Impact of SARS-CoV-2 lineages and mutations on the survival of hospitalized patients

Carlos Loucera , **Javier Perez-Florido** , Carlos S. Casimiro-Soriguer , Francisco Ortuño , Rosario Carmona , Gerrit Bostelmann , Luis Javier Martinez-Gonzalez , Dolores Muñoyerro-Muñiz , Román Villegas , Jesus Rodriguez-Baño , Manuel Romero , Nicola Lorusso , Javier Garcia-León , Jose M. Navarro-Marí , Pedro Camacho-Martinez , Laura Merino-Diaz , Adolfo de Salazar , Laura Viñuela , Jose A Lepe , Federico Garcia , and Joaquin Dopazo

*

Jornadas Andaluzas de Bioinformática (JABI), 2023

Introduction

The combined use of SARS-CoV-2 genome sequences from the Andalusian surveillance circuit and detailed clinical information of the patients allowed us to assess the impact of both, the SARS-CoV-2 lineage and the mutations each virus harbors on mortality rate among patients hospitalized for COVID-19. This study illustrates how a combination of genomic and clinical data provide solid evidences that can help to entablish VOIs and VOCs.

The Project in numbers

SARS-CoV-2 genomes sequenced in Andalusia: near **36 500**
764 matched to BPS inpatients (February 2020 - April 2021)
18 `PANGO` lineages
5 `PANGO` lineages eligible for causal analysis: (**A**, **A.2**, **B.1**, **B.1.177** and **B.1.1.7**)
594 nucleotide mutations with respect to the SARS-CoV-2 reference genome
49 nucleotide mutations eligible for causal analysis

Bioinformatics

Sequencing data were analyzed using:
`nf-core/viralrecon` pipeline
SARS-CoV-2 reference **MN908947.3**
Genomic variants and consensus sequences were identified through `iVar` software:
Minimum allele frequency threshold of 0.75
Lineage assignment to each consensus genome was generated by `Pangolin` tool

Base Poblacional de Salud (BPS) clinical data

764 genomes could be matched to hospitalization events in BPS
Outcome: certified death event during first 30 days of hospitalization
Stays that imply one or more changes of hospital units are combined in a single stay
To reduce possible confounding effects, reinfections are excluded
Outcome confounders grouped to mititigate the curse of dimensionality:
sex, age, diabetes, circulatory, respiratory, neoplasms, dementia, anxiety and mood disorders, other mental diseases

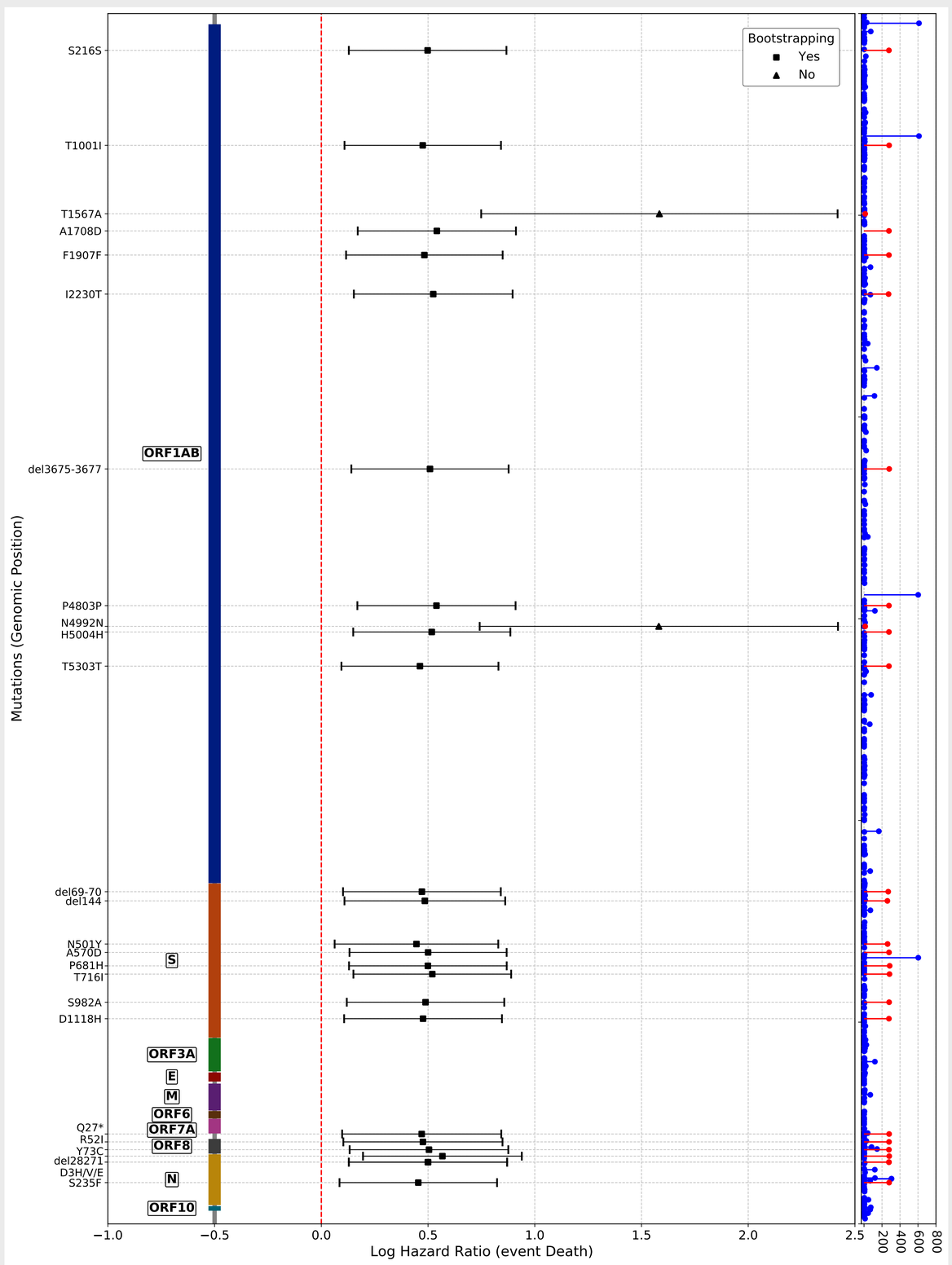
Modeling

For each mutation/lineage:
A covariate balance analysis to determinate the viability for a causal adjustment
A covariate-adjusted analysis using a closed-form estimator for the hazard ratio
An IPW covariate-adjusted bootstrapped (n=1001) hazard ratio

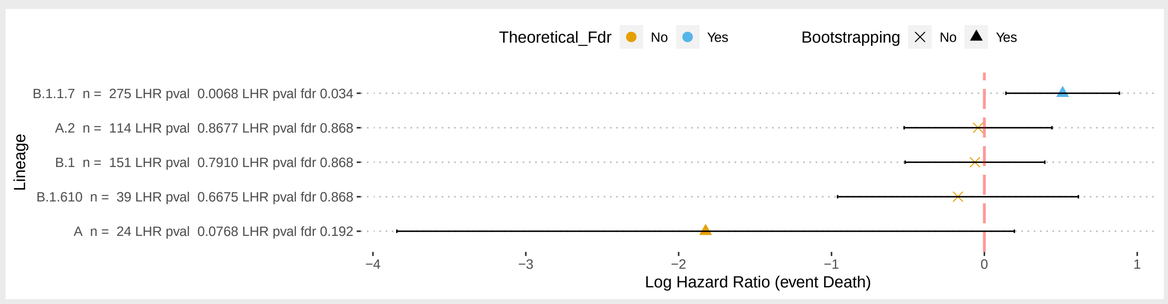
Conclusions

The combined use of SARS-CoV-2 genome sequences and clinical information allowed us to assess the mortality impact of lineage and mutations
Only the **B.1.1.7** lineage has rendered a significant impact on patient survival
Among 27 (**25 FDR**) mutations with significant association:
S:T716I, **ORF8:Y73C**, **ORF8:R52I**, **S:N501Y** and **ORF8:Q27*** affects PFAM motifs
ORF1ab:I2230T does not affect PFAM motif and it is uncorrelated to other variants
Some disrupted motifs are of unknown function, as PF19211 (NSP1) or PF12379 (NSP2)
ORF1ab:del3674-3676, **S:del69-70** and **S:del144** are present in **BA.1**
S:N501Y and **S:P681H** are present in **BA.1** and **BA.2**
Synonymous mutations with a significant impact are correlated to other coding mutations

Covariate-adjusted (closed and 1001 bootstrapped) LHR of the different mutations, plotted along the structure of the protein



Covariate-adjusted (closed and 1001 bootstrapped) LHR of the different lineages



Main author Funding

Junta de Andalucía
Consejería de Universidad, Investigación e Innovación
UNIÓN EUROPEA
Fondo Social Europeo
Andalucía se mueve con Europa

Main author Affiliations

Fundación Progreso y Salud
Consejería de Salud y Consumo
IBiS
INSTITUTODEBIOMEDICINADESEVILLA