Genomic determinants associated with SARS-CoV-2 virulence

P. Ruiz-Rodriguez[1], N. Chaturvedi[2], M.A. Bracho[1,3], Á. Chiner-Oms[4], S. Jiménez-Serrano[4], M. G. López[4], I. Carcino-Muñoz[4], P. Ruiz-Hueso[5], M. Torres-Puente[4], G. D'Auria[1,5], L. Martinez-Priego[5], SeqCOVID-SPAIN consortium, F. González-Candelas[1,3], I. Comas[3,4], M. Coscolla[1]

1.I[2]SysBio, University of Valencia-CSIC, FISABIO Joint Unit - Paterna (Spain)

2.ACCELOM, Antibes (France)

3.CIBER in Epidemiology and Public Health (Spain)

4.Instituto de Biomedicina de Valencia (IBV-CSIC) - València (Spain)

5.Sequencing and Bioinformatic Service of FISABIO (Valencian Region Foundation for the Promotion of Health and Biomedical Research - València (Spain)

31st ECCMID Online 9 – 12 July 2021

ESCMID EUROPEAN SOCIETY OF CLINICAL MICROBIOLOGY AND INFECTIOUS DISEASES

INTRODUCTION

SARS-CoV-2 genomic determinants of virulence should be defined and characterized in order to detect viral mutants that could pose a risk. Previous studies have shown the benefits of using Cycle threshold (Ct) as a surrogate of viral load to infer SARS-CoV-2 transmission.

OBJECTIVE

We have evaluated if we can detect viral mutations associated with high or low viral load, using Ct (Cycle threshold) of real-time PCR used for diagnosis as a surrogate.

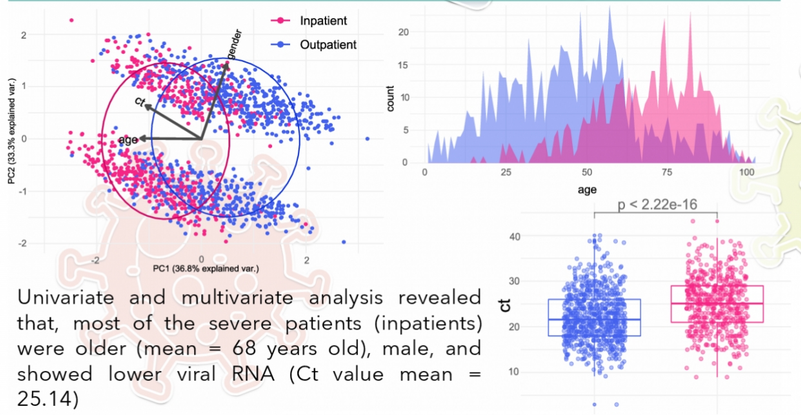
MATERIAL & METHODS

We have analyzed SARS-CoV-2 sequences from 6,102 individuals diagnosed using RT-PCR of viral cDNA in 48 hospitals in Spain, and these sequences are linked to clinical data such as sex, age, or clinical outcome for COVID-19 disease (hospitalization or intensive care).

We performed the analysis with two types of consensus sequences: 1) intra-host variable SNPs are determined when SNP shows alternative allele between 20-80% of the reads and 2) intra-host variable SNPs are defined when alternative allele is present between 5-20% of the reads.

We applied linear mixed model approaches (implemented in PLINK) to test for genetic associations of whole genome SNPs with viral load using two approaches: 1) high or low Ct value (threshold 27) and 2) numerical Ct value.

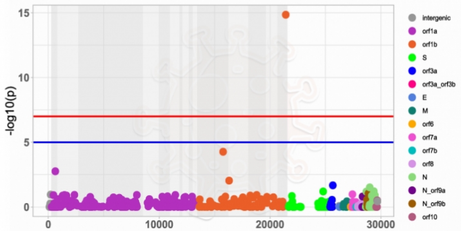
RESULTS I



Univariate and multivariate analysis revealed that, most of the severe patients (inpatients) were older (mean = 68 years old), male, and showed lower viral RNA (Ct value mean = 25.14)

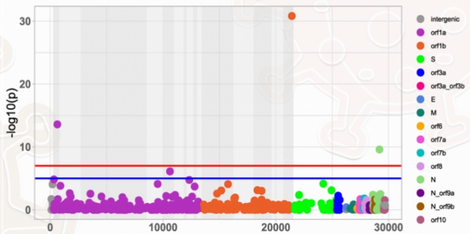
RESULTS II

INTRA-HOST SNPS > 20% reads



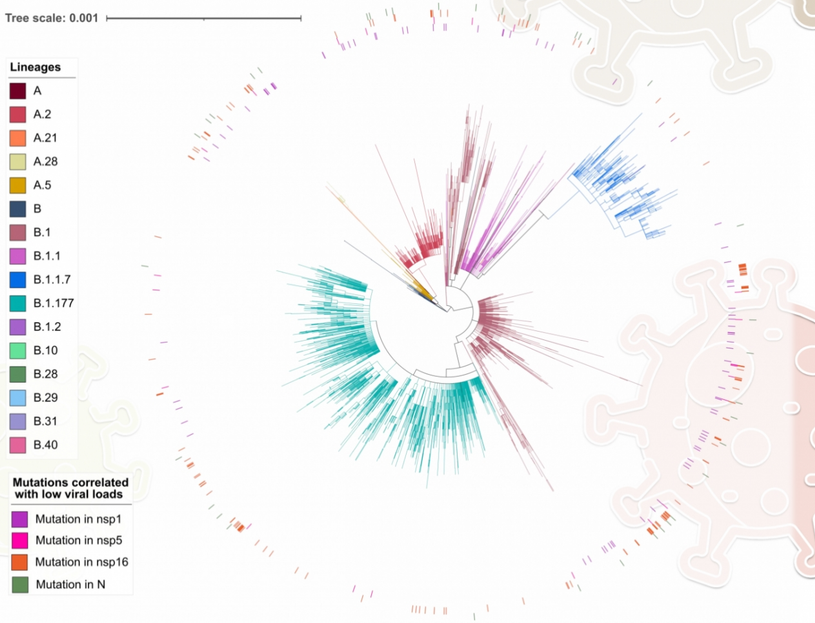
We obtained a correlation between one SNP and binarized Ct belonging to low viral loads. This missense SNP is located at orf1b which corresponds to nsp16 encoding a 2'-O-methyltransferase, crucial for RNA cap formation.

INTRA-HOST SNPS < 20% reads



We obtained a correlation between four SNPs and low Ct values. We found 2 missense SNPs in orf1a. First one affecting nsp1 that suppresses host innate immune functions, and second one in nsp5, a main protease of the virus. We detect the same SNP as previous analysis in nsp16. Finally, we detect one missense SNP in N gene, which encodes to nucleocapsid of the virus.

RESULTS III



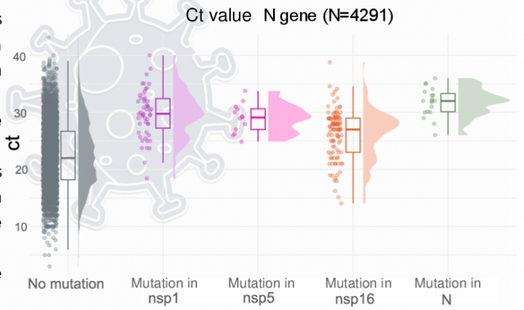
SNPs associated with lower viral loads appear independently over SARS-CoV-2 phylogeny, without fixing into population or forming clusters of transmission.

RESULTS IV

We found significant differences of Ct between individuals with the different mutations found in GWAS, except between nsp1 and nsp5 (Wilcoxon test p-value < 0.05).

Mutated sequences shows higher Cts compared with non mutated sequences for all these positions.

Mutation in nsp16 shows the lowest Ct values.



CONCLUSIONS

Spanish data from SEQCOVID consortium shows that the profile of hospitalized patients corresponds to older men with lower viral load, confirming that older people are more vulnerable to the virus.

We found 4 nonsynonymous SNPs correlated with low viral loads affecting orf1ab and N genes, 3 of them are at in intra-host frequencies lower than 20%, and one of them either is fixed in some individuals or shows a variability in 80-20% of the reads. These SNPs appear independently along SARS-CoV-2 phylogeny and do not form big clusters.

This work was funded by the Instituto de Salud Carlos III project COV20/00140 and COV20/00437, Spanish National Research Council project CSIC-COV19-021 and CSIC-COVID19-082, and the Generalitat Valenciana (SEJI/2019/011 and Covid_19-SCI). Action co-financed by the European Union through the Operational Program of the European Regional Development Fund (ERDF) of the Valencian Community 2014-2020. MC is supported by Ramón y Cajal program from Ministerio de Ciencia, grant RTI2018-094399-A-I00.