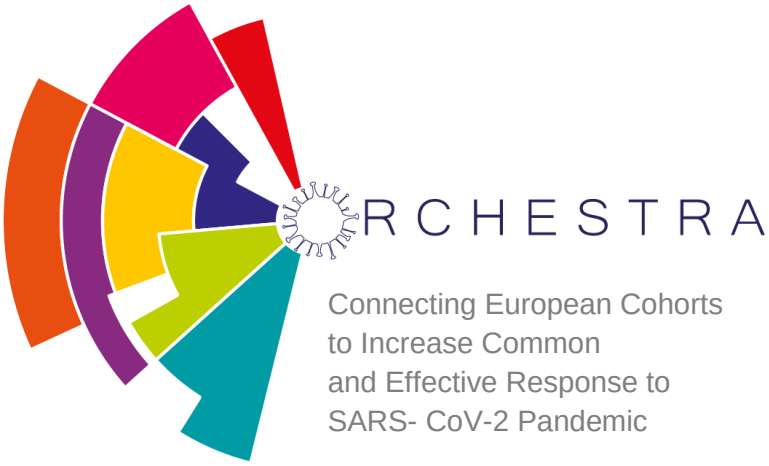


WP6 Deliverable 6.11

# Dataset of human genomic sequencing outcome of patients from prospective COVID-19 cohorts

University of Bologna (UNIBO)



ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

## Project Classification

| | |
|---|---|
| **Project Acronym:** | ORCHESTRA |
| **Project Title:** | Connecting European Cohorts to Increase Common and Effective Response to SARS- CoV-2 Pandemic |
| **Coordinator:** | UNIVR |
| **Grant Agreement Number:** | 101016167 |
| **Funding Scheme:** | Horizon 2020 |
| **Start:** | 1st December 2020 |
| **Duration:** | 48 months |
| **Website:** | www.orchestra-cohort.eu |
| **Email:** | info@orchestra.eu |

## Document Classification

| | |
|---|---|
| **WP No:** | WP6 |
| **Deliverable No:** | D6.11 |
| **Title:** | Dataset of human genomic sequencing outcome of patients from prospective COVID-19 cohorts |
| **Lead Beneficiary:** | UNIBO |
| **Other Involved Beneficiaries:** | INSERM |
| **Nature:** | Report |
| **Dissemination Level:** | Public |
| **Due Delivery Date:** | M40 |
| **Submission Date:** | M42 |
| **Justification of delay:** | Internal revision |
| **Status:** | Completed |
| **Version:** | 1.0 |
| **Author(s):** | Alessandro Mattiaccio, Tommaso Pippucci, Laurent Abel, Alessia Fiorentino, Valerio Carelli, Marco Seri |

## History of Changes

| Version | Date | Created/Modified by |
|---|---|---|
| 0.1 | 27/03/2024 | Alessandro Mattiaccio, Tommaso Pippucci, Laurent Abel, Alessia Fiorentino, Valerio Carelli, Marco Seri |
| 1.0 | 15/05/2024 | Alessandro Mattiaccio, Tommaso Pippucci, Laurent Abel, Alessia Fiorentino, Valerio Carelli, Marco Seri |



ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

# Table of contents





ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

# Executive summary

## Project and deliverable context

ORCHESTRA is an international research project aimed at addressing the coronavirus pandemic and involves 26 partners from 15 countries. The project consists of 11 Work Packages (WPs), each with its own tasks; this deliverable is part of WP6 “Biobanking, genomics and virus-host interactions”. The ORCHESTRA project investigates the genetic, immunological, microbial, and viral characteristics with the goal of identifying markers of disease severity and the long-term impact of SARS-CoV-2 infection in different patient cohorts. The project involved the recruitment of retrospective and prospective samples from existing biobanking facilities participating in the project, containing samples from patients with various degrees of severity of COVID-19 disease and selected for analysis of host genomic DNA, host microbiome (fecal sample), viral genome (nasopharyngeal swab), and analysis of host immune cytokines and chemokines (blood, plasma, and serum sample). Our group aims to identify human genomic variants associated to susceptibility and severity of the disease. Together with INSERM in Paris, the goal was to sequence the exome/genome of 1500 patients (900 from UNIBO and 600 from INSERM).

## Content of the document

The present document describes a report of genomic variation in patients with varying degrees of COVID-19 disease severity within the ORCHESTRA project. All variants from whole exome and whole genome were considered. More specifically, this report focuses on the role of rare single nucleotide variants at 13 influenza susceptibility loci involved in TLR3-dependent type I IFN immunity and the recently reported TYK2 and TLR7 COVID-19 loci in the development of severe COVID-19. In addition, a specific analysis investigated the association of HLA alleles with asymptomatic COVID-19 infection.

*Dissemination level: public*

# Core content

## UNIBO analysis

We have completed the whole genome sequencing of 300 retrospective and 600 prospective patients from three different cohorts. The Spanish cohort (SAS and NEUMO) includes 268 hospitalized patients and 78 healthcare workers; the UNIBO cohort has 385 hospitalized patients; the COVID-HOME cohort includes 169 individuals, ranging from not infected, asymptomatic and mild disease.

We have produced data for all samples for the main types of genomic variants: Single Nucleotide Variants (SNVs), Copy Number Variants (CNVs), HLA haplotypes and mitochondrial variants (mtDNA) and haplotypes.



ORCHESTRA has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101016167

## Methods

During the project we received peripheral blood samples from 900 COVID-19 patients from Italy (UNIBO), Spain (SAS and NEUMO) and the Netherlands (COVID-HOME). Clinical data were collected but, given the difficulty of combining different cohorts in which several patients were not included in REDCap, it was not always available to have the same and complete clinical variables. Based on the World Health Organization Clinical Progression Scale (WHO-CPS) score, patients were ranked from 1 (absence of infection) to 10 (severe illness and death) as shown in figure 1. Most patients are in class 5 and from this class onwards patients are treated with oxygen therapy.

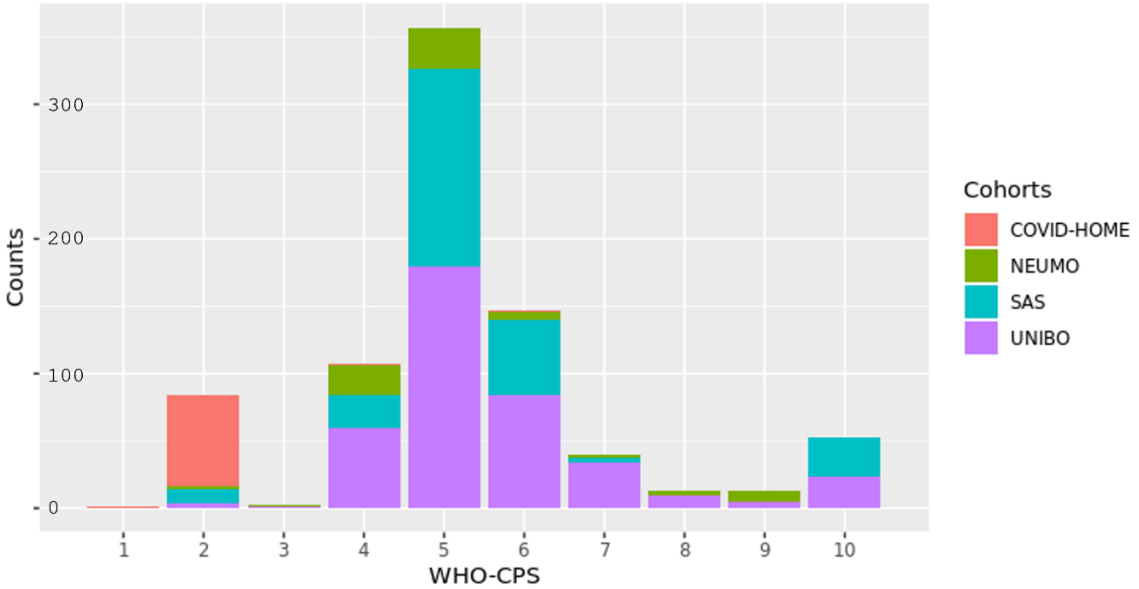


**Figure 1. Distribution of all patients along the WHO Clinical Progressive Scale score (WHO-CPS).** The colors in the legend correspond to COVID-HOME (Netherlands), SAS + NEUMO (Spain) and UNIBO (Italy).

DNA was extracted and normalized and high-coverage whole genome sequencing (WGS) was performed on Illumina platform. WGS reads were processed and variants called using DRAGEN Germline v3.10.4. After the first part on the Illumina Basespace Sequence Hub platform, which produced gVCFs file for each sample, we worked on the multisample VCF on the Sant'Orsola computational genomics platform to retrieve variants for downstream analysis. All variants in the multisample VCFs for SNVs/Indels and SVs/CNVs were functionally annotated and prioritized according to the impacted genes being implicated in the IFN-pathway or associated with the HPO term "Primary immunodeficiency". Finally, we selected rare variants. For CNVs an overlapping of calls from the two main bioinformatics tools, Manta and CANVAS, was carried out to increase the likelihood of true calls.

Mitochondrial (mtDNA) variants were retrieved from the WGS VCF files and mtDNA haplogroup was assigned to each patient using HaploGrep v2.4.0. The mtDNA copy number was calculated using the nuclear and mitochondrial mean coverage.

## Sample pruning

Samples were filtered out based on the residuals of eight quality control metrics and relatedness. Samples were removed that were four median absolute deviations (MADs) above



ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

or below the median for the following metrics: ratio heterozygous-homozygous, ratio transitions-transversions, total deletions, total insertions, total snps, total transitions, total transversions. For the number of total singletons, samples were removed that were more than 8 MADs above the median. Individuals with relatedness phi score equal or greater than 0,0625 were also filtered out (**table 1**).

| Cohort | het/hom | ti/tv | dels | ins | snps | ti | tv | singleton | actually | related | total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **UNIBO** | 26 | 0 | 3 | 3 | 12 | 6 | 3 | 30 | 35 | 2 | 37 |
| **SAS** | 31 | 0 | 1 | 1 | 17 | 2 | 2 | 2 | 33 | 29 | 62 |
| **COVID-HOME** | 9 | 0 | 2 | 2 | 3 | 4 | 3 | 5 | 9 | 27 | 36 |

***Table 1. Sample filtering.*** Number of samples filtered out based on relatedness and several quality control metrics. The *actually* column is the sum of the eight QC steps minus samples who have failed in more than one stage.

## Key findings

For SNVs we focused on rare (MAF≤0.001) mutations in type I IFN pathway immunity genes plus *TYK2* and *TLR7* [1]. In the UNIBO cohort we identified 46 mutations. We found stop gained mutations in *IFNAR1* (p.Gln329Ter) and *TLR3* (p.Tyr756Ter) genes, frameshift variant in *IRF7* (p.Ala177CysfsTer15) gene, splice acceptor variant in *STAT2* (c.1035-2A>G) and *IRF7* (c.887-1G>C) genes in patients treated with oxygen therapy (**table 2**). 41 Missense variants in *IFNAR1*, *IFNAR2*, *IRF3*, *IRF7*, *IRF9*, *STAT2*, *TICAM1*, *TLR3*, *TLR7*, *TRAF3*, *TYK2* and *UNC93B1* were found in patients from different WHO-CPS classes. All variants were heterozygous.

| N | Variant | Gene | Cons | Exon | Intron | cDNA | Protein | CADD | Max_AF | Carriers | WHO-CPS | Sex | Age |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | chr21-33349287-C-T | **IFNAR1** | stop_gained | 7/11 | | 985C>T | Q329* | 33 | -1 | 055C024:0/1 | 6 | M | 72 |
| 2 | chr4-186083954-T-A | **TLR3** | stop_gained | 4/5 | | 2268T>A | Y756* | 35 | -1 | WGS_97:0/1 | 5 | M | 52 |
| 3 | chr11-614363-C-CA | **IRF7** | Frameshift_variant | 4/9 | | 528dup | A177Cfs*15 | 13.55 | -1 | 074C016:0/1 | 6 | M | 77 |
| 4 | chr12-56350890-T-C | **STAT2** | splice_acceptor | | 10/23 | 1035-2A>G | | 34 | -1 | 012DWB:0/1 | 5 | M | 36 |
| 5 | chr11-613596-C-G | **IRF7** | splice_acceptor | | 6/8 | 887-1G>C | | 28.3 | $2*10^{-4}$ | WGS_13:0/1 | 5 | F | 38 |

***Table 2. Top variants in UNIBO cohort.*** Variants with MAF≤0.001. Only stop, frameshift and splice acceptor variants are shown, 41 missense variants not shown.

For SAS cohort we found the same mutational pattern, however, in this cohort the analysis was performed only on patients in REDCap of whom we had the WHO class available (n = 96). We are waiting to receive the missing data from the other patients in order to complete the analysis. We didn't pursue this analysis on COVID-HOME patients because it's not a cohort composed of severe cases.

In a “Primary immunodeficiency” panel of 541 genes from Genomics England PanelApp [2], which also includes type I IFN pathway immunity genes plus *TYK2* and *TLR7*, we performed a rare variant collapsing analysis in patients with oxygen therapy treatment and patients without this therapy in all three cohorts and no genes reached significance after Benjamini & Hochberg correction.



ORCHESTRA has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101016167

We also tested for 62 SNPs with top associations with hospitalization and severity of the disease from Covid-19 Host Genetics Initiative Browser [3] in UNIBO and SAS cohorts, which have hospitalized patients with different severity scores. Instead, SNPs with top associations for susceptibility to SARS-CoV-2 infection were tested in the COVID-HOME cohort. In **table 3** are shown SNPs with a p<0.05. After adjustment for multiple testing with Bonferroni, Sidak, FDR and others, no SNPs confirmed their significance. For SAS cohort we found the same mutational pattern, however, in this cohort the analysis was performed only on patients in REDCap of whom we had the WHO class available (n = 96). We are waiting to receive the missing data from the other patients in order to complete the analysis. We didn't pursue this analysis on COVID-HOME patients because it's not a cohort composed of severe cases.

In a “Primary immunodeficiency” panel of 541 genes from Genomics England PanelApp [2], which also includes type I IFN pathway immunity genes plus *TYK2* and *TLR7*, we performed a rare variant collapsing analysis in patients with oxygen therapy treatment and patients without this therapy in all three cohorts and no genes reached significance after Benjamini & Hochberg correction.

We also tested for 62 SNPs with top associations with hospitalization and severity of the disease from Covid-19 Host Genetics Initiative Browser [3] in UNIBO and SAS cohorts, which have hospitalized patients with different severity scores. Instead, SNPs with top associations for susceptibility to SARS-CoV-2 infection were tested in the COVID-HOME cohort. In **table 3** are shown SNPs with a p<0.05. After adjustment for multiple testing with Bonferroni, Sidak, FDR and others, no SNPs confirmed their significance.

| CHR | Risk allele | SNP | Mapped gene | Unadj | F_A | F_U | OR | Cohort |
|---|---|---|---|---|---|---|---|---|
| chr3:146517122 | A | rs343320 | *PLSCR1* | 0.02948 | 0.1065 | 0.03846 | 2.981 | UNIBO |
| chr19:50379362 | C | rs1405655 | *NR1H2* | 0.04028 | 0.3076 | 0.4135 | 0.6301 | UNIBO |
| chr5:131995059 | T | rs56162149 | *ACSL6* | 0.03 | 0.2053 | 0.1266 | 1.783 | SAS |
| chr21:33287378 | T | rs8178521 | *IFNAR2, IL10RB* | 0.0314 | 0.2868 | 0.4333 | 0.5259 | COVID-HOME |

***Table 3. SNPs.*** Polymorphisms with top associations with severity in UNIBO and SAS and susceptibility in COVID-HOME. F_A: risk allele frequency of affected group, oxygen therapy; F_U: risk allele frequency of the unaffected group, no oxygen therapy. For COVID-HOME cohort F_A means infected and F_U is not infected.

There is a lack of evidence for associations of CNVs, and poor evidence of HLA involvements.

The analysis of CNVs is still ongoing and follows the method used for SNVs; it will be reported in D6.15. Also, HLA alleles/haplotypes analysis is still ongoing but preliminary data indicate no association with severity and susceptibility in our cohort.

Regarding mtDNA, the analysis on the entire cohort is still ongoing; however, here we report preliminary data on 460 patients. We highlighted a significant difference in mtDNA copy number, which decreased in relation to the increase of disease severity, suggesting that most severe cases were relatively depleted of mtDNA, possibly reflecting into OXPHOS deficiency (**figure 2**). The haplogroup H frequency resulted to be significantly higher in mild patients, protecting from severe forms of COVID-19. Haplogroup H has been previously associated with sepsis as a protective background and it has been suggested that it underwent positive



ORCHESTRA has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101016167

selection from the great pandemics during middle age. This possibly explains why haplogroup H is currently the most frequent in the European population (**figure 3**).

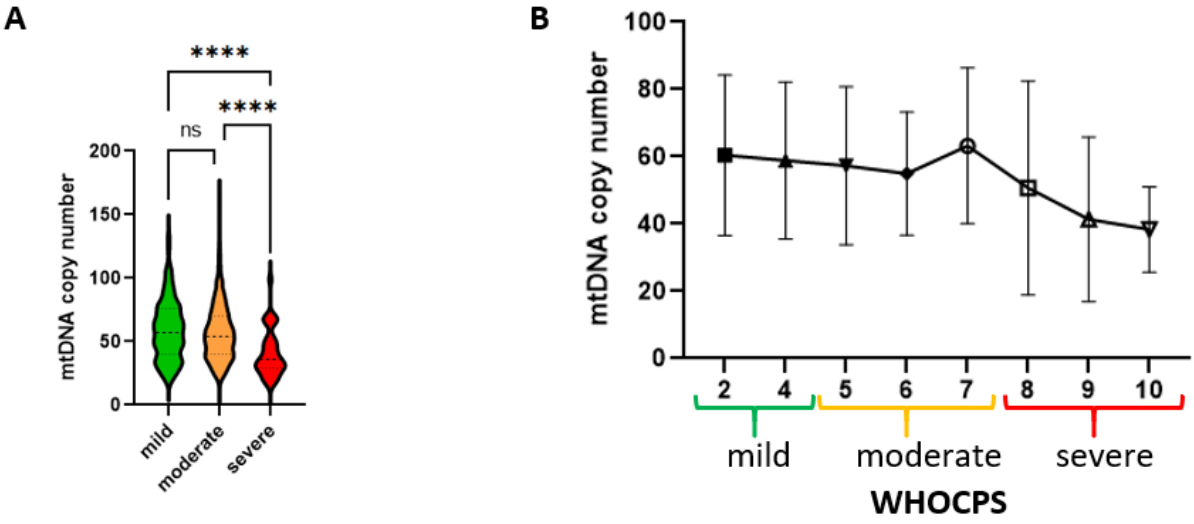


***Figure 2. mtDNA copy number.*** mtDNA copy number in COVID patients divided into groups) (A) and kept separate in each category (B).

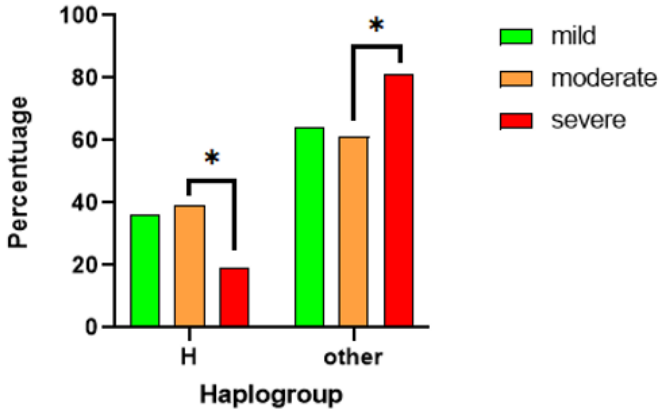


***Figure 3. Haplogroup frequency.*** Percentuage of patients with H and other haplogroups divided by WHO-CPS categories.

## INSERM analysis

### Sample description

We have completed the sequencing of 200 retrospective and 400 prospective patients from the WP2 French Covid cohort. Analysis of these patients did not identify any potential deleterious gene variants that were previously reported to be involved in severe COVID-19. To perform an enrichment analysis in a sufficient number of patients, and to have appropriate SARS-CoV-2 infected controls, we combined those data with previously available data from the COVID Human Genetic effort consortium for a final sample size of 3 269 severe cases and 1 373 infected controls.



ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

## Type I IFN gene analysis

No gene reached genome-wide significance. Under a recessive model, the most significant gene with at-risk variants was *TLR7,* with an OR of 27.68 (95%CI 1.5-528.7, p=1.1x10$^{-4}$) for biochemically loss-of-function (bLOF) variants. We replicated the enrichment in rare predicted LOF (pLOF) variants at 13 influenza susceptibility loci involved in TLR3-dependent type I IFN immunity (OR 3.70 [95%CI 1.3-8.2], p 2.1x10$^{-4}$) as shown in **table 4**.

This enrichment was further strengthened by:

1) adding the recently reported *TYK2* and *TLR7* COVID-19 loci, particularly under a recessive model (OR 19.65, 95%CI 2.1-2635.4, p=3.4x10$^{-3}$)
2) considering as pLOF branchpoint variants with potentially strong impacts on splicing among the 15 loci (OR=4.40, 95%CI:2.3-8.4, *P*=7.7x10$^{-8}$).

Finally, the patients with pLOF/bLOF variants at these 15 loci were significantly younger (mean age±SD 43.3±20.3 years) than the other patients (56.0±17.3 years; p=1.68x10$^{-5}$). Overall, we found that rare variants of TLR3- and TLR7-dependent type I IFN immunity genes can underlie life-threatening COVID-19, particularly with recessive inheritance, in patients under 60 years old. This work was published in Genome Medicine [1].

| Gene set | Cohort | Model | No. carriers | | Joint analysis | | Trans-pipeline meta-analysis | Trans-ethnic meta-analysis |
|---|---|---|---|---|---|---|---|---|
| | | | Cases | Controls | P value | OR [95%CI] | *P* value | *P* value |
| 13 genes [a] | Samples independent of [14] [b] | Co-dominant | 25 | 5 | 5.97x10$^{-3}$ | 3.21 [1.3-8.2] | 9.15x10$^{-3}$ | 0.01 |
| 13 genes | Full [c] | Co-dominant | 34 | 6 | 2.13x10$^{-4}$ | 3.70 [1.7-9.5] | 7.45x10$^{-4}$ | 6.52x10$^{-4}$ |
| 13 genes | Full | Heterozygous only [d] | 31 | 6 | 5.21x10$^{-3}$ | 3.11 [1.3-8.6] | 7.88x10$^{-3}$ | 5.98x10$^{-3}$ |
| 13 genes | Full | Recessive | 3 | 0 | 0.02 | 15.79 [1.4-2170.4] | 0.05 | 0.03 |
| 13 genes + *TYK2* | Full | Co-dominant | 37 | 7 | 1.40x10$^{-4}$ | 3.30 [1.6-7.8] | 5.77x10$^{-4}$ | 5.64x10$^{-4}$ |
| 13 genes + *TYK2* | Full | Heterozygous only | 32 | 7 | 0.02 | 2.53 [1.1-6.6] | 0.03 | 0.02 |
| 13 genes + *TYK2* | Full | Recessive | 5 | 0 | 3.36x10$^{-3}$ | 19.65 [2.1-2635.4] | 9.84x10$^{-3}$ | 0.03 |
| 13 genes + *TYK2* + bLOF *TLR7* | Full | Co-dominant | 57 | 9 | 1.27x10$^{-7}$ | 3.82 [2.0-7.2] | 1.99x10$^{-7}$ | 2.20x10$^{-6}$ |
| 13genes + *TYK2* + bLOF *TLR7* | Full | Heterozygous only | 32 | 9 | 0.04 | 2.27 [1.0-5.2] | 0.04 | 0.02 |
| 13genes + *TYK2* + bLOF *TLR7* | Full | Recessive | 25 | 0 | 4.69x10$^{-7}$ | 39.19 [5.2-5037.01] | 2.39x10$^{-6}$ | 6.66x10$^{-5}$ |
| 13genes + *TYK2* + bLOF *TLR7* + BP variants | Full | Co-dominant | 67 | 9 | 7.7x10$^{-8}$ | 4.40 [2.3-8.4] | 3.5x10$^{-8}$ | 6.5x10$^{-7}$ |

[a] IFNAR1, IFNAR2, IRF3, IRF7, IRF9, IKBKG, STAT1, STAT2, TBK1, TICAM1, TLR3, TRAF3 and UNC93B1

[b] 2718 patients and 1059 controls newly recruited and not screened in [14]

[c] The full cohort includes 3269 patients and 1373 controls

[d] In this model, only subjects with heterozygous variants are considered as carriers

## HLA analysis



ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

## HLA analyses

More recently, these patients were also used to perform an association analysis of HLA alleles in asymptomatic SARS-CoV-2 infection following a publication in Nature in August 2023 showing the role of HLA-B*15:01 in this phenotype [4]. We found a lack of association of classical HLA alleles, including HLA-B*15:01, with pre-Omicron asymptomatic SARS-CoV-2 infection in unvaccinated participants in a prospective population-based study in the US (191 asymptomatic vs. 945 symptomatic COVID-19 cases). Moreover, we found no such association in the international COVID Human Genetic Effort cohort including the Orchestra sequenced patients (206 asymptomatic vs. 574 mild or moderate COVID-19 cases and 1625 severe or critical COVID-19 cases). Detailed results for HLA-B*15:01 are shown in **figure 3**. As with other acute primary infections, no classical HLA alleles apparently favor an asymptomatic course of SARS-CoV-2 infection. This work is in medRxiv [5], and is under revision for publication in Human Genetics and Genomics Advances.

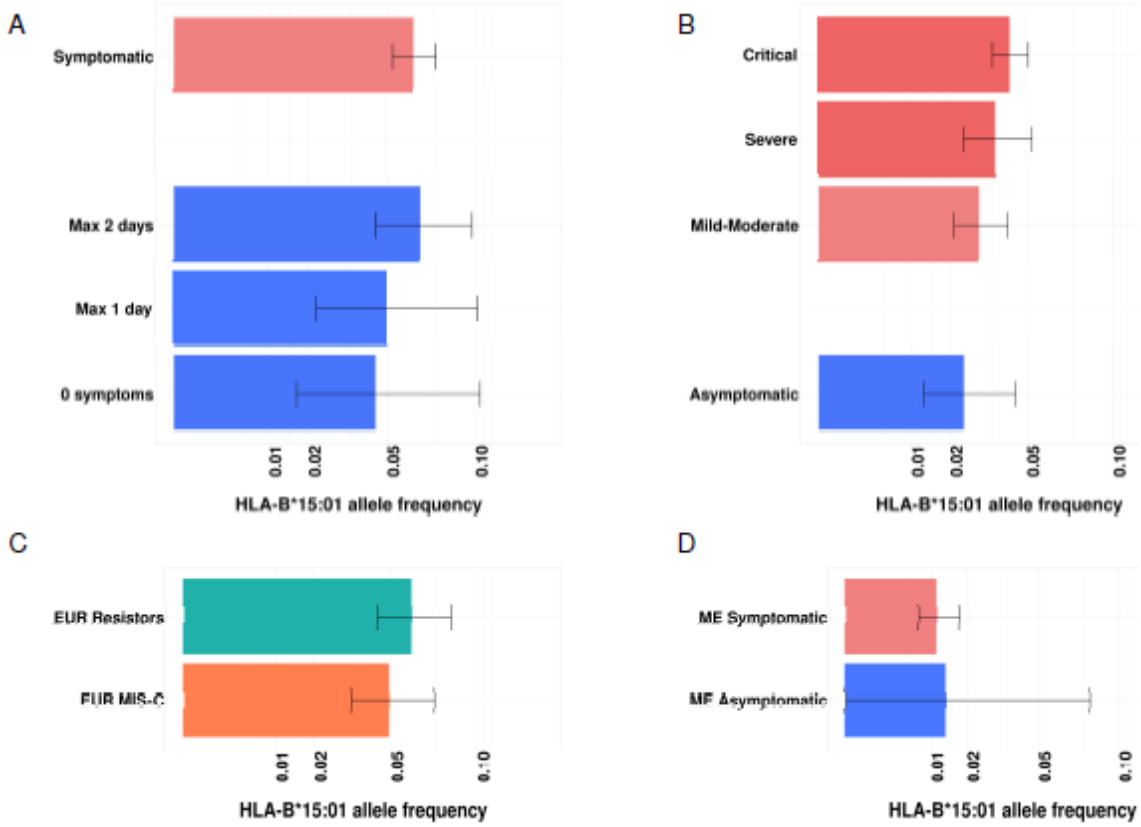


***Figure 4. HLA-B*15:01 is not enriched in asymptomatic cases.***

A: Allele frequency and 95% confidence intervals in the US prospective cohort subgroups.

B: Allele frequency and 95% CI in the CHGE European sample.

C: Allele frequency and 95% CI in individuals highly exposed to SARS-CoV-2 who never tested positive ('Resistors', n=291) and children with SARS-CoV-2 infection complicated by multisystem inflammatory syndrome ('MIS-C', n=235) from the European CHGE sample.

D: Allele frequency and 95% CI in Middle Eastern (ME) individuals from the CHGE cohort (Symptomatic, n=895; Asymptomatic, n=37).



ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

## Conclusions

The investigation of genomic variations associated with COVID-19 severity and susceptibility within the ORCHESTRA project yielded significant insights. Analysis of rare single nucleotide variants (SNVs) in type I IFN pathway immunity genes revealed mutations in patients with higher severity scores (WHO-CPS≥5). A combined data enrichment analysis revealed significant associations between rare variants in genes related to type I IFN immunity and severe COVID-19 outcomes, particularly in patients under 60 years old. Preliminary analysis of mtDNA revealed a significant decrease in mtDNA copy number with increasing disease severity, indicating a potential link between mitochondrial dysfunction and severe COVID-19 cases. Analysis of HLA alleles in asymptomatic SARS-CoV-2 infection did not find associations with classical HLA alleles, including HLA-B*15:01. This suggests that classical HLA alleles may not significantly influence the asymptomatic course of SARS-CoV-2 infection. Further investigation is warranted to elucidate the role of CNVs in COVID-19 outcomes.

Overall, the findings underscore the multifaceted nature of genetic factors influencing COVID-19 severity and susceptibility. Further research is needed to elucidate the complex interplay between host genetics and disease outcomes, which can inform strategies for prevention, treatment, and management of COVID-19.

# References


1. Matuozzo, D., Talouarn, E., Marchal, A. et al. Rare predicted loss-of-function variants of type I IFN immunity genes are associated with life-threatening COVID-19. *Genome Med* 15, 22 (2023). https://doi.org/10.1186/s13073-023-01173-8
2. Genomics England PanelApp, https://panelapp.genomicsengland.co.uk/panels/398/
3. Covid-19 Host Genetics Initiative Browser, https://www.covid19hg.org/
4. Augusto, D.G., Murdolo, L.D., Chatzileontiadou, D.S.M. et al. A common allele of HLA is associated with asymptomatic SARS-CoV-2 infection. Nature 620, 128–136 (2023). https://doi.org/10.1038/s41586-023-06331-x
5. Marchal A, Cirulli ET, Neveux I, Bellos E, Thwaites RS, Schiabor Barrett KM, Zhang Y, Nemes-Bokun I, Kalinova M, Catchpole A, Tangye SG, Spaan AN, Lack JB, Ghosn J, Burdet C, Gorochov G, Tubach F, Hausfater P, Effort CHG, Group COS, French CCSG, Co VCC, Clinicians C-S, Clinicians C, Orchestra Working Group, Amsterdam UMCC-B, Group N-UCS, Dalgard CL, Zhang SY, Zhang Q, Chiu C, Fellay J, Grzymski JJ, Sancho-Shimizu V, Abel L, Casanova JL, Cobat A, Bolze A. Lack of association between HLA and asymptomatic SARS-CoV-2 infection. (2023) *medRxiv* 10.1101/2023.12.06.23299623


# Acknowledgments


The WP6 Human Genomics group wants to thanks UNIBO, SAS, COVID-HOME, and the COVID Human Genetic Effort consortium for collecting and providing us patients' blood samples.





ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167