Genetic diversity of the Ig Heavy chain locus in SARS-CoV-2 infection

Michela Biancolella[1], Patrizia Malaspina[1], Carla Jodice[1], Bianca Maria Ciminelli[1], Vito Luigi Colona[2], Andrea Latini[3], Francesca Leonardis[4], Paola Rogliani[4], Antonio Novelli[2], Giuseppe Novelli[4], Andrea Novelletto[1]

[1] *Department of Biology, Tor Vergata University, ITALY* [2]*Bambino Gesù Hospital, ITALY* [3]*Dept. Biomedicine and Prevention, Tor Vergata University, ITALY*
[4]*Tor Vergata University Hospital, ITALY*

This study was designed to explore the association between a genomic region relevant for B-cell maturation and antibody production and the outcomes of the Covid-19.

Early observations revealed that antibody production in SARS-CoV-2 infection is highly variable, in terms of Ig class and timing of seroconversion [1,2].

The Ig heavy chain locus (Figure 1) includes the gene segments encoding for the carboxy termini of different Ig classes as well as two enhancers named 3'RR1 and 3'RR2 [3]. The central portion of 3'RR1 is polymorphic in length, with strong disequilibrium with flanking variants [4]. Variation in this genomic region has implications on interindividual diversity in the ability to make Ig switches [5]. The polymorphisms located in the HS1.2 surrounding region (IgHA2, IgHE, IgHG4, IgHG2, 3'RR1, IgHA1) are poorly considered in the literature, while evidence is accumulating on the importance of HS1.2 as an enhancer in model organisms.

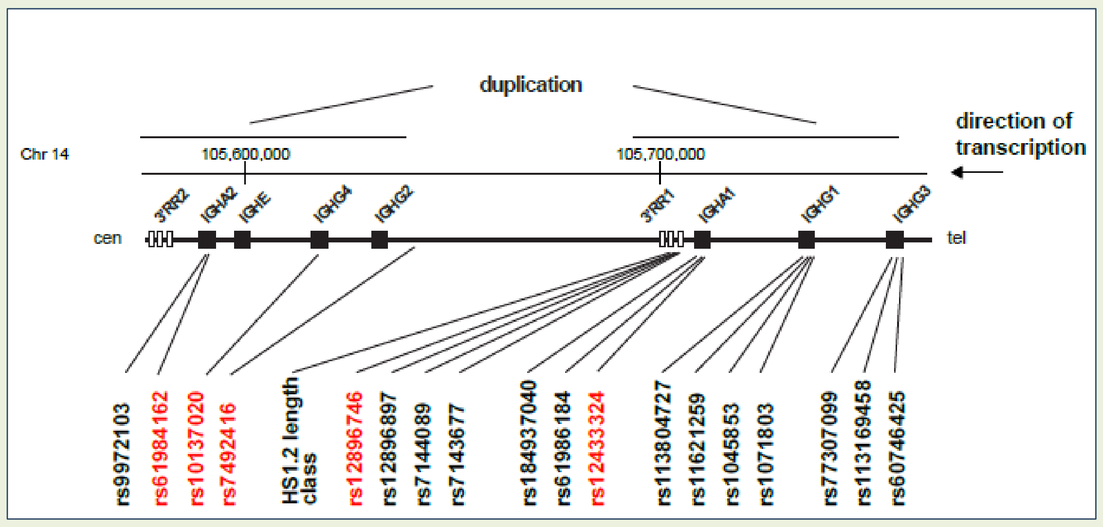

Figure 1. Map of the constant IgH gene segments, with a selection of variable positions associated with variable expression and/or amino acid substitutions. In red the SNPs examined in this work.

We genotyped for 5 SNPs the following cohorts of subjects (Table 1):

A) Covid-19 patients admitted and treated at the Microbiology and Virology unit of the Tor Vergata University Hospital, classified as Severe (N = 111) (suffering from respiratory failure, requiring invasive ventilation and intensive care unit admission), Moderate (N = 182) (showing respiratory impairment, requiring non-invasive ventilation and continuous positive airway pressure or bilevel positive airway pressure cycles) or Mild (N = 7) (paucisymptomatic, not requiring respiratory help).

B) Resistors (N = 139 unvaccinated individuals who, despite prolonged exposure, did not become infected). This category showed strong excess of young subjects (age <= 50).

C) Control subjects (N = 107) matched for population affiliation, as represented by in-house controls and Tuscans of the 1000 Genomes project (1KG).

Table 1. Results of genotyping - all subjects

COVID status	rs61984162 IGHA2		rs10137020 IGHG4		rs7492416		rs12896746 3'RR1		rs12433324 IGHA1	
	n. typed	MAF (A-Ref)	n. typed	MAF (G-ref)	n. typed	MAF (A-Alt)	n. typed	MAF (A-Ref)	n. Typed	MAF (G-Alt)
Resistor	139	6.8%	139	48.2%	139	42.1%	139	43.2%	137	31.4%
Mild	7	7.1%	7	64.3%	7	50.0%	7	57.1%	7	42.9%
Moderate	182	4.9%	180	38.1%	182	31.3%	182	34.3%	182	28.0%
Severe	111	5.4%	112	41.5%	111	29.7%	110	32.3%	106	27.4%
Tuscan controls (1KG)	107	2.8%	107	47.2%	107	41.1%	107	42.5%	107	32.2%
In-house controls	97	3.6%					97	31.9%	96	22.9%
		p=0.07		p=0.013		p=0.006		p=0.02		

HS1.2

Statistical analyses showed that allele frequency differences between resistor and moderate for rs10137020 and between resistor and severe patients for rs7492416 and rs12896746 were statistically significant.

The genotypic distribution of all SNPs was in accordance with Hardy Weinberg equilibrium

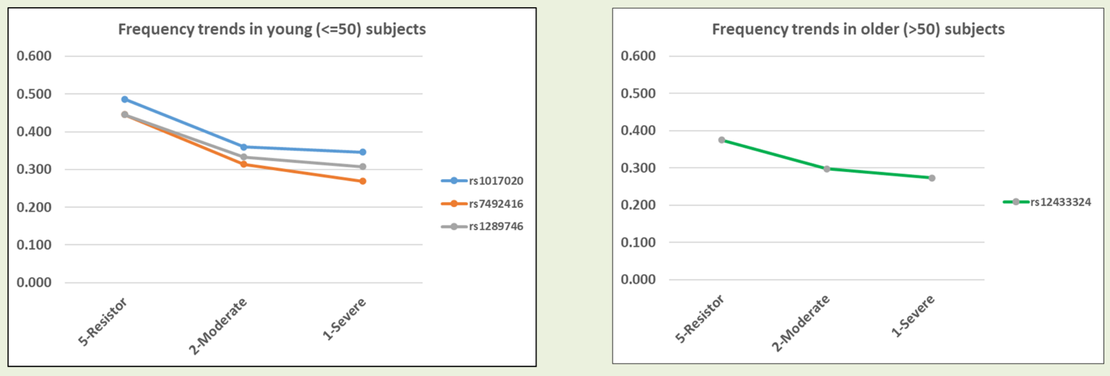

Figure 2. Frequency trends according to disease status

When subjects of all classes were partitioned according to age, three markers (rs1017020, rs7492416 and rs1289746) displayed a frequency trend in accordance with the severity of the disease in younger subjects(<=50), while one marker (rs12433324) showed a trend in older patients (>50) (Figure 2).

We also phased the 5 loci, to analyze associations at the level of IgH haplotype. The occurrence of the GGAAA haplotype declined significantly according to disease severity (Table 2).

Table 2. Haplotype frequencies - all subjects

		5-locus haplotype				
COVID status	n. typed	AGAAG	GGAAG	GGAAA	GGGGA	GAGGA
Resistor	139	2.9%	25.2%	12.2%	3.6%	49.3%
Mild	7	0.0%	35.7%	14.3%	7.1%	35.7%
Moderate	182	2.7%	20.3%	6.3%	4.1%	59.1%
Severe	110	2.3%	19.6%	4.1%	7.3%	56.4%
Tuscan controls (1KG)	107	2.8%	27.1%	8.4%	4.7%	49.5%
				p=0.001		

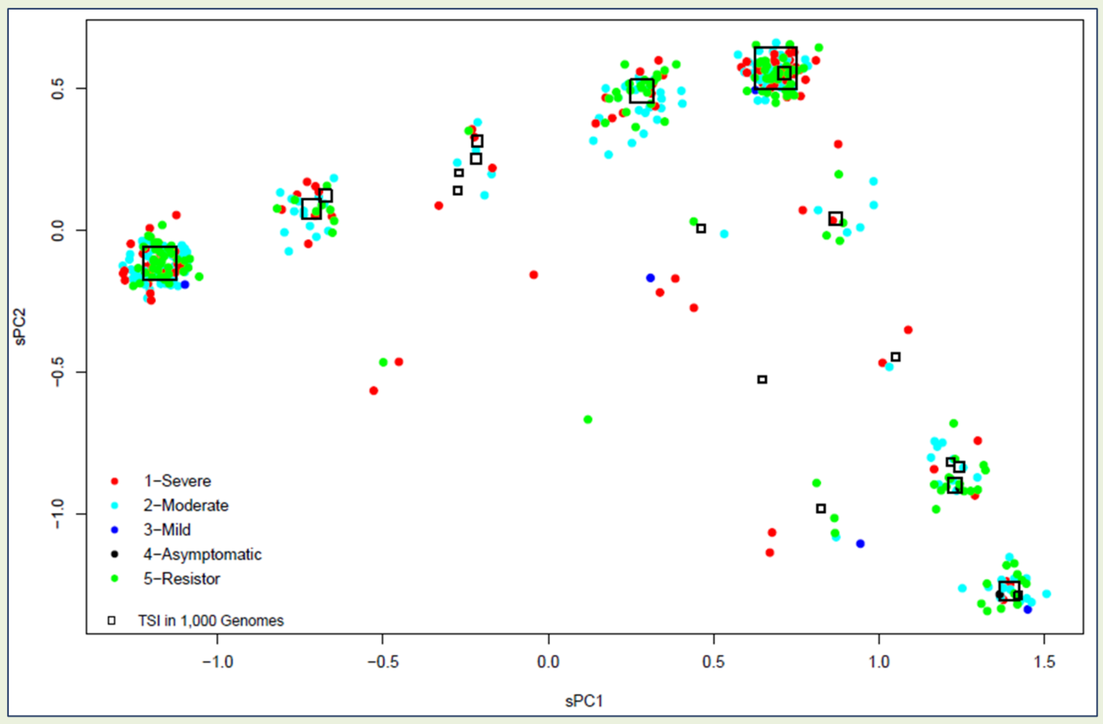

Figure 3. Plot of Sparse Principal Components 1 and 2 of 413 subjects colored according to clinical status and 107 Tuscans of the 1KG project.

When genotypes were represented in the space of sPC1 and sPC2 (Figure 3), the most numerous clusters were the same as detected in the 1KG Tuscans (controls). However, the cohort of severe patients turned out to be enriched in unusual genotypes, not recovered in the 1KG (central part of the plot). Haplotypes contributing to these genotypes (mainly GGGGA and GGGNN) are thus to be considered to confer risk for Covid-19 complications.

CONCLUSIONS

Our results show that:

- **No significant allele frequency differences were observed between sexes in any of the status category.**
- **The resistor cohort differed only marginally from control subjects.**
- **The two closest markers that flank HS1.2 centromerically and telomerically displayed significant shifts of allele frequencies in the cohort of severe patients as compared to resistors; this points to HS1.2 as a strong candidate to drive the observed association.**
- **Four markers displayed frequency trends according to severity.**
- **Carriers of haplotypes GGGGA and GGGNN are to be considered at high risk for a severe presentation**

Acknowledgments

This research was supported by European Commission UNDINE project, grant number 101057100

Funded by the Horizon Europe Framework Programme of the European Union
GA no.: 101057100

References

1) Long et al (2020) Nat Med 26:845.
2) Sun et al (2020) Emerg Microbes Infect 9:940.
3) D'Addabbo (2011) Evol Biol 11:71.
4) Frezza et al (2020) Am J Phys Anthropol 171:671.
5) Cianci et al (2023) Gene 30:862.