

Shorter androgen receptor polyQ alleles protect against life-threatening COVID-19 disease in males.

Margherita Baldassarri MD^{1,2,#}, Nicola Picchiotti MS^{3,4,#}, Francesca Fava MD^{1,2,5}, Chiara Fallerini PhD^{1,2}, Elisa Benetti MS², Sergio Daga PhD^{1,2}, Floriana Valentino MS^{1,2}, Gabriella Doddato MS^{1,2}, Simone Furini PhD², Annarita Giliberti MS^{1,2}, Rossella Tita PhD⁵, Sara Amitrano MS⁵, Mirella Bruttini MS^{1,2,5}, Susanna Croci MS^{1,2}, Ilaria Meloni PhD^{1,2}, Anna Maria Pinto MD⁵, Chiara Gabbi MD⁶, Francesca Sciarra PhD⁷, Mary Anna Venneri PhD⁷, Marco Gori PhD^{3,8}, Maurizio Sanarico MS⁹, Francis P. Crawley Mr¹⁰, Uberto Pagotto MD¹¹, Flaminia Fanelli PhD¹¹, Marco Mezzullo PhD¹¹, Elena Dominguez-Garrido PhD¹², Laura Planas-Serra MS^{13,14,15}, Agatha Schlüter PhD^{13,14,15}, Roger Colobran PhD¹⁶, Pere Soler-Palacin PhD^{16,17}, Pablo Lapunzina PhD^{14,18}, Jair Tenorio PhD^{14,18}, Spanish Covid HGE, Aurora Pujol PhD^{13,14,19}, Maria Grazia Castagna MD²⁰, Marco Marcelli MD²¹, Andrea M. Isidori PhD⁷, GEN-COVID Multicenter Study, Alessandra Renieri MD^{1,2,5*}, Elisa Frullanti PhD^{1,2,^}, Francesca Mari MD^{1,2,5,^}.

Co-first authors

^ Co-last authors

- 1) Medical Genetics, University of Siena, Italy
- 2) Med Biotech Hub and Competence Center, Department of Medical Biotechnologies, University of Siena, Italy
- 3) University of Siena, DIISM-SAILAB, Siena, Italy
- 4) Department of Mathematics, University of Pavia, Pavia, Italy
- 5) Genetica Medica, Azienda Ospedaliero-Universitaria Senese, Italy
- 6) Independent Medical Scientist, Milan, Italy
- 7) Department of Experimental Medicine, Sapienza University of Rome, Rome, Italy
- 8) Université Côte d'Azur, Inria, CNRS, I3S, Maasai
- 9) Independent Data Scientist, Milan, Italy
- 10) Good Clinical Practice Alliance-Europe (GCPA) and Strategic Initiative for Developing Capacity in Ethical Review-Europe (SIDCER), Leuven, Belgium
- 11) Unit of Endocrinology and Prevention and Care of Diabetes, Center for Applied Biomedical Research, Department of Medical and Surgical Sciences, University of Bologna, S. Orsola-Malpighi Hospital, Bologna, Italy
- 12) Molecular Diagnostic Unit, Fundación Rioja Salud, Logroño, La Rioja, Spain
- 13) Neurometabolic Diseases Laboratory, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain
- 14) CIBERER, Centro de Investigación Biomédica en Red de Enfermedades Raras, ISCIII, Melchor Fernández Almagro, 3, 28029 Madrid, Spain
- 15) Spanish Covid HGE
- 16) Immunology Division, Genetics Department. Hospital Universitari Vall d'Hebron. Vall d'Hebron Research Institute. Vall d'Hebron Barcelona Hospital Campus. Universitat Autònoma de Barcelona (UAB). Barcelona, Catalonia, Spain, EU
- 17) Pediatric Infectious Diseases and Immunodeficiencies Unit, Hospital Universitari

- Vall d'Hebron, Vall d'Hebron Research Institute, Barcelona, Spain
- 18) Institute of Medical and Molecular Genetics (INGEMM)-IdiPAZ, Hospital Universitario La Paz-UAM Paseo de La Castellana, 261, 28046 Madrid, Spain.
 - 19) Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain
 - 20) Department of Medical, Surgical and Neurological Sciences, University of Siena, Italy
 - 21) Department of Medicine, Baylor College of Medicine, Houston TX, USA

*** Corresponding author:**

Professor Alessandra Renieri
Medical Genetics Unit
University of Siena
Policlinico Le Scotte
Viale Bracci, 2
53100 Siena, Italy
Phone: +39 0577 233303
FAX +39 0577 233325
E.mail: alessandra.renieri@unisi.it

Boxes

Evidence before this study

We searched on Medline, EMBASE, and Pubmed for articles published from January 2020 to August 2020 using various combinations of the search terms “sex-difference”, “gender” AND SARS-Cov-2, or COVID. Epidemiological studies indicate that men and women are similarly infected by COVID-19, but the outcome is less favorable in men, independently of age. Several studies also showed that patients with hypogonadism tend to be more severely affected. A prompt intervention directed toward the most fragile subjects with SARS-Cov2 infection is currently the only strategy to reduce mortality. glucocorticoid treatment has been found cost-effective in improving the outcome of severe cases. Clinical algorithms have been proposed, but little is known on the ability of genetic profiling to predict outcome and disclose novel therapeutic strategies.

Added-value of this study

In a cohort of 1178 men and women with COVID-19, we used a supervised machine learning approach on a synthetic representation of the uncovered variability of the human genome due to poly-amino acid repeats. Comparing the genotype of patients with extreme manifestations (severe vs. asymptomatic), we found that the poly-glutamine repeat of the androgen receptor (AR) gene is relevant for COVID-19 disease and defective AR signaling identifies an association between male sex, testosterone exposure, and COVID-19 outcome. Failure of the endocrine feedback to overcome AR signaling defect by increasing testosterone levels during the infection leads to the fact that polyQ becomes dominant to T levels for the clinical outcome.

Implications of all the available evidence

We identify the first genetic polymorphism predisposing some men to develop a more severe disease irrespectively of age. Based on this, we suggest that sizing the AR poly-glutamine repeat has important implications in the diagnostic pipeline of patients affected by life-threatening COVID-19 infection. Most importantly, our studies open to the potential of using testosterone as adjuvant therapy for severe COVID-19 patients having defective androgen signaling, defined by this study as ≥ 23 PolyQ repeats and inappropriate levels of circulating androgens.

ABSTRACT

Background COVID-19 presentation ranges from asymptomatic to fatal. The variability in severity may be due in part to impaired Interferon type I response due to specific mutations in the host genome or to autoantibodies, explaining about 15% of the cases when combined. Exploring the host genome is thus warranted to further elucidate disease variability. **Methods** We developed a synthetic approach to genetic data representation using machine learning methods to investigate complementary genetic variability in COVID-19 infected patients that may explain disease severity, due to poly-amino acids repeat polymorphisms. Using host whole-exome sequencing data, we compared extreme phenotypic presentations (338 severe versus 300 asymptomatic cases) of the entire (men and women) Italian GEN-COVID cohort of 1178 subjects infected with SARS-CoV-2. We then applied the LASSO Logistic Regression model on Boolean gene-based representation of the poly-amino acids variability. **Findings** Shorter polyQ alleles (≤ 22) in the androgen receptor (*AR*) conferred protection against a more severe outcome in COVID-19 infection. In the subgroup of males with age < 60 years, testosterone was higher in subjects with *AR* long-polyQ (≥ 23), possibly indicating receptor resistance ($p=0.004$ Mann-Whitney U test). Inappropriately low testosterone levels for the long-polyQ alleles predicted the need for intensive care in COVID-19 infected men. In agreement with the known anti-inflammatory action of testosterone, patients with long-polyQ (≥ 23) and age > 60 years had increased levels of C Reactive Protein ($p=0.018$). **Interpretation** Our results may contribute to design reliable clinical and public health measures and provide a rationale to test testosterone treatment as adjuvant therapy in symptomatic COVID-19 men expressing *AR* polyQ longer than 23 repeats. **Funding** MIUR project “Dipartimenti di Eccellenza 2018-2020” to Department of Medical Biotechnologies University of Siena, Italy (Italian D.L. n.18 March 17, 2020). Private donors for COVID research and charity funds from Intesa San Paolo.

INTRODUCTION

This study was based on a preliminary descriptive analysis of a cohort of 35 hospitalized COVID-19 patients that was used to pinpoint a combined model of rare and common host genetic variants.¹ After this initial study, we extended the cohort to more than 1,000 Italian SARS-CoV-2 PCR-positive subjects.² Among them, 25% were asymptomatic, 25% were very severe, and the remaining presented with different levels of respiratory failure and multi-organ involvement.²

Alongside the mode of transmission, viral load, comorbidities, and demographic factors (such as age and sex), the host genetic background appears to play an important role in COVID-19 severity and progression.³ Data from classical methods of analysis, such as Genome-Wide Association Study (GWAS) or burden gene test on whole-exome sequencing (WES) data, have pinpointed a limited number of genetic factors that do not entirely explain how the host's genetic background contributes to COVID-19 severity⁴. Two groups have recently identified rare variants in the interferon type I pathway that are responsible for inborn errors of immunity in a small percentage of patients and auto-antibodies against type I interferon genes in up to 10% of with severe COVID-19^{5,6,7}. We hypothesized that most patients might instead have a more complex genetic background consisting of mutations arising in common susceptibility genes where additional private, rare, or low-frequency mutations provide the virus with the most favorable environment for replication, spread, and organ damage. Thus, we have considered the possibility that poly-amino acids repeat polymorphisms may contribute to COVID-19 severity. Our hypothesis was tested using a method of analysis based on a synthetic approach to genetic data representation. Specifically, we considered that polymorphisms in the polyQ tract of the Androgen Receptor (*AR*) may protect against the development of severe COVID-19. *AR* contains in its N-terminus domain a polymorphic polyQ tract, ranging between 9 and

36 repeated CAG units in the normal population⁸. *In vitro* and *in vivo* studies have demonstrated that the transactivation potential of AR is inversely correlated to repeat length and Q-tract length can significantly influence androgen-dependent physiological processes⁸⁻¹¹.

Several lines of evidence lead to the concept that androgens are relevant to both SARS-CoV-2 infection and COVID-19 disease presentation; however, they seem to have a Janus bifacial way of action^{12,13}. On one side, androgens promote the transcription of the *TMPRSS2* gene that encodes a serine protease known to prime the spike (S) protein of coronaviruses, facilitating viral entry into the cells¹⁴. On the other hand, hypogonadism and not hypergonadism is known to correlate with severe COVID-19¹⁵ and other chronic conditions, partly due to the loss of attenuation of the inflammatory immune response by testosterone¹⁶⁻¹⁸. Functional hypogonadism is also very common in male obesity and diabetes mellitus, two recognized factors of poorer prognosis in COVID-19 patients. The synthetic supervised genetic approach reported here contributes to disentangling this complex androgen-COVID-19 relationship. It may also have significant implications for further development of precision medicine in COVID-19 and identifies the use of androgens as a potential way to shape disease outcome as earlier proposed for glucocorticoids¹⁹, and subsequently demonstrated in a clinical trial²⁰.

RESULTS

Testing the role of common poly-amino acid repeat polymorphisms in COVID-19 outcome

In order to test the role of common poly-amino acid repeat polymorphisms in determining COVID-19 clinical severity, we performed a nested case-control study (NCC), selecting the extreme phenotypic ends of our entire GENCOVID cohort (**Table 1**). (**Figure 1**). Among 18,439 annotated genes we selected those with amino acid repeats,

namely 43 genes, and represented them as a boolean. Logistic regression with LASSO regularization analysis identified *AR* as the only protective gene (**Figure 1, panel A**). As expected, the grid search curve of the cross-validation score (**Figure 1, panel C**) shows a maximum for an intermediate value of the L1 regularization parameters, and the chosen parameter is 6.31. With this calibration setting, the 10-fold cross-validation provides good performances in terms of accuracy (77%), precision (81%), sensitivity (77%), and specificity (78%) as shown in **Figure 1, panel D**. The confusion matrix is reported in **Figure 1, panel B**, whereas the Receiver Operating Characteristic (ROC) curve (**Figure 1, panel E**) provides an Area Under the Curve (AUC) score of 86%.

Validation of polyQ polymorphism by sizing the PolyQ repeat of the *AR* receptor gene

In order to validate the results on *AR* obtained by LASSO logistic regression, we seized the male subset (351 subjects) using the gold standard technique that uses a fluorescent PCR reaction followed by the use of GeneScan Analysis software[®] (Applied Biosystems) for sizing the number of triplets²¹. Based on the *AR* polyQ length, male patients were subdivided into two categories, those having a number of PolyQ repeats less than or equal to 22 repeats, and those having a number of PolyQ repeats greater than or equal to 23 repeats, being 23 repeats the reference sequence on genome browsers. We found that PolyQ repeats below 22 are enriched in the asymptomatic cohort of males. The difference was statistically significant in the group of males younger than 60 years of age in which genetic factors are expected to have a major impact (p-value 0,024) (**Table 2; Supplementary Table S1; Supplementary Table S2**).

Validation of polyQ polymorphism in the Spanish Cohort

We then seized the polyQ repeat in an independent cohort of 158 males from Spain, with age < 60 years and without known comorbidities (117 cases and 41 controls). The association with shorter repeats (≤ 22) and protection was confirmed (p-value 0,014139 by χ^2 test) (**Table 3**).

Males with longer polyQ have receptor resistance

To functionally link the length of the PolyQ repeats to AR functionality, we measured total testosterone (TT) in 183 men using LCMS/MS (**Supplementary Table S1**). TT was higher in patients carrying ≥ 23 vs ≤ 22 glutamines (13.45 vs 11.23 nmol/L, p-value of 0.0422), reflecting reduced negative feedback from the less active receptors present in patients carrying a PolyQ repeat of ≥ 23 (**Fig. 2a**).

Unbalanced Testosterone-AR axis in males with longer polyQ repeats

Hormonal status of the entire male set sized for polyQ according to the age group revealed lower TT and calculated free T levels and higher SHBG levels with increasing age (**Supplementary Table S3**).

To evaluate whether the AR receptor reduced activity resulted in signs and symptoms of hypogonadism, subjects were interviewed, post-infection, using a modified version of the Androtest®²². Interviews were available for 61 subjects (representative of the extremes genotypes of the cohort: 43 with ≤ 19 repeats and 18 with ≥ 25 repeats). An Androtest score ≥ 8 was found in 38% of men with ≥ 25 repeats as compared to 16% of those with ≤ 19 glutamines (likelihood ratio, p=0.046). Similarly, cryptorchidism (11% vs. 2%), anemia (11% vs 2%) and severe erectile dysfunction (22% vs. 9%) were more frequently reported in subjects with ≥ 25 repeats, but not osteopenia/osteoporosis (6% vs 7%). These results indicate a trend toward clinical hypogonadism for those with longer repeats. Conversely, in the entire male dataset, 6 cases of prostate cancer were found

annotated in the past-medical history, all in the ≤ 22 glutamines group, suggesting an increased prostate sensitivity to androgens. No difference was found in the prevalence of BPH or 5- α -reductase inhibitors use.

As the reduced signal transduction of AR might be compensated by higher testosterone levels, as shown in **Fig 2a**, we tested whether the feedback was sufficient to overcome polyQ repeats. A binomial logistic regression was performed to ascertain the interaction between age, testosterone levels and polyglutamine receptor length on the likelihood that subjects require intensive care during COVID infections. The logistic regression model was highly significant ($\chi^2(3) = 18,881$, $p < 0.0001$), with the model explaining 7.5% (Nagelkerke R^2) of the variance in COVID19 outcome (**Supplementary Table S4**). A significant testosterone by polyglutamine length interaction was found ($p = 0.018$), suggesting impaired feedback as predictor of the worst outcome, namely intubation or CPAP/BiPAP versus hospitalization not requiring respiratory assistance. To provide a graphical representation, we plotted the ratio between TT serum concentrations and polyQ number vs. clinical outcome (**Fig. 2b**). Results show a decreased mean ratio, a sign of an inappropriate rise of TT for increasing polyglutamine repeats, was associated with a worse outcome ($p = 0.00038$).

Older severely affected males with longer polyQ repeats display a more pronounced inflammatory phenotype

Finally, we tested the relationship between the AR polyQ repeat size and 5 laboratory markers of immunity/inflammation, including CRP, Fibrinogen, IL6, CD4 and NK count. We found that older (≥ 60) males with AR polyQ tract ≥ 23 have a higher (55.92 versus 48.21 mg/dl) mean value of CRP ($p = 0.018$) and lower mean value of Fibrinogen and a trend of higher IL6 (**Table 4**).

DISCUSSION

We employed machine learning methodologies to identify a set of genes involved in the severity of COVID-19. In the presence of very high dimensionality, as for instance in a WES study, it is crucial to select the most predictive genes representing patterns of variation (mutations or variants) in subjects with different classes of response (i.e., disease state: from asymptomatic to severe cases). This problem is even more complex in diseases where multiple genes are involved in determining the severity and clinical variability of the pathology. Here, we wanted to represent poly-amino acids repeat polymorphisms that are typically missed in classical GWAS analysis, which concentrates on bi-allelic polymorphisms.

We used a machine learning approach and logistic regression with a LASSO regularization to test if using such a simplified representation could lead to a reliable prediction of extreme clinical outcomes (asymptomatic versus severely affected). This approach enabled us to predict such clinical outcomes with 77% sensitivity.

AR contains a highly variable polyglutamine repeat (poly-Q) located in the N-terminal domain of the protein, spanning from 9 to 36 glutamine residues in the normal population⁵. AR polyQ length correlates with receptor functionality, with shorter polymorphic glutamine repeats typically associated with higher and longer PolyQ tracts with lower receptor activity.⁵ AR is expressed in both males and females, but the bioavailability of its ligands testosterone (T) and dihydrotestosterone (DHT) differs significantly, being much higher in males. As previous studies linked male hypogonadism to a poorer outcome we decided to focus on male patients and demonstrated that shorter polymorphic glutamine repeats (≤ 22) confer protection against life-threatening COVID-19 in a subpopulation of individuals with age <60 years.

We also confirmed the functional association between polyQ size and receptor activity. Specifically, we showed that longer polyQ size (≥ 23) is associated with higher serum T levels, suggestive of impaired negative feedback ($p=0.0046$ at Mann-Whitney U test) at the level of the hypothalamus and pituitary gland, therefore confirming the association between functional hypogonadism and severe COVID-19 infection.

As T is known to have an immunomodulatory activity attenuating inflammatory immune responses,²³⁻²⁷ we hypothesized that a long PolyQ repeat would lead to a pro-inflammatory status heralded by increased proinflammatory markers^{18,28} by conferring decreased AR transcriptional activity. Conversely, men with a more active receptor (short PolyQ tract) would be protected because they can tame the inflammatory response and increase survival regardless of serum T levels. We found that -CRP-, one of the main inflammatory markers, was higher in subjects with a long AR PolyQ tract. This observation not only is in line with the known anti-inflammatory function of testosterone, but also reinforces the functional importance of the AR PolyQ tract and its association with COVID-19 clinical outcome. Furthermore, this observation suggests that CRP is hierarchically more indicative than serum T level, which can be inappropriately normal and mask a functional hypogonadism in men with a long PolyQ repeat.

The allele distribution of the PolyQ repeat length varies among different populations, with the shortest in Africans, medium in Caucasians, and longest in Asians²⁹. Interestingly, WHO data on mortality rates during the first pandemic wave indicated a higher fatality rate in China and Italy (<https://covid19.who.int/>)³⁰ with respect to African. Hence, AR polyQ length variability could represent an explanation for the observed differences in death rate. Moreover, Africans seem to be more prone to infection³¹. This observation could be due to a more active AR receptor, leading to a higher expression of *TMPRSS2*, a protease essential for SARS-CoV-2 spread¹⁴.

Different studies have shown an association between hypogonadism and severe COVID-19¹⁵ and other chronic obstructive pulmonary diseases^{16,17}. Our results are in line with these initial observations and contribute to explain a possible mechanism leading to these associations. The present study brings the observations to the upper level, revealing that is the overall androgenic effect -resulting from the interaction of polyQ polymorphism and circulating testosterone levels- to predict the need for intensive care. In infected men, we observed impaired feedback no longer sufficient to compensate for the reduced AR transcriptional activity, leading to the fact that polyQ becomes dominant to T levels. The latter helps to solve some inconsistencies, including the early reports of a slightly better outcome in prostate cancer patients -who tend to have lower polyQ, as in our cohort - when compared to other cancers.

An improvement in peak oxygen saturation in men receiving testosterone replacement therapy has been demonstrated in a randomized controlled trial³² and could be one of the mechanisms responsible for the observed protective effect of AR's with shorter polyQ tract in COVID-19 patients. Thus, our study has important implications and suggests that androgen therapy could be tested as an adjuvant treatment in COVID-19-infected men with functional hypogonadism. Based on the evidence discussed in this paper, a proper randomized controlled trial (RCT) is warranted in COVID-19 male patients with signs of hypoandrogenism and longer AR polyQ tract to test the safety and efficacy of AR agonists, like testosterone, in improving outcome. A simple genetic test measuring the AR receptor PolyQ repeat in male patients could then be introduced to screen individuals more likely to benefit from testosterone therapy.

Variants of another X-linked gene, TLR7, have been associated with severe COVID-19 outcomes in young men⁶. In the 2 reported families, the rare TLR7 mutations segregated as a highly penetrant monogenic X-linked recessive trait. While variants in TLR7 gene are expected to account for a small number of severely affected cases, our

findings involve a much larger number of subjects, as although the OR is relatively small, long polyQ alleles are relatively common (27%)³³. Overall, X-linked genetic variants keep coming up as important for defining severe COVID-19 cases in males.

In conclusion, we present a method that can predict if subjects infected by SARS-CoV-2 are at risk for life-threatening complications. This approach has 77% accuracy, 81% precision, 77% sensitivity, and 78% specificity. Furthermore, we present evidence suggesting that a more active AR has the potential to confer protection against COVID-19 severity. If confirmed, these observations should be followed by properly conducted clinical trials studying if testosterone helps decrease morbidity and mortality in patients affected by the most severe forms of the disease. Finally, as shown by regression analysis, ORs ranges between 1.26 and 1.45, therefore the risk of carrying a longer AR is much smaller than other already known strong predictors such as age and sex, but still is highly significant, relatively common, and among the very few known genetic predictors of COVID19 outcome.

METHODS

Patients.

We performed a nested case-control study (NCC). Cases and controls were drawn from the Italian GENCOVID² cohort of 1178 subjects infected with SARS-CoV-2 diagnosed either by RT-PCR on nasopharyngeal swab, or by serology test.² Demographic characteristics of patients enrolled in the cohort are summarized in Table 1 according to their clinical status. In particular, the cohort was subdivided into 5 groups, based on the type of the respiratory and medical support that was given to the patients: *i.* endotracheal intubation; *ii.* CPAP-biPAP ventilation; *iii.* oxygen administration; *iv.* hospitalization without respiratory support; *v.* no hospitalization either because asymptomatic or with minor symptoms of COVID-19 (mild fever, cough, sore throat, etc.). Subjects of the cohort

were recruited by Italian hospitals and primary care clinics participating in the GENCOVID Multicenter study previously described ².

In the current NCC study, cases were selected according to the following inclusion criteria: *i.* CPAP/biPAP ventilation (230 subjects); *ii.* endotracheal intubation (108 subjects). As controls, 300 subjects were selected using the sole criterion of not requiring hospitalization. Cases and controls show the extreme phenotypic presentations of the GENCOVID cohort. Exclusion criteria for both cases and controls were *i.* SARS-CoV-2 infection not confirmed by PCR; *ii.* not-caucasian ethnicity. Demographic characteristics of the subjects in the NCC study are summarized in Table 1.

Analysis of triplets size in the AR locus

To establish allele sizes of the polymorphic triplet in the AR locus, we used the HUMARA assay with minor modifications²¹. Specifically, we performed a fluorescent PCR followed by capillary electrophoresis on an ABI3130 sequencer. Allele size was established using the Genescan Analysis software.

Binary representation of WES data

In this paper, one of several synthetic approaches to genetic data variability representation is used: poly-amino acids repeat polymorphisms (Benetti et al., Topological Data Analysis on boolean representation of genome variability as a method for discovering the genetic bases of complex disorders. Paper in preparation). Reads were mapped to the hg19 reference genome by the Burrow-Wheeler aligner BWA. Variants calling was performed according to the GATK4 best practice guidelines. Namely, duplicates were first removed by MarkDuplicates, and base qualities were recalibrated using BaseRecalibration and ApplyBQSR. HaplotypeCaller was used to calculate Genomic VCF files for each sample, which were then used for multi-sample calling by GenomicDBImport and

GenotypeGVCF. In order to improve the specificity-sensitivity balance, variants quality scores were calculated by VariantRecalibrator and ApplyVQSR, and only variants with estimated truth sensitivity above 99.9% were retained. Variants were annotated by ANNOVAR. WES data were represented in a binary mode on a gene-by-gene basis.

Representation of poly-amino acids triple repeats: A set of binary variables was included to correctly describe non-frameshift variants in gene portions with repeated triplets. A total of 43 genes with triplet repeat variability were taken from UniProtKB (<https://www.uniprot.org/uniprot/?query=reviewed:yes%20keyword:%22Triplet%20repeat%20expansion%20%5BKW-0818%5D%22>). For any of these genes a variant D_{ij} was defined as equal to 1 if the i -th gene presented a deletion in the region characterized by repeated triplets. Non-informative genes, with either all 0 or all 1, are removed from the dataset.

LASSO logistic regression

The problem of extracting knowledge on the most relevant genes involved in the classification tasks of COVID-19 disease can be interpreted in the classical framework of feature selection analysis. Due to the specificities of the problem (i.e., the complex classification task), we adopted the embedded method of the LASSO logistic regression model. By denoting with β_k the coefficients of the logistic regression and by lambda (λ) the strength of the regularization, the LASSO (Least Absolute Shrinkage and Selection Operator) regularization³⁴ term of the loss, $\sum_{k=1}^p |\beta_k|$, has the effect of shrinking the estimated coefficients to 0. This provides a feature selection method for sparse solutions within the classification tasks able to enforce both the sparsity and the interpretability of the results. The weights of the logistic regression algorithm can be interpreted as the feature importances of the subset of the most relevant features for the task³⁵. The input features are the poly-amino acids triplet repeats presented in the previous section as well as

gender, comorbidity (1 if there is at least one comorbidity) and the age, the latter as a continuous variable normalized between 0 and 1.

The fundamental hyper-parameter of the logistic regression algorithm is the strength of the LASSO term, which is tuned with a grid search method on the average area under the Receiver Operating Characteristic (ROC) curve for the 10-fold cross-validation. The regularization hyperparameter varies in the range $[10^{-3}, 10^2]$ with 50 equally spaced values in the logarithmic scale. The optimal regularization parameter is chosen by selecting the most parsimonious parameter whose cross-validation score falls in the range of the best one along with its standard deviation. This choice is related to the aim of the work that is to select the most important genes (and not necessarily the entire set of genes contributing to COVID-19 variability). During the fitting procedure, the class slight unbalancing is tackled by penalizing the misclassification of the minority class with a multiplicative factor inversely proportional to the class frequencies. The data pre-processing was coded in Python, whereas for the logistic regression model the scikit-learn module with the liblinear coordinate descent optimization algorithm was used. Performances of the model were evaluated using the cross-validation confusion matrix as well as by computing precision, sensitivity, specificity, and the Receiver Operating Characteristic (ROC) curve.

Total Testosterone measurement.

Blood samples were collected after an overnight fast, immediately centrifuged at 4 °C and stored at -20 °C until assayed. Blood samples were collected after an overnight fast, immediately centrifuged at 4 °C and stored at -20 °C until assayed. Serum total testosterone was measured using the Access testosterone assay (Beckman Coulter Inc., Fullerton, CA, USA) with a minimum detection limit of 0.35 nmol/L. Reference range for this assay was 6.07-27.1 nmol/L. Thawed plasma underwent 15 min incubation at 56°C for

virus inactivation. Afterward, plasma total testosterone was measured by liquid chromatography - tandem mass spectrometry (LC-MS/MS) according to a previously validated method provided with reference values between 9.8 - 28.4 nmol/L³⁶. In the present study, measurement was performed in 100 µl plasma, with sensitivity limit being 0.270 nmol/L, imprecision ranging 9.8 to 0.7% and accuracy 90.6 to 101.5% at concentration levels between 1.12 and 39.2 nmol/L. A stability test under viral inactivation conditions was performed in 6 samples, revealing a testosterone mean (min-max) % loss of 9.7% (4.6-16.7%). SHBG levels were measured in plasma samples using Quantikine ELISA Kit (DSHBG0B, R&D Systems, Minneapolis, MN, USA) according to the manufacturers' instructions. Serum LH was measured using “Access LH assay “ a chemiluminescent, two-step enzyme immunoassay (Beckman Coulter Inc., Fullerton, CA, USA). Sensitivity for the LH determination is 0.2 mIU/mL. Reference range in adult males for this assay is 1.2-8.6 mIU/mL

Statistical analysis.

Since serum and plasma Testosterone values were not normally distributed, the statistical analyses were performed using non-parametric tests. In particular, we used the Mann-Whitney U test to compare testosterone levels in males with AR long-polyQ (≥ 23) versus males with short polyQ (≤ 22). Logistic regression analysis was performed to ascertain the interaction between age, testosterone, and the number of polyglutamine repetitions and COVID19 outcomes. Box-Tidwell procedure was used to assess linearity and the Hosmer and Lemeshow to assess goodness of fit test.

ACKNOWLEDGEMENTS

This study is part of the GEN-COVID Multicenter Study, <https://sites.google.com/dbm.unisi.it/gen-covid>, the Italian multicenter study aimed at

identifying the COVID-19 host genetic bases. Specimens were provided by the COVID-19 Biobank of Siena, which is part of the Genetic Biobank of Siena, member of BBMRI-IT, of Telethon Network of Genetic Biobanks (project no. GTB18001), of EuroBioBank, and of RD-Connect. We thank the CINECA consortium for providing computational resources and the Network for Italian Genomes (NIG) <http://www.nig.cineca.it> for its support. We thank private donors for the support provided to A.R. (Department of Medical Biotechnologies, University of Siena) for the COVID-19 host genetics research project (D.L n.18 of March 17, 2020). We also thank the COVID-19 Host Genetics Initiative (<https://www.covid19hg.org/>), MIUR project “Dipartimenti di Eccellenza 2018-2020” to the Department of Medical Biotechnologies University of Siena, Italy. We also thank Intesa San Paolo for the 2020 charity fund dedicated to the project N. B/2020/0119 “Identificazione delle basi genetiche determinanti la variabilità clinica della risposta a COVID-19 nella popolazione italiana”.

ETHICS APPROVAL

The GEN-COVID study was approved by the University Hospital of Siena Ethics Review Board (Protocol n. 16917, dated March 16, 2020). This observational study has been inserted in www.clinicaltrials.org (NCT04549831). The Spanish CovidHGE cohort is under IRB approval PR127/20 from Bellvitge University Hospital, Barcelona Spain.

AUTHOR CONTRIBUTIONS STATEMENT

EF, FM, AR designed the study. CF and IM, were in charge of biological samples' collection and biobanking. MB, FF were in charge of clinical data collection. MB, FF, AR, and FM performed analysis/interpretation of clinical data. UP, FF and MM performed testosterone measurement by LC-MS/MS. EDG, AS, AP and LPS performed the

validation of association between shorter repeats and protection in a Spanish cohort. MM and AI critically reviewed the manuscript and interpreted clinical data/androgen physiopathological processes. SA and MB were in charge of DNA isolations from peripheral blood samples. FV, GD, AG, RT carried the sequencing experiments. EB, NP, SF, CG, MG and MS, performed bioinformatics and statistical analyses. EB, NP, SD, CF, and SC prepared Figures and Tables. EB, NP, AMP, FPC, AR, EF and FM wrote the manuscript. All authors have reviewed and approved the manuscript.

DATA AVAILABILITY AND DATA SHARING STATEMENT

The data and samples referenced here are housed in the GEN-COVID Patient Registry and the GEN-COVID Biobank and are available for consultation. For further information, you may contact the corresponding author, Prof. Alessandra Renieri (e-mail: alessandra.renieri@unisi.it).

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

1. Benetti E, Giliberti A, Emiliozzi A, et al. Clinical and molecular characterization of COVID-19 hospitalized patients. *medRxiv doi: <https://doi.org/10.1101/2020.05.22.20108845>*.
2. Daga S, Fallerini C, Baldassarri M, et al. Employing a Systematic Approach to Biobanking and Analyzing Genetic and Clinical Data for Advancing COVID-19 Research. *medRxiv doi: <https://doi.org/10.1101/2020.07.24.20161307>*

3. Zhang X, Tan Y, Ling Y, et al. Viral and host factors related to the clinical outcome of COVID-19. *Nature* 2020; **583**(7816):437-440.
doi:10.1038/s41586-020-2355-0.
4. Ellinghaus D, Degenhardt F, Bujanda L, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure [published online ahead of print, 2020 Jun 17]. *N Engl J Med.* 2020;NEJMoa2020283.
doi:10.1056/NEJMoa2020283
5. Zhang Q, Bastard P, Liu Z, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science.* 2020 Sep 24:eabd4570.
doi: 10.1126/science.abd4570. Epub ahead of print. PMID: 32972995.
6. van der Made CI, Simons A, Schuurs-Hoeijmakers J, et al., Presence of Genetic Variants Among Young Men With Severe COVID-19. *JAMA.* 2020 Jul 24;324(7):1–11. doi: 10.1001/jama.2020.13719. Epub ahead of print. PMID: 32706371; PMCID: PMC7382021
7. Bastard P, Rosen LB, Zhang Q, et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science.* 2020 Oct 23; **370**(6515):eabd4585. doi: 10.1126/science.abd4585. Epub 2020 Sep 24. PMID: 32972996.
8. Callewaert L, Christiaens V, Haelens A, Verrijdt G, Verhoeven G, Claessens F. Implications of a polyglutamine tract in the function of the human androgen receptor. *Biochem Biophys Res Commun.* 2003; **306**(1): 46-52. doi:10.1016/s0006-291x(03)00902-1
9. Simanainen U, Brogley M, Gao Y R, et al. Length of the human androgen receptor glutamine tract determines androgen sensitivity in vivo. *Mol Cell Endocrinol.* 2011; **342**: 81–86. doi: 10.1016/j.mce.2011.05.011

10. Tirabassi G, Cignarelli A, Perrini S, et al. Influence of CAG Repeat Polymorphism on the Targets of Testosterone Action. *Int J Endocrinol.* 2015; 2015: 298107. doi:10.1155/2015/298107
11. Lindström S, Ma J, Altshuler D, et al. A large study of androgen receptor germline variants and their relation to sex hormone levels and prostate cancer risk. Results from the National Cancer Institute Breast and Prostate Cancer Cohort Consortium. *J Clin Endocrinol Metab.* 2010; **95**(9): E121-E127. doi:10.1210/jc.2009-1911
12. Wambier CG, Goren A. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is likely to be androgen mediated. *J Am Acad Dermatol.* 2020; **83**(1): 308-309. doi:10.1016/j.jaad.2020.04.032.
13. Pozzilli P, Lenzi A. Commentary: Testosterone, a key hormone in the context of COVID-19 pandemic. *Metabolism.* 2020; **108**: 154252. doi:10.1016/j.metabol.2020.154252.
14. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell.* 2020; **181**(2): 271-280.e8. doi:10.1016/j.cell.2020.02.052.
15. Rastrelli G, Di Stasi V, Inglese F, et al. Low testosterone levels predict clinical adverse outcomes in SARS-CoV-2 pneumonia patients [published online ahead of print, 2020 May 20]. *Andrology.* 2020; 10.1111/andr.12821. doi:10.1111/andr.12821.
16. Van Vliet M, Spruit MA, Verleden G, et al. Hypogonadism, quadriceps weakness, and exercise intolerance in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2005; **172**(9): 1105-1111. doi:10.1164/rccm.200501-114OC.

17. Mohan SS, Knuihan MW, Divitini ML, et al. Higher serum testosterone and dihydrotestosterone, but not oestradiol, are independently associated with favourable indices of lung function in community-dwelling men. *Clin Endocrinol (Oxf)*. 2015; **83**(2): 268-276. doi:10.1111/cen.12738.
18. Mohamad NV, Wong SK, Wan Hasan W N, et al. The relationship between circulating testosterone and inflammatory cytokines in men. *Aging Male*. 2019; **22**(2): 129-140. doi:10.1080/13685538.2018.1482487V.
19. Isidori AM, Pofi R, Hasenmajer V, Lenzi A, Pivonello R. Use of glucocorticoids in patients with adrenal insufficiency and COVID-19 infection. *Lancet Diabetes Endocrinol*. 2020; **8**(6):472-473. doi:10.1016/S2213-8587(20)30149-2
20. WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group, Sterne JAC, Murthy S, et al. Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis [published online ahead of print, 2020 Sep 2]. *JAMA*. 2020;10.1001/jama.2020.17023. doi:10.1001/jama.2020.17023
21. Allen RC, Zoghbi HY, Moseley AB, et al., Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet*. 1992; **51**(6):1229-39.
22. Millar AC, Lau ANC, Tomlinson G, Kraguljac A, Simel DL, Detsky AS, Lipscombe LL. Predicting low testosterone in aging men: a systematic review. *CMAJ*. 2016 Sep 20; **188**(13):E321-E330.

23. Gubbels Bupp MR, Jorgensen TN. Androgen-Induced Immunosuppression. *Front Immunol.* 2018; **9**: 794. Published 2018 Apr 17.
doi:10.3389/fimmu.2018.00794.
24. Hoebe K, Janssen E, Beutler B. The interface between innate and adaptive immunity. *Nat Immunol.* 2004; **5**(10): 971-974. doi:10.1038/ni1004-971.
25. Janeway CA Jr, Medzhitov R. Innate immune recognition. *Annu Rev Immunol.* 2002; **20**: 197-216.
doi:10.1146/annurev.immunol.20.083001.084359.
26. Lai JJ, Lai KP, Zeng W, Chuang KH, Altuwaijri S, Chang C. Androgen receptor influences on body defense system via modulation of innate and adaptive immune systems: lessons from conditional AR knockout mice. *Am J Pathol.* 2012; **181**(5): 1504-1512. doi:10.1016/j.ajpath.2012.07.008.
27. Medzhitov R, Janeway C Jr. Innate immunity. *N Engl J Med.* 2000; **343**(5): 338-344. doi:10.1056/NEJM200008033430506.
28. Pierotti S, Lolli F, Lauretta R, et al. Androgen modulation of pro-inflammatory and anti-inflammatory cytokines during preadipocyte differentiation. *Horm Mol Biol Clin Investig.* 2010; **4**(1): 483-488.
doi:10.1515/HMBCI.2010.059.
29. Ackerman CM, Lowe LP, Lee H, et al. Ethnic variation in allele distribution of the androgen receptor (AR) (CAG)_n repeat. *J Androl.* 2012; **33**(2): 210-215. doi:10.2164/jandrol.111.013391
30. Porcheddu R, Serra C, Kelvin D, Kelvin N, Rubino S. Similarity in Case Fatality Rates (CFR) of COVID-19/SARS-COV-2 in Italy and China. *J Infect Dev Ctries.* 2020; **14**(2): 125-128. Published 2020 Feb 29.
doi:10.3855/jidc.12600.

31. de Lusignan S, Dorward J, Correa A, et al. Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary care network: a cross-sectional study. *Lancet Infect Dis.* 2020; **20**(9): 1034-1042. doi:10.1016/S1473-3099(20)30371-6.
32. Caminiti G, Volterrani M, Iellamo F, et al. Effect of long-acting testosterone treatment on functional exercise capacity, skeletal muscle performance, insulin resistance, and baroreflex sensitivity in elderly patients with chronic heart failure a double-blind, placebo-controlled, randomized study. *J Am Coll Cardiol.* 2009; **54**(10): 919-927. doi:10.1016/j.jacc.2009.04.078.
33. Canale D, Caglieresi C, Moschini C, et al. Androgen receptor polymorphism (CAG repeats) and androgenicity [published correction appears in *Clin Endocrinol (Oxf)*. 2005 Oct; **63**(4):482]. *Clin Endocrinol (Oxf)*. 2005; **63**(3):356-361. doi:10.1111/j.1365-2265.2005.02354.x
34. Molnar, R. Interpretable machine learning. A Guide for Making Black Box Models Explainable. 2020.
35. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 1996; **58**(21):267-288.
36. Fanelli F, Belluomo I, Di Lallo VD, et al. Serum steroid profiling by isotopic dilution-liquid chromatography-mass spectrometry: comparison with current immunoassays and reference intervals in healthy adults. *Steroids.* 2011 Feb; **76**(3):244-53. doi: 10.1016/j.steroids.2010.11.005.

FIGURE LEGENDS

Figure 1 LASSO logistic regression.

This histogram of the LASSO logistic regression weights represents the importance of each feature for the classification task (**Figure 1**) (**Panel A**). The positive weights reflect a susceptible behaviour of the features to the target COVID-19 disease, whereas the negative weights a protective action. The calculated odd ratio of AR short repeats (≤ 22) is 0,79 i.e. protective. Therefore, the odd ratio of long repeats (≥ 23) is $1/0,79 = 1,27$ i.e. severity.

Panel B: Confusion matrix for the aggregation of the logistic regression predictions in the 10 folds of the cross-validation. **Panel C:** Cross-validation ROC-AUC score for the grid of LASSO regularization parameters; the error bar is given by the standard deviation of the score within the 10 folds; the red point corresponds to the parameter chosen for the fitting procedure. **Panel D:** Boxplot of accuracy, precision, sensitivity, specificity, and ROC-AUC score for the 10-fold of the cross-validation. The box extends from the Q1 to Q3 quartile, with a line at the median (Q2) and a triangle for the average. **Panel E:** ROC curve for the 10 folds of the cross-validation.

Figure 2 Relationship between Total Testosterone and polyQ repeats and clinical outcome

Panel A: Increased Total Testosterone levels in patients with longer polyQ repeats.

Box-plot showing values of Total Testosterone (TT), expressed in nmol/L, in patients with shorter (≤ 22) and longer (≥ 23) polyQ repeats in AR gene. A subset of 183 males is studied: 122 males with polyQ repeats ≤ 22 and 61 with polyQ repeats ≥ 23 . The TT median value, represented by the black horizontal line, is higher in patients with ≥ 23 polyQ repeats. (*p-value = 0.0422; Mann-Whitney U test).

Panel B: Reduced Testosterone/PolyQ ratio in patients with the most severe COVID-19 clinical presentation.

Box-plot showing the ratio Total Testosterone (TT), over the number of polyQ repeats, in cases and controls. A subset of 183 males (109 cases, 74 controls) was studied. The median of TT Ratio (nmol/L/polyQ number), represented by the black horizontal line, is lower in patients that underwent intubation/CPAP/BiPAP ventilation (cases). (***)p-value = 0,00038; Mann-Whitney U test).

ADDITIONAL INFORMATION

GEN-COVID Multicenter Study (<https://sites.google.com/dbm.unisi.it/gen-covid>)

Francesca Montagnani^{2,22}, Laura Di Sarno^{1,2}, Andrea Tommasi^{1,2,5}, Maria Palmieri^{1,2}, Arianna Emiliozzi^{2,22}, Massimiliano Fabbiani²², Barbara Rossetti²², Giacomo Zanelli^{2,22}, Fausta Sestini²⁰, Laura Bergantini²³, Miriana D'Alessandro²³, Paolo Cameli²³, David Bennet²³, Federico Anedda²⁴, Simona Marcantonio²⁴, Sabino Scolletta²⁴, Federico Franchi²⁴, Maria Antonietta Mazzei²⁵, Susanna Guerrini²⁵, Edoardo Conticini²⁶, Luca Cantarini²⁶, Bruno Frediani²⁶, Danilo Tacconi²⁷, Chiara Spertilli²⁷, Marco Feri²⁸, Alice Donati²⁸, Raffaele Scala²⁹, Luca Guidelli²⁹, Genni Spargi³⁰, Marta Corridi³⁰, Cesira Nencioni³¹, Leonardo Croci³¹, Gian Piero Caldarelli³², Maurizio Spagnesi³³, Paolo Piacentini³³, Maria Bandini³³, Elena Desantis³³, Silvia Cappelli³³, Anna Canaccini³⁴, Agnese Verzuri³⁴, Valentina Anemoli³⁴, Agostino Ognibene³⁵, Massimo Vaghi³⁶, Antonella D'Arminio Monforte³⁷, Esther Merlini³⁷, Federica Gaia Miraglia³⁷, Mario U. Mondelli^{38,39}, Stefania Mantovani³⁸, Serena Ludovisi^{38,39}, Massimo Girardis⁴⁰, Sophie Venturelli⁴⁰, Marco Sita⁴⁰, Andrea Cossarizza⁴¹, Andrea Antinori⁴², Alessandra Vergori⁴², Stefano Rusconi^{43,44}, Matteo Siano⁴⁴, Arianna Gabrieli⁴⁴, Agostino Riva^{43,44}, Daniela Francisci^{45,46}, Elisabetta Schiaroli⁴⁵, Pier Giorgio Scotton⁴⁷, Francesca Andretta⁴⁷, Sandro Panese⁴⁸, Stefano Baratti⁴⁸, Renzo Scaggiante⁴⁹, Francesca Gatti⁴⁹, Saverio Giuseppe Parisi⁵⁰, Francesco Castelli⁵¹, Eugenia Quiros-Roldan⁵¹, Melania Degli Antoni⁵¹, Isabella Zanella⁵², Matteo Della Monica⁵³, Carmelo Piscopo⁵³, Mario Capasso^{54,55,56}, Roberta Russo^{54,55}, Immacolata Andolfo^{54,55}, Achille Iolascon^{54,55}, Giuseppe Fiorentino⁵⁷, Massimo Carella⁵⁸, Marco Castori⁵⁸, Giuseppe Merla⁵⁸, Filippo Aucella⁵⁹, Pamela Raggi⁶⁰, Carmen Marciano⁶⁰, Rita Perna⁶⁰, Matteo Bassetti^{61,62}, Antonio Di Biagio⁶², Maurizio Sanguinetti^{63,64}, Luca Masucci^{63,64}, Serafina Valente⁶⁵, Maria Antonietta Mencarelli⁵, Caterina Lo Rizzo⁵, Elena Bargagli²³, Marco Mandalà⁶⁶, Alessia Giorli⁶⁶, Lorenzo Salerno⁶⁶, Patrizia Zucchi⁶⁷, Pierpaolo Parravicini⁶⁷, Elisabetta Menatti⁶⁸, Tullio Trotta⁶⁹, Ferdinando Giannattasio⁶⁹, Gabriella Coiro⁶⁹, Fabio Lena⁷⁰, Domenico A. Coviello⁷¹, Cristina Mussini⁷², Giancarlo Bosio⁷³, Enrico Martinelli⁷³, Sandro Mancarella⁷⁴, Luisa Tavecchia⁷⁴, Lia Crotti^{75,76,77, 78}

- 22) Dept of Specialized and Internal Medicine, Tropical and Infectious Diseases Unit, Azienda Ospedaliera Universitaria Senese, Siena, Italy
- 23) Unit of Respiratory Diseases and Lung Transplantation, Department of Internal and Specialist Medicine, University of Siena, Italy
- 24) Dept of Emergency and Urgency, Medicine, Surgery and Neurosciences, Unit of Intensive Care Medicine, Siena University Hospital, Italy
- 25) Department of Medical, Surgical and Neuro Sciences and Radiological Sciences, Unit of Diagnostic Imaging,

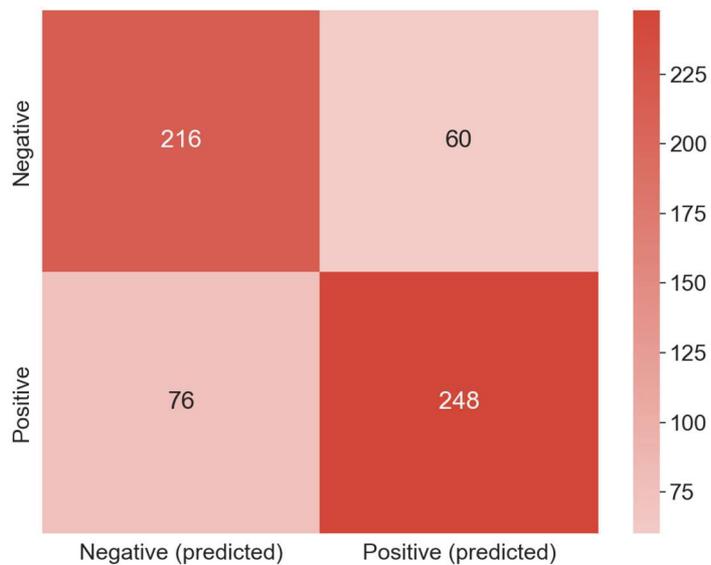
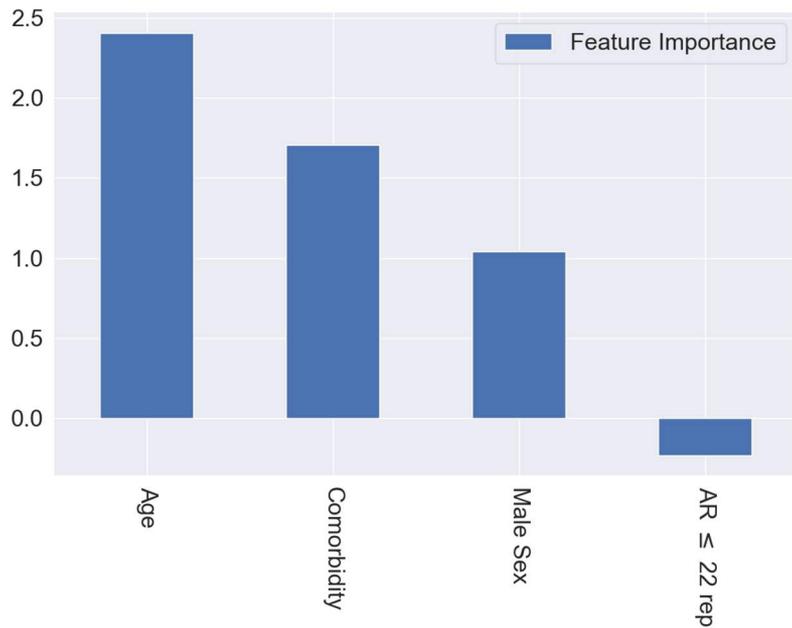
- University of Siena, Italy
- 26) Rheumatology Unit, Department of Medicine, Surgery and Neurosciences, University of Siena, Policlinico Le Scotte, Italy
 - 27) Department of Specialized and Internal Medicine, Infectious Diseases Unit, San Donato Hospital Arezzo, Italy
 - 28) Dept of Emergency, Anesthesia Unit, San Donato Hospital, Arezzo, Italy
 - 29) Department of Specialized and Internal Medicine, Pneumology Unit and UTIP, San Donato Hospital, Arezzo, Italy
 - 30) Department of Emergency, Anesthesia Unit, Misericordia Hospital, Grosseto, Italy
 - 31) Department of Specialized and Internal Medicine, Infectious Diseases Unit, Misericordia Hospital, Grosseto, Italy
 - 32) Clinical Chemical Analysis Laboratory, Misericordia Hospital, Grosseto, Italy
 - 33) Department of Preventive Medicine, Azienda USL Toscana Sud Est, Italy
 - 34) Territorial Scientific Technician Department, Azienda USL Toscana Sud Est, Italy
 - 35) Clinical Chemical Analysis Laboratory, San Donato Hospital, Arezzo, Italy
 - 36) Chirurgia Vascolare, Ospedale Maggiore di Crema, Italy
 - 37) Department of Health Sciences, Clinic of Infectious Diseases, ASST Santi Paolo e Carlo, University of Milan, Italy
 - 38) Division of Infectious Diseases and Immunology, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy
 - 39) Department of Internal Medicine and Therapeutics, University of Pavia, Italy
 - 40) Department of Anesthesia and Intensive Care, University of Modena and Reggio Emilia, Modena, Italy
 - 41) Department of Medical and Surgical Sciences for Children and Adults, University of Modena and Reggio Emilia, Modena, Italy
 - 42) HIV/AIDS Department, National Institute for Infectious Diseases, IRCCS, Lazzaro Spallanzani, Rome, Italy
 - 43) III Infectious Diseases Unit, ASST-FBF-Sacco, Milan, Italy
 - 44) Department of Biomedical and Clinical Sciences Luigi Sacco, University of Milan, Milan, Italy
 - 45) Infectious Diseases Clinic, Department of Medicine 2, Azienda Ospedaliera di Perugia and University of Perugia, Santa Maria Hospital, Perugia, Italy
 - 46) Infectious Diseases Clinic, "Santa Maria" Hospital, University of Perugia, Perugia, Italy
 - 47) Department of Infectious Diseases, Treviso Hospital, Local Health Unit 2 Marca Trevigiana, Treviso, Italy
 - 48) Clinical Infectious Diseases, Mestre Hospital, Venezia, Italy.
 - 49) Infectious Diseases Clinic, ULSS1, Belluno, Italy
 - 50) Department of Molecular Medicine, University of Padova, Italy
 - 51) Department of Infectious and Tropical Diseases, University of Brescia and ASST Spedali Civili Hospital, Brescia, Italy
 - 52) Department of Molecular and Translational Medicine, University of Brescia, Italy; Clinical Chemistry Laboratory, Cytogenetics and Molecular Genetics Section, Diagnostic Department, ASST Spedali Civili di Brescia, Italy
 - 53) Medical Genetics and Laboratory of Medical Genetics Unit, A.O.R.N. "Antonio Cardarelli", Naples, Italy
 - 54) Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II, Naples, Italy
 - 55) CEINGE Biotecnologie Avanzate, Naples, Italy
 - 56) IRCCS SDN, Naples, Italy
 - 57) Unit of Respiratory Physiopathology, AORN dei Colli, Monaldi Hospital, Naples, Italy
 - 58) Division of Medical Genetics, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy
 - 59) Department of Medical Sciences, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy
 - 60) Clinical Trial Office, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy
 - 61) Department of Health Sciences, University of Genova, Genova, Italy
 - 62) Infectious Diseases Clinic, Policlinico San Martino Hospital, IRCCS for Cancer Research Genova, Italy
 - 63) Microbiology, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Catholic University of Medicine, Rome, Italy
 - 64) Department of Laboratory Sciences and Infectious Diseases, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy
 - 65) Department of Cardiovascular Diseases, University of Siena, Siena, Italy
 - 66) Otolaryngology Unit, University of Siena, Italy
 - 67) Department of Internal Medicine, ASST Valtellina e Alto Lario, Sondrio, Italy
 - 68) Study Coordinator Oncologia Medica e Ufficio Flussi Sondrio, Italy
 - 69) First Aid Department, Luigi Curto Hospital, Polla, Salerno, Italy
 - 70) Local Health Unit-Pharmaceutical Department of Grosseto, Toscana Sud Est Local Health Unit, Grosseto, Italy

- 71) U.O.C. Laboratorio di Genetica Umana, IRCCS Istituto G. Gaslini, Genova, Italy.
- 72) Infectious Diseases Clinics, University of Modena and Reggio Emilia, Modena, Italy.
- 73) Department of Respiratory Diseases, Azienda Ospedaliera di Cremona, Cremona, Italy
- 74) U.O.C. Medicina, ASST Nord Milano, Ospedale Bassini, Cinisello Balsamo (MI), Italy
- 75) Istituto Auxologico Italiano, IRCCS, Department of Cardiovascular, Neural and Metabolic Sciences, San Luca Hospital, Milan, Italy.
- 76) Department of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy
- 77) Istituto Auxologico Italiano, IRCCS, Center for Cardiac Arrhythmias of Genetic Origin, Milan, Italy.
- 78) Istituto Auxologico Italiano, IRCCS, Laboratory of Cardiovascular Genetics, Milan, Italy

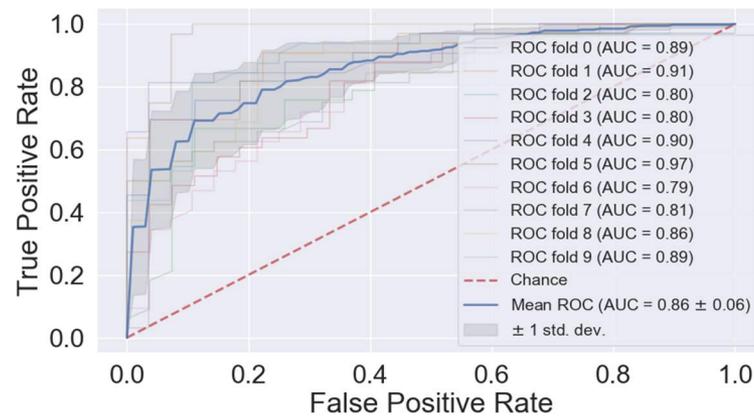
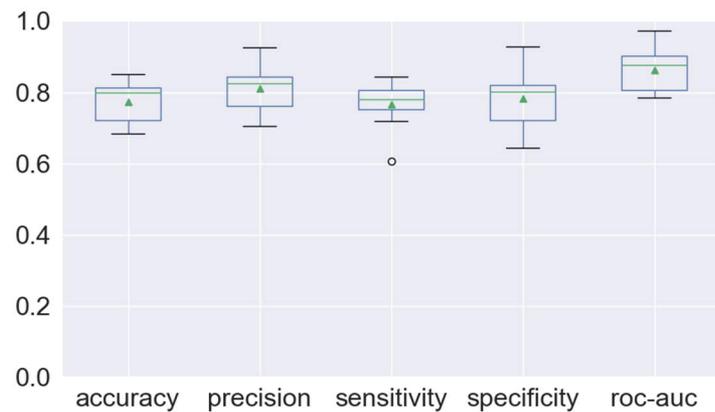
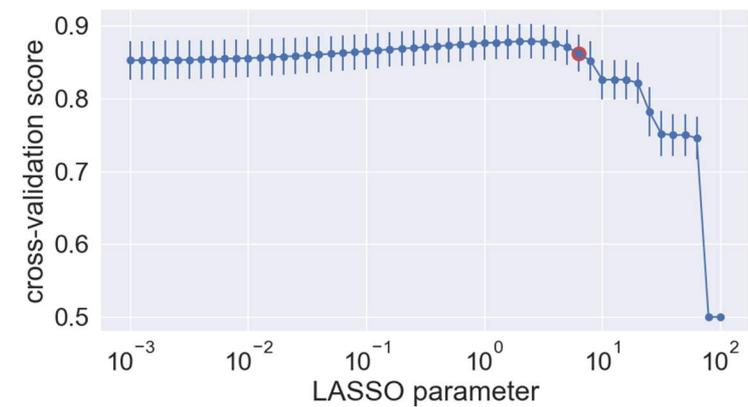
Spanish COVID Human Genetic Effort

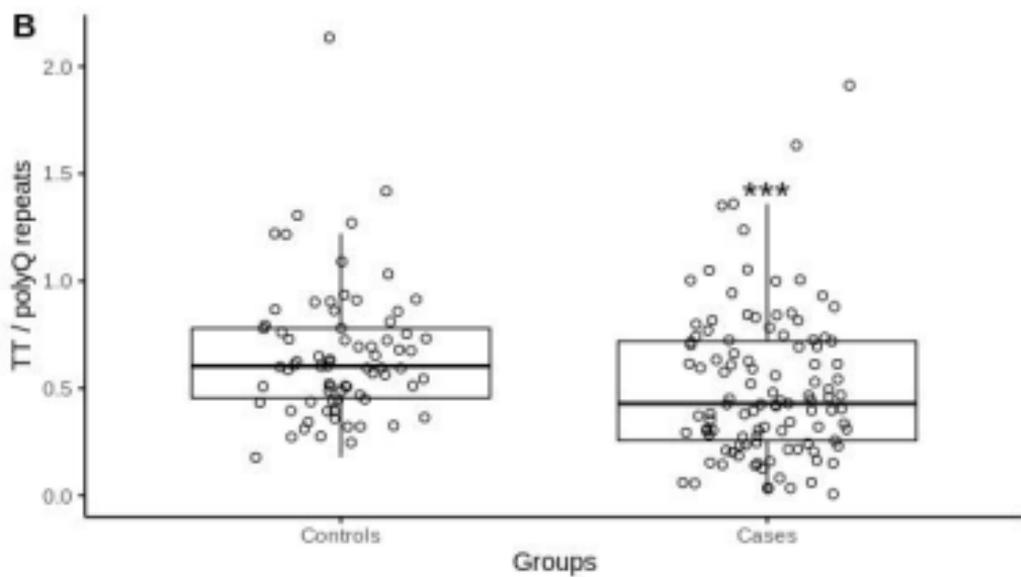
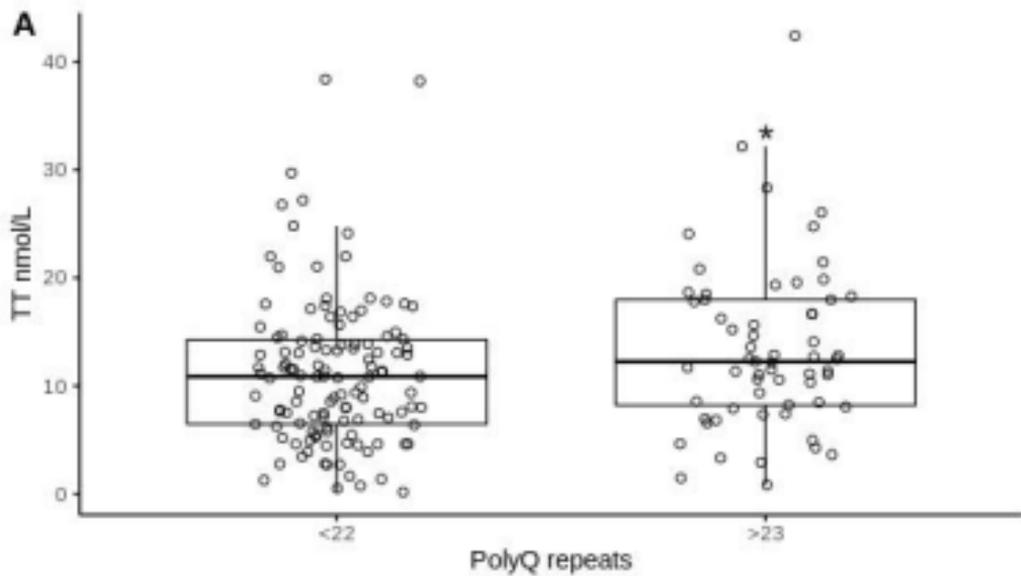
Sergio Aguilera-Albesa⁷⁹, Sergiu Albu⁸⁰, Carlos Casasnovas^{81,13}, Valentina Vélez-Santamaria^{81,13}, Juan Pablo Horcajada⁸², Judit Villar⁸², Agustí Rodríguez-Palmero^{83,13,14}, Montserrat Ruiz^{13,14}, Luis M Seijo⁸⁴, Jesús Troya⁸⁵, Juan Valencia-Ramos⁸⁶, Marta Gut⁸⁷

- 79) Navarra Health Service Hospital, Pamplona, Spain
- 80) Institut Guttmann Foundation, Badalona, Barcelona, Spain
- 81) Bellvitge University Hospital, L'Hospitalet de Llobregat, Barcelona, Spain
- 82) Hospital del Mar, Parc de Salut Mar, Barcelona, Spain
- 83) University Hospital Germans Trias i Pujol, Badalona, Barcelona, Spain
- 84) Clínica Universitaria de Navarra, Madrid, Spain
- 85) Infanta Leonor University Hospital, Madrid, Spain
- 86) University Hospital of Burgos, Burgos, Spain
- 87) CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Carrer Baldiri i Reixac 4, 08028, Barcelona, Spain



B





| | | Intubation | CPAP/BiPAP Ventilation | Oxygen Therapy | No respiratory support | Oligo-asymptomatics without hospitalization |
|------------------|---------------------|-------------------|-------------------------------|-----------------------|-------------------------------|--|
| GEN-COVID | Number of Subjects | 108 | 230 | 352 | 188 | 300 |
| | Male/Female | 80/28 | 157/73 | 208/144 | 104/84 | 116/184 |
| | Age males (years) | 61,52±11,43 | 62,75±13,48 | 63,41±14,53 | 55,99±15,44 | 47,40±13,23 |
| | Age females (years) | 63,71±13,96 | 66,23±15,25 | 68,40±14,74 | 52,88±16,39 | 48,61±11,06 |
| | | Cases | | | | Controls |
| NCC study | Number of Subjects | 338 | | | | 300 |
| | Male/Female | 237/101 | | | | 116/184 |
| | Age males (years) | 62,34±12,84 | | | | 47,40±13,23 |
| | Age females (years) | 65,53±14,94 | | | | 48,61±11,06 |

| Males <60 | | | |
|-------------------------------|---------------|---------------|----------------------------|
| | <22 | >23 | Marginal Row Totals |
| Cases | 52 (59,15%) | 36 (28,85%) | 88(48,08%) |
| Controls | 71 (63,85%) | 24 (31,15%) | 95(51,91%) |
| Marginal Column Totals | 123 (67,21%) | 60 (32,78%) | 183 (Grand Total) |

* p-value (cases vs controls) = 0,024

Table 3. Validation in Spanish cohort

| Spanish validation (χ^2) | | | |
|---|-----------------------------|-----------------------------|-----------------------------------|
| Males global | | | |
| | ≤ 22 | ≥ 23 | <i>Marginal Row Totals</i> |
| Cases | 51 (32,27%) | 66 (41,77%) | 117 (74,05%) |
| Controls | 27 (17,08%)* | 14 (8,86%) | 41 (25,94%) |
| <i>Marginal Column Totals</i> | 78 (49,36%) | 80 (50,63%) | 158 (Grand Total) |

* p-value (cases vs controls) = 0,014139 (Significant at $p < 0.05$)

Table 4. Correlation between polyA repeats in AR gene and laboratory values.

CRP M \geq 60y cases

| Triplets | Mean | Count | p-value |
|-----------|-------|-------|---------------------------------|
| ≤ 22 | 48,21 | 78 | 0,01804 (Significant at p<0,05) |
| ≥ 23 | 55,92 | 38 | |

CRP M<60y cases

| Triplets | Mean | Count | p-value |
|-----------|-------|-------|---------|
| ≤ 22 | 54,5 | 43 | 0,2 |
| ≥ 23 | 26,41 | 29 | |

Fibrinogen M \geq 60y cases

| Triplets | Mean | Count | p-value |
|-----------|--------|-------|---------|
| ≤ 22 | 401,33 | 57 | 0,09277 |
| ≥ 23 | 320,34 | 27 | |

Fibrinogen M<60y cases

| Triplets | Mean | Count | p-value |
|-----------|--------|-------|---------|
| ≤ 22 | 316,93 | 22 | 0,53 |
| ≥ 23 | 356,91 | 19 | |

IL6 times the upper limit of normal M \geq 60y cases

| Triplets | Mean | Count | p-value |
|-----------|-------|-------|---------|
| ≤ 22 | 54,56 | 40 | 0,249 |
| ≥ 23 | 75,78 | 16 | |

IL6 times the upper limit of normal M<60y cases

| Triplets | Mean | Count | p-value |
|-----------|-------|-------|---------|
| ≤ 22 | 40,43 | 17 | 0,81 |
| ≥ 23 | 31,8 | 14 | |

CD4 Lymphocytes M \geq 60y cases

| Triplets | Mean | Count | p-value |
|-----------|--------|-------|---------|
| ≤ 22 | 264,06 | 32 | 0,22 |
| ≥ 23 | 357,52 | 21 | |

CD4 Lymphocytes M<60y cases

| Triplets | Mean | Count | p-value |
|-----------|--------|-------|---------|
| ≤ 22 | 503,68 | 16 | 0,45 |
| ≥ 23 | 396,13 | 15 | |

NK Cells M \geq 60y cases

| Triplets | Mean | Count | p-value |
|-----------|--------|-------|---------|
| ≤ 22 | 70,71 | 28 | 0,179 |
| ≥ 23 | 102,25 | 16 | |

NK Cells M<60y cases

| Triplets | Mean | Count | p-value |
|-----------|--------|-------|---------|
| ≤ 22 | 147,3 | 13 | 0,098 |
| ≥ 23 | 107,14 | 14 | |