# Genetic Analysis of Severe COVID-19 Patients in Madrid Cohort

Whole exome analysis of 100 Spanish cohort of patients prior COVID-19 infection

during Madrid’s first wave.

# Aims

1. Identify variants associated with patients displaying severe symptoms
2. Identify variants associated with patients displaying mild/no symptoms
3. Evaluate the presence of genetic variants in a gene panel context analysis
4. Examine any differences of affected genes in sample subgroups (male/female; symptomatic/asymptomatic)

# Sample selection

We are aware that sex may have implications in the severity of symptoms and therefore will account for it when performing analyses.

Only European Iberian Spanish individuals are included, to minimise genetic variance not related to COVID-19 symptoms.

To control for the age factor contributing to severity in COVID-19, patients older than 60 were excluded. Only patients deemed able to give legal informed consent (>18 years old) were included. This makes a dataset of

* 25 whole exomes of severe COVID-19 women
* 25 whole exomes of severe COVID-19 men
* 25 whole exomes of asymptomatic COVID-19 women
* 25 whole exomes of asymptomatic COVID-19 men

# Analysis description

We will segregate patients according to the presence/ absence of severe symptoms and sex to find enrichment of common variants in any of these sample subgroups against any of the others.

Our analysis will encompass screening of three gene panels designed to offer the most comprehensive consensus on genetic contributions to infection susceptibility, immunity and disease severity:

* 1. ‘Panel A’, genes involved in **infection susceptibility and immunological response**;
  2. ‘Panel B’ encompassing a selection of genes suspected to be involved in (non-immune) COVID-19 **disease severity**;
  3. ‘Panel C’, **OMIM morbid genes** known to cause pathogenic phenotype if mutated (n>3,900).

# Implementation plan

The following stages are envisaged for the implementation of this analysis:

1. **Agreement of analysis plan**.
   1. All PIs asked to input their feedback on this proposal
   2. Discuss and incorporate changes to this proposal as agreed
2. **Setting up of the environment and storage**.
   1. Setting up of AWS server
   2. Install required analysis tools
   3. Plan and test the appropriate layers of storage and procedures for moving data across from the sequencing centre to this instance
3. **Data transfer**.
   1. Bring all variant data from the sequencing centre into the AWS server for analysis
4. **Quality control statistics**.
   1. Check the quality of variant call files
   2. Check reported sex of samples is matched in variant call file
   3. Check that samples are not family relatives
   4. Summarise sample and quality control statistics
   5. Present results to PIs
5. **Gene-based annotation**.
   1. Classify variants according to affected genes
   2. Infer type of variant: exonic, intronic, splicing, 3’-untranslated region (UTR), 5’-UTR, etc.
   3. When the variant is exonic, identify the functional role on protein coding and affected transcripts
   4. Assign variant annotations found in ClinVar and ExAC/gnomAD databases
   5. Explore / assign any other appropriate variant annotation algorithms
6. **Prioritisation of top signals**.
   1. Prioritise variants according to mutagenic impact and frequency
   2. Manually select variant candidates for classification
   3. Classify variants according to pathogenicity
   4. Select variants for reporting
   5. Provide list of selected variants for reporting for each patient
7. **Follow up search for causal and functional variants**.
   1. Research reported variants by clinicians
   2. Relate reported variants to patient phenotypes
   3. Relate reported variants to gene panels and sample subgroups
8. **Interpretation and analysis of results**.
   1. Agree, implement appropriate enrichment statistical test performed to check whether any particular variant is statistically overrepresented in any of the sample subgroups
   2. Identify variants associated with patients displaying severe symptoms
   3. Identify variants associated with patients displaying mild/no symptoms
   4. Evaluate the presence of genetic variants in a gene panel context analysis
   5. Examine any differences of affected genes in sample subgroups (male/female; symptomatic/asymptomatic)
   6. Write up and publication

Below is a Gantt chart with the estimated times for each of the stages presented here.



# Research outcomes

1. List of overrepresented pathogenic variants in patients with severe symptoms
2. List of overrepresented pathogenic variants in patients with mild/no symptoms
3. Documentation of biochemical functions in affected genes
4. Finding associations between affected genes and patient phenotypes
5. Identification of genetic variants in a gene panel context analysis
6. Reporting of differences in affected genes for sample subgroups (male/female; symptomatic/asymptomatic)
7. Selection of patients from samples for follow up whole genome sequencing

Annex

# Project design

UNIR has partnered with hospitals and institutions in Madrid to collect samples from patients diagnosed with COVID-19, following a strict clinical examination process to ensure that the recruited patients showed a radiological image compatible with a typical pneumonia and characteristic clinical symptoms (cough, fever, breathless, anosmia, dysgeusia, etc.) for >3 days of evolution and/or positive SARS-CoV-2 PCR from nasopharyngeal exudate or positive SARS-CoV-2 serology. It is therefore confirmed that every sampled patient has been SARS-CoV-2+ confirmed by PCR or additional clinical criteria. All individuals with chronic underlying diseases as recorded in medical reports, such as diabetes mellitus, high blood pressure, obesity (BMI >30), kidney insufficiency, cirrhosis, cancer, HIV infection, alcoholism, moderate-severe physical and/or mental disability, were excluded. Likewise, we removed from our series those taking medications with immunosuppressive effect for any reason (i.e. autoimmune diseases, etc.).

All severely affected patients (cases group; 50 samples) had a diagnosis of COVID-19 based on clinical presentation (radiological image compatible with atypical pneumonia and characteristic clinical symptoms (cough, fever, breathless, anosmia, dysgeusia, etc.) for >3 days of evolution) and/or positive SARS-CoV-2 PCR from nasopharyngeal exudate or positive SARS-CoV-2 serology (confirmed cases). Severe cases showed respiratory insufficiency with Saturation of oxygen/Fraction of inspired oxygen (SAFI) below 300. Patients that developed arterial thrombosis episodes, either systemic or pulmonary, and either during the acute or convalescent phase, were also considered severe. Our control group (50 samples) of COVID-19 patients was obtained from individuals with positive PCR for SARS-CoV-2 and/or seroreactivity for specific antibodies that did not need hospitalisation nor have vascular complications as described above.