR1 and IFNAR2 play distinct roles in initiating type I interferon-induced JAK–STAT signaling and activating STATs. *Sci. Signal.* **2021**, *14*, eabe4627. [CrossRef]
28. Gaub, A.; Sheikh, B.N.; Basilicata, M.F.; Vincent, M.; Nizon, M.; Colson, C.; Bird, M.J.; Bradner, J.E.; Thevenon, J.; Boutros, M.; et al. Evolutionary conserved NSL complex/BRD4 axis controls transcription activation via histone acetylation. *Nat. Commun.* **2020**, *11*, 2243. [CrossRef]
29. Dias, J.; Nguyen, N.V.; Georgiev, P.; Gaub, A.; Brettschneider, J.; Cusack, S.; Kadlec, J.; Akhtar, A. Structural analysis of the KANSL1/WDR5/KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex. *Genes Dev.* **2014**, *28*, 929–942. [CrossRef]
30. Zhong, F.L.; Robinson, K.; Teo, D.E.T.; Tan, K.-Y.; Lim, C.; Harapas, C.R.; Yu, C.-H.; Xie, W.H.; Sobota, R.M.; Au, V.B.; et al. Human DPP9 represses NLRP1 inflammasome and protects against autoinflammatory diseases via both peptidase activity and FIIND domain binding. *J. Biol. Chem.* **2018**, *293*, 18864–18878. [CrossRef]
31. Huang, M.; Zhang, X.; Toh, G.A.; Gong, Q.; Wang, J.; Han, Z.; Wu, B.; Zhong, F.; Chai, J. Structural and biochemical mechanisms of NLRP1 inhibition by DPP9. *Nature* **2021**, *592*, 773–777. [CrossRef]
32. Schips, T.G.; Vanhoutte, D.; Vo, A.; Correll, R.N.; Brody, M.J.; Khalil, H.; Karch, J.; Tjondrokoesoemo, A.; Sargent, M.A.; Maillet, M.; et al. Thrombospondin-3 augments injury-induced cardiomyopathy by intracellular integrin inhibition and sarcolemmal instability. *Nat. Commun.* **2019**, *10*, 76. [CrossRef] [PubMed]
33. Piggin, C.L.; Roden, D.L.; Gallego-Ortega, D.; Lee, H.J.; Oakes, S.R.; Ormandy, C.J. ELF5 isoform expression is tissue-specific and significantly altered in cancer. *Breast Cancer Res.* **2016**, *18*, 4. [CrossRef]
34. Oakes, S.R.; Naylor, M.J.; Asselin-Labat, M.-L.; Blazek, K.D.; Gardiner-Garden, M.; Hilton, H.N.; Kazlauskas, M.; Pritchard, M.A.; Chodosh, L.A.; Pfeffer, P.L.; et al. The Ets transcription factor Elf5 specifies mammary alveolar cell fate. *Genes Dev.* **2008**, *22*, 581–586. [CrossRef]
35. Adams, J.C.; Lawler, J. The thrombospondins. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, a009712. [CrossRef]
36. Engelhart, D.C.; Granados, J.C.; Shi, D.; Saier, M.H.; Baker, M.E.; Abagyan, R.; Nigam, S.K. Systems biology analysis reveals eight SLC22 transporter subgroups, including OATs, OCTs, and OCTNs. *Int. J. Mol. Sci.* **2020**, *21*, 1791. [CrossRef] [PubMed]
37. Kastens, G.; Berger-Santangelo, H.; Gerstner, S.; Ufartes, R.; Mischak, M.; Borchers, A.; Pauli, S. FBRSL1 regulates the expression of chromatin regulators BRPF1 and KAT6A. *Hum. Genet.* **2025**, *144*, 809–826. [CrossRef]
38. Chaudhary, R.; Meher, A.; Krishnamoorthy, P.; Kumar, H. Interplay of host and viral factors in inflammatory pathway mediated cytokine storm during RNA virus infection. *Curr. Res. Immunol.* **2023**, *4*, 100062. [CrossRef]
39. Maddaloni, L.; Bugani, G.; Fracella, M.; Bitossi, C.; D'Auria, A.; Aloisi, F.; Azri, A.; Santinelli, L.; Ben M'Hadheb, M.; Pierangeli, A.; et al. Pattern Recognition Receptors (PRRs) Expression and Activation in COVID-19 and Long COVID: From SARS-CoV-2 Escape Mechanisms to Emerging PRR-Targeted Immunotherapies. *Microorganisms* **2025**, *13*, 2176. [CrossRef] [PubMed]
40. Lee, A.J.; Feng, E.; Chew, M.V.; Balint, E.; Poznanski, S.M.; Giles, E.; Zhang, A.; Marzok, A.; Revill, S.D.; Vahedi, F.; et al. Type I interferon regulates proteolysis by macrophages to prevent immunopathology following viral infection. *PLoS Pathog.* **2022**, *18*, e1010471. [CrossRef]

41. Crouse, J.; Kalinke, U.; Oxenius, A. Regulation of antiviral T cell responses by type I interferons. *Nat. Rev. Immunol.* **2015**, *15*, 231–242. [CrossRef]
42. Nie, J.; Zhou, L.; Tian, W.; Liu, X.; Yang, L.; Yang, X.; Zhang, Y.; Wei, S.; Wang, D.W.; Wei, J. Deep insight into cytokine storm: From pathogenesis to treatment. *Signal Transduct. Target. Ther.* **2025**, *10*, 112. [CrossRef]
43. Boroujeni, M.E.; Sekrecka, A.; Antonczyk, A.; Hassani, S.; Sekrecki, M.; Nowicka, H.; Lopacinska, N.; Olya, A.; Kluzek, K.; Wesoly, J.; et al. Dysregulated interferon response and immune hyperactivation in severe COVID-19: Targeting STATs as a novel therapeutic strategy. *Front. Immunol.* **2022**, *13*, 888897. [CrossRef]
44. Zhou, Y.; Liu, Y.; Gupta, S.; Paramo, M.I.; Hou, Y.; Mao, C.; Luo, Y.; Judd, J.; Wierbowski, S.; Bertolotti, M.; et al. A comprehensive SARS-CoV-2–human protein–protein interactome reveals COVID-19 pathobiology and potential host therapeutic targets. *Nat. Biotechnol.* **2022**, *41*, 128–139. [CrossRef]
45. Babačić, H.; Christ, W.; Araújo, J.E.; Mermelekas, G.; Sharma, N.; Tynell, J.; García, M.; Varnaite, R.; Asgeirsson, H.; Glans, H.; et al. Comprehensive proteomics and meta-analysis of COVID-19 host response. *Nat. Commun.* **2023**, *14*, 41159. [CrossRef] [PubMed]
46. Gordon, D.E.; Jang, G.M.; Bouhaddou, M.; Xu, J.; Obernier, K.; O'Meara, M.J.; Guo, J.Z.; Swaney, D.L.; Tummino, T.A.; Hüttenhain, R.; et al. A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature* **2020**, *583*, 459–468. [CrossRef] [PubMed]
47. Wang, K.; Ni, X.; Deng, X.; Nan, J.; Ma-Lauer, Y.; von Brunn, A.; Zeng, R.; Lei, J. The CoV-Y Domain of SARS-CoV-2 Nsp3 Interacts with BRAP to Stimulate NF-κB Signaling and Induce Host Inflammatory Responses. *Int. J. Biol. Macromol.* **2024**, *266*, 136123. [CrossRef] [PubMed]
48. Read, S.A.; Obeid, S.; Ahlenstiel, C.; Ahlenstiel, G. The Role of Zinc in Antiviral Immunity. *Adv. Nutr.* **2019**, *10*, 696–710. [CrossRef]
49. Maares, M.; Hackler, J.; Haupt, A.; Heller, R.A.; Bachmann, M.; Diegmann, J.; Moghaddam, A.; Schomburg, L.; Haase, H. Free Zinc as a Predictive Marker for COVID-19 Mortality Risk. *Nutrients* **2022**, *14*, 1407. [CrossRef]
50. Jin, D.; Shen, X.; Liu, H.; Xu, Y.; Jiang, C.; Lin, Y. The Nutritional Roles of Zinc for the Immune System and COVID-19. *Front. Nutr.* **2024**, *11*, 1385591. [CrossRef]
51. Yadav, D.; Kumar, P.V.S.N.K.; Tomo, S.; Sankanagoudar, S.; Charan, J.; Purohit, A.; Nag, V.; Bhatia, P.; Singh, K.; Dutt, N.; et al. Association of Iron-Related Biomarkers with Severity and Mortality in COVID-19 Patients. *J. Trace Elem. Med. Biol.* **2022**, *74*, 127075. [CrossRef]
52. Chen, R.; Zhang, L.; Zhong, B.; Tan, B.; Liu, Y.; Shu, H.B. The ubiquitin-specific protease 17 is involved in virus-triggered type I IFN signaling. *Cell Res.* **2010**, *20*, 802–811. [CrossRef] [PubMed]
53. Thomas, T.; Stefanoni, D.; Reisz, J.A.; Nemkov, T.; Bertolone, L.; Francis, R.O.; Hudson, K.E.; Zimring, J.C.; Hansen, K.C.; Hod, E.A.; et al. COVID-19 Infection Alters Kynurenine and Fatty Acid Metabolism, Correlating with IL-6 Levels and Renal Status. *JCI Insight* **2020**, *5*, e140327. [CrossRef]
54. V'kovski, P.; Kratzel, A.; Steiner, S.; Stalder, H.; Thiel, V. Coronavirus Biology and Replication: Implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **2021**, *19*, 155–170. [CrossRef]
55. Pickrell, A.M.; Youle, R.J. The Roles of PINK1, Parkin, and Mitochondrial Fidelity in Parkinson's Disease. *Neuron* **2015**, *85*, 257–273. [CrossRef]
56. Agnew, T.; Munnur, D.; Crawford, K.; Palazzo, L.; Mikoc, A.; Ahel, I. MacroD1 Is a Promiscuous ADP-Ribosyl Hydrolase Localized to Mitochondria. *Front. Microbiol.* **2018**, *9*, 20. [CrossRef]
57. Robinot, R.; Hubert, M.; de Melo, G.D.; Lazarini, F.; Bruel, T.; Smith, N.; Levallois, S.; Larrous, F.; Fernandes, J.; Gellenoncourt, S.; et al. SARS-CoV-2 Infection Induces the Dedifferentiation of Multiciliated Cells and Impairs Mucociliary Clearance. *Nat. Commun.* **2021**, *12*, 4354. [CrossRef]
58. Cheng, G.; Zhong, M.; Kawaguchi, R.; Kassai, M.; Al-Ubaidi, M.; Deng, J.; Ter-Stepanian, M.; Sun, H. Identification of PLXDC1 and PLXDC2 as the Transmembrane Receptors for the Multifunctional Factor PEDF. *eLife* **2014**, *3*, e05401. [CrossRef]
59. Feyaerts, D.; Luyten, T.; van der Vaart, A.; Smits, E.; de Vries, R.D.; Mortier, G.; Heylen, L.; Meersseman, P.; van Braeckel, E.; Lambrecht, B.N.; et al. Integrated plasma proteomic and single-cell immune signaling network signatures demarcate mild, moderate, and severe COVID-19. *Cell Rep. Med.* **2022**, *3*, 100680. [CrossRef] [PubMed]
60. Litovchick, L.; Florens, L.A.; Swanson, S.K.; Washburn, M.P.; DeCaprio, J.A. DYRK1A Protein Kinase Promotes Quiescence and Senescence through DREAM Complex Assembly. *Genes Dev.* **2011**, *25*, 801–813. [CrossRef] [PubMed]
61. Gabrielson, M.; Reizer, E.; Stål, O.; Tina, E. Mitochondrial Regulation of Cell Cycle Progression through SLC25A43. *Biochem. Biophys. Res. Commun.* **2016**, *469*, 1090–1096. [CrossRef]
62. Monticelli, M.; Hay Mele, B.; Benetti, E.; Fallerini, C.; Baldassarri, M.; Furini, S.; Frullanti, E.; Mari, F.; Andreotti, G.; Cubellis, M.V.; et al. Protective Role of a TMPRSS2 Variant on Severe COVID-19 Outcome in Young Males and Elderly Women. *Genes* **2021**, *12*, 596. [CrossRef]
63. Sun, H.; Yang, Q.; Zhang, Y.; Cui, S.; Zhou, Z.; Zhang, P.; Jia, L.; Zhang, M.; Wang, Y.; Chen, X.; et al. Syntaxin-6 Restricts SARS-CoV-2 Infection by Facilitating Virus Trafficking to Autophagosomes. *J. Virol.* **2025**, *99*, e00002-25. [CrossRef] [PubMed]

64. Su, H.C.; Jing, H.; Zhang, Y.; Casanova, J.-L. Interfering with Interferons: A Critical Mechanism for Critical COVID-19 Pneumonia. *Annu. Rev. Immunol.* **2023**, *41*, 561–585. [CrossRef]
65. Matuozzo, D.; Talouarn, E.; Marchal, A.; Zhang, P.; Manry, J.; Seeleuthner, Y.; Zhang, Y.; Bolze, A.; Chaldebas, M.; Milisavljevic, B.; et al. Rare Predicted Loss-of-Function Variants of Type I IFN Immunity Genes Are Associated with Life-Threatening COVID-19. *Genome Med.* **2023**, *15*, 22, Correction in *Genome Med.* **2024**, *16*, 6. [CrossRef]
66. Horowitz, J.E.; Kosmicki, J.A.; Damask, A.; Sharma, D.; Roberts, G.H.L.; Justice, A.E.; Banerjee, N.; Coignet, M.V.; Yadav, A.; Leader, J.B.; et al. Genome-Wide Analysis Provides Genetic Evidence That ACE2 Influences COVID-19 Risk and Yields Risk Scores Associated with Severe Disease. *Nat. Genet.* **2022**, *54*, 382–392. [CrossRef] [PubMed]
67. The Severe Covid-19 GWAS Group Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.* **2020**, *383*, 1522–1534. [CrossRef]
68. Pairo-Castineira, E.; Clohisey, S.; Klaric, L.; Bretherick, A.D.; Rawlik, K.; Pasko, D.; Walker, S.; Parkinson, N.; Fourman, M.H.; Russell, C.D.; et al. Genetic Mechanisms of Critical Illness in COVID-19. *Nature* **2021**, *591*, 92–98. [CrossRef] [PubMed]
69. Lai, G.; Liu, H.; Deng, J.; Li, K.; Xie, B. A Novel 3-Gene Signature for Identifying COVID-19 Patients Based on Bioinformatics and Machine Learning. *Genes* **2022**, *13*, 1602. [CrossRef]
70. Zheng, W.; Zhang, Y.; Lai, G.; Xie, B. Landscape of infiltrated immune cell characterization in COVID-19. *Heliyon* **2024**, *10*, e28174. [CrossRef]
71. Kollár, L.; Horváth, I. Manual for the prevention and treatment of infections caused by a new coronavirus (SARS-CoV-2) identified in 2020 (COVID-19). 50. Available online: https://msotke.hu/2020/12/30/magyar-koronavirus-kezikonyv/ (accessed on 25 August 2025).
72. Statsenko, Y.; Al Zahmi, F.; Habuza, T.; Almansoori, T.M.; Smetanina, D.; Simiyu, G.L.; Neidl-Van Gorkom, K.; Ljubisavljevic, M.; Awawdeh, R.; Elshekhali, H.; et al. Impact of Age and Sex on COVID-19 Severity Assessed From Radiologic and Clinical Findings. *Front. Cell. Infect. Microbiol.* **2022**, *11*, 777070. [CrossRef]
73. Kang, S.H.; Kim, S.W.; Kim, A.Y.; Cho, K.H.; Park, J.W.; Do, J.Y. Association between Chronic Kidney Disease or Acute Kidney Injury and Clinical Outcomes in COVID-19 Patients. *J. Korean Med. Sci.* **2020**, *35*, e434. [CrossRef]
74. Ssentongo, P.; Ssentongo, A.E.; Heilbrunn, E.S.; Ba, D.M.; Chinchilli, V.M. Association of Cardiovascular Disease and 10 Other Pre-Existing Comorbidities with COVID-19 Mortality: A Systematic Review and Meta-Analysis. *PLoS ONE* **2020**, *15*, e0238215. [CrossRef]
75. Zhang, J.; Wu, J.; Sun, X.; Xue, H.; Shao, J.; Cai, W.; Jing, Y.; Yue, M.; Dong, C. Association of Hypertension with the Severity and Fatality of SARS-CoV-2 Infection: A Meta-Analysis. *Epidemiol. Infect.* **2020**, *148*, e106. [CrossRef]
76. Centers for Disease Control and Prevention. Brief Summary of Findings on the Association between Thalassemia and Severe COVID-19 Outcomes. CDC COVID-19 Science Brief, June 2022. Available online: https://www.cdc.gov/covid/media/pdfs/science-briefs/Thalassemia_review.pdf (accessed on 25 August 2025).
77. Centers for Disease Control and Prevention. Brief Summary of Findings on the Association between Disabilities and Severe COVID-19 Outcomes. CDC COVID-19 Science Brief, October 2021. Available online: https://archive.cdc.gov/www_cdc_gov/coronavirus/2019-ncov/hcp/clinical-care/underlyingconditions.html (accessed on 25 August 2025).
78. Pilgram, L.; Eberwein, L.; Wille, K.; Koehler, F.C.; Stecher, M.; Rieg, S.; Kielstein, J.T.; Jakob, C.E.M.; Rüthrich, M.; Burst, V.; et al. Clinical Course and Predictive Risk Factors for Fatal Outcome of SARS-CoV-2 Infection in Patients with Chronic Kidney Disease. *Infection* **2021**, *49*, 725–737. [CrossRef]
79. Yang, Z.; Wang, M.; Zhu, Z.; Liu, Y. Coronavirus Disease 2019 (COVID-19) and Pregnancy: A Systematic Review. *J. Matern.-Fetal Neonatal Med.* **2022**, *35*, 1619–1622. [CrossRef] [PubMed]
80. Zuin, M.; Guasti, P.; Roncon, L.; Cervellati, C.; Zuliani, G. Dementia and the Risk of Death in Elderly Patients with COVID-19 Infection: Systematic Review and Meta-analysis. *Int. J. Geriatr. Psychiatry* **2021**, *36*, 697–703. [CrossRef]
81. Wang, B.; Li, R.; Lu, Z.; Huang, Y. Does Comorbidity Increase the Risk of Patients with COVID-19: Evidence from Meta-Analysis. *Aging* **2020**, *12*, 6049–6057. [CrossRef]
82. Aziz, F.; Mandelbrot, D.; Singh, T.; Parajuli, S.; Garg, N.; Mohamed, M.; Astor, B.C.; Djamali, A. Early Report on Published Outcomes in Kidney Transplant Recipients Compared to Nontransplant Patients Infected with Coronavirus Disease 2019. *Transpl. Procs.* **2020**, *52*, 2659–2662. [CrossRef]
83. Ssentongo, P.; Heilbrunn, E.S.; Ssentongo, A.E.; Advani, S.; Chinchilli, V.M.; Nunez, J.J.; Du, P. Epidemiology and Outcomes of COVID-19 in HIV-Infected Individuals: A Systematic Review and Meta-Analysis. *Sci. Rep.* **2021**, *11*, 6283. [CrossRef] [PubMed]
84. Banerjee, A.; Chen, S.; Pasea, L.; Lai, A.G.; Katsoulis, M.; Denaxas, S.; Nafilyan, V.; Williams, B.; Wong, W.K.; Bakhai, A.; et al. Excess Deaths in People with Cardiovascular Diseases during the COVID-19 Pandemic. *Eur. J. Prev. Cardiol.* **2021**, *28*, 1599–1609. [CrossRef] [PubMed]

85. Saini, K.S.; Tagliamento, M.; Lambertini, M.; McNally, R.; Romano, M.; Leone, M.; Curigliano, G.; de Azambuja, E. Mortality in Patients with Cancer and Coronavirus Disease 2019: A Systematic Review and Pooled Analysis of 52 Studies. *Eur. J. Cancer* **2020**, *139*, 43–50. [CrossRef]
86. Fadini, G.P.; Morieri, M.L.; Boscari, F.; Fioretto, P.; Maran, A.; Busetto, L.; Bonora, B.M.; Selmin, E.; Arcidiacono, G.; Pinelli, S.; et al. Newly-Diagnosed Diabetes and Admission Hyperglycemia Predict COVID-19 Severity by Aggravating Respiratory Deterioration. *Diabetes Res. Clin. Pract.* **2020**, *168*, 108374. [CrossRef] [PubMed]
87. Yang, J.; Hu, J.; Zhu, C. Obesity Aggravates COVID-19: A Systematic Review and Meta-analysis. *J. Med. Virol.* **2021**, *93*, 257–261. [CrossRef] [PubMed]
88. Földi, M.; Farkas, N.; Kiss, S.; Zádori, N.; Váncsa, S.; Szakó, L.; Dembrovszky, F.; Solymár, M.; Bartalis, E.; Szakács, Z.; et al. Obesity Is a Risk Factor for Developing Critical Condition in COVID-19 Patients: A Systematic Review and Meta-analysis. *Obes. Rev.* **2020**, *21*, e13095. [CrossRef]
89. Hamer, M.; Gale, C.R.; Kivimäki, M.; Batty, G.D. Overweight, Obesity, and Risk of Hospitalization for COVID-19: A Community-Based Cohort Study of Adults in the United Kingdom. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 21011–21013. [CrossRef]
90. So, C.N.; Green, R.F.; Drzymalia, E.; Hill, A.L.; Okasako, D.L.; Stone, E.C.; Taliano, J.; Koumans, E.; Sircar, K.D. *Brief Summary of Findings on the Association Between Cystic Fibrosis and Severe COVID-19 Outcomes*; CDC: Atlanta, GA, USA, 2022.
91. Fadini, G.P.; Morieri, M.L.; Longato, E.; Avogaro, A. Prevalence and Impact of Diabetes among People Infected with SARS-CoV-2. *J. Endocrinol. Invest.* **2020**, *43*, 867–869. [CrossRef]
92. Yang, J.; Zheng, Y.; Gou, X.; Pu, K.; Chen, Z.; Guo, Q.; Ji, R.; Wang, H.; Wang, Y.; Zhou, Y. Prevalence of Comorbidities and Its Effects in Patients Infected with SARS-CoV-2: A Systematic Review and Meta-Analysis. *Int. J. Infect. Dis.* **2020**, *94*, 91–95. [CrossRef]
93. Michelon, I.; Vilbert, M.; Pinheiro, I.S.; Costa, I.L.; Lorea, C.F.; Castonguay, M.; Tran, T.H.; Forté, S. COVID-19 Outcomes in Patients with Sickle Cell Disease or Sickle Cell Trait: A Systematic Review and Meta-analysis. *EClinicalMedicine* **2023**, *62*, 102152.
94. Zheng, Z.; Peng, F.; Xu, B.; Zhao, J.; Liu, H.; Peng, J.; Li, Q.; Jiang, C.; Zhou, Y.; Liu, S.; et al. Risk Factors of Critical & Mortal COVID-19 Cases: A Systematic Literature Review and Meta-Analysis. *J. Infect.* **2020**, *81*, e16–e25. [CrossRef]
95. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)*, 2nd ed.; WHO Press: Geneva, Switzerland, 2004; Available online: https://icd.who.int/browse10/2019/en (accessed on 25 September 2025).
96. Leggett, R.M.; Ramirez-Gonzalez, R.H.; Clavijo, B.J.; Waite, D.; Davey, R.P. Sequencing Quality Assessment Tools to Enable Data-Driven Informatics for High Throughput Genomics. *Front. Genet.* **2013**, *4*, 288. [CrossRef] [PubMed]
97. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]
98. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]
99. Okonechnikov, K.; Conesa, A.; García-Alcalde, F. Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data. *Bioinformatics* **2016**, *32*, 292–294. [CrossRef] [PubMed]
100. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.1–11.10.33. [CrossRef]
101. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain w 1118; Iso-2; Iso-3. *Fly* **2012**, *6*, 80–92. [CrossRef] [PubMed]
102. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* **2010**, *38*, e164. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.