The role of thrombophilia related polymorphisms in Coronavirus pandemic mortality

Marcos Henrique Damasceno Cantanhede (ma_cantanhede@hotmail.com)
Federal University of Para: Universidade Federal do Para
https://orcid.org/0000-0003-2263-8403

Kevin Matheus de Lima Sarges
Federal University of Para: Universidade Federal do Para

Mauro de Meira Leite
Federal University of Para: Universidade Federal do Para

Fábio Miyajima
Oswaldo Cruz Foundation: branch Ceará

Eduardo José Melo dos Santos
Universidade Federal do Pará: Universidade Federal do Para

Research Article

Keywords: COVID-19, Polymorphisms, Thrombophilia, Death Rates

DOI: https://doi.org/10.21203/rs.3.rs-279902/v1

License: ☇️ This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

The current study represents a comprehensive analysis in several databases and extensive bibliographic review, aiming to evaluate the correlation between the frequency of thrombophilia related polymorphisms with mortality rates by COVID-19 worldwide. We found 8 polymorphisms at 8 genes statistically correlated with Daily Death Rates of COVID-19, being an important clue to assignment of genes involved in poor prognosis of COVID-19, specifically the ones related to thromboembolic events.

Main Text

COVID-19 was initially described as an acute respiratory distress syndrome (ARDS) (1). As further understanding about the disease has been gained, it became increasingly apparent the existence of other relevant hallmark features, particularly cytokine storm and systemic thromboembolism (1). Probable mechanisms linking COVID-19 with thromboembolic events have been highlighted by recent reports, amongst them: i) imbalance of Angiotensin-Converting Enzyme-2 (ACE2) caused by direct viral action on their receptors; ii) the interplay between inflammatory cytokines, like IL-6 and IL-17A (1), platelet hyper-reactivity and thrombosis; iii) overexpression of platelets due to the presence of extramedullary megakaryocytes that produces platelets during COVID-19 infection and iv) increase of Von Willebrand Factor and Factor VIII expression (1,2). Despite the importance of thromboembolism in systemic manifestation of the disease, only a fraction of patients develops thromboembolic disease, reaching 20-30% in some study series and being a poor prognosis factor (2). Additional studies also point to D-dimer as an important marker in disease severity (3).

Intrinsic individual factors may contribute to thromboembolic disease in COVID-19: i) comorbidities like hypertension, diabetes and obesity; ii) advanced age that are related to these comorbidities; iii) inflammatory status; iv) predisposing genetic factors (1,4).

There are a few studies that associates COVID-19 with genetics factors, mostly related to the immune response, like the HLA Class I genes (5). ABO blood group and single nucleotide polymorphisms (SNP) in 3p21.31 and 9q34.2 (6); 19p13.3, 12q24.13 and 21q22.1 (4) regions were also recently reported as associated in genome wide association studies (GWAS). Ongoing GWAS initiatives may reveal further key host genetics players underpinning susceptibility-resistance to the disease. Encoded proteins related to the coagulation cascade are now being described as critically influenced by pathological alterations brought by COVID-19, modulating key elements in this pathways, such as von Willebrand factor (vWF) and the antithrombin III binding agent (2). Interestingly, a study from Manne et al. reported the increase of P-selectin, a major platelet activator, occurs mainly during acute phase of COVID-19, which may unveil key gene modulations accounting for the thrombophilia (2).

Moreover, an whole exome sequencing approach identified polymorphisms at immune related genes in regions 4q35.1, 11q13.2, 19p13.3, 12q14.2, 19q13.33, 11p15.5 and 21q22.11 (7), where three of them at chromosomes 12, 19 and 21 are coincident with previous GWAS studies.

Although blood screening tests routinely used for monitoring hematological parameters, such as D-dimer, coupled with the assessment of thromboembolic comorbidities (1) are considered relevant risk factors for thrombosis-related events, little is still known about the genetic basis underlying the clotting cascade in the presence of COVID-19. In this context, genetic polymorphisms that predispose to thromboembolism have the potential to generate clinically relevant knowledge on COVID-19 pathogenesis that can be used for prognosis and stratification of therapy.
Hence, we performed a comprehensive list of functional polymorphisms in 24 key genes or clusters related to thrombophilia that were obtained from surveys in OMIM (8) and Orphanet (9) databases, listed in notes of Table 1. Since very low frequency SNPs are unlikely to contribute to general epidemiological findings on COVID-19, SNPs that have frequency below 1% were not considered for the metaanalysis. Additionally, a number of SNPs have no available frequencies in populations. The frequencies in populations worldwide of the remaining 18 SNPs mapping 15 genes were retrieved from Ensembl database (10) along with a comprehensive survey using the SNP IDs as keyword (raw data and bibliographic sources are given in Additional Files 1 and 2) and are presented in Table 1, with their respective gene location, frequency ranges and effects on thrombophilia.

| Genes | Chromosome Localization | Min Freq | Max Freq | SNP     | Effects               |
|-------|-------------------------|----------|----------|---------|-----------------------|
| MPL   | 1p34.2                  | 0        | 0.071    | rs17292650 | Susceptibility to thrombus formation |
| ADAMTS13 | 9q34.2               | 0.004    | 0.529    | rs2301612 | Susceptibility to thrombus formation |
| F13B  | 1q31.3                  | 0.015    | 0.424    | rs6003   | Susceptibility to thrombus formation |
| FGA   | 4q31.3                  | 0.188    | 0.567    | rs6050   | D-Dimers Levels       |
| F9    | Xq27.1                  | 0        | 0.32     | rs6048   | Susceptibility to thrombus formation |
| THBD  | 20p11.21                | 0        | 0.034    | rs41348347 | Thrombophilic         |
| HABP2 | 10q25.3                 | 0        | 0.055    | rs7080536 | Marburg I Disease (Thrombophilic) |
| SERPINE1 | 7q22.1               | 0.12     | 0.859    | rs1799762 | Thrombophilic         |
| SERPINE1 | 7q22.1               | 0        | 0.121    | rs6092   | Protection to thrombus formation |
| MTHFR | 1p36.22                 | 0.062    | 0.66     | rs1801133 | Susceptibility to thrombus formation |
| MTHFR | 1p36.22                 | 0.082    | 0.712    | rs1801131 | Susceptibility to thrombus formation |
| F13A  | 6p25.1                  | 0        | 0.312    | rs5985   | Protection to thrombus formation |
| LOC105378861 | 1p21.3    | 0.0525   | 0.332    | rs12029080 | D-Dimers Levels       |
| Intergenic | Chr1: 169508336     | 0        | 0.202    | rs6687813 | D-Dimers Levels       |
| FGA   | 4q31.3                  | 0.182    | 0.4660   | rs13109457 | D-Dimers Levels       |
| FGG   | 4q32.1                  | 0.01     | 0.476    | rs2066865 | D-Dimers Levels       |
| F5    | 1q24.2                  | 0        | 0.1361   | rs6025   | Factor V of Leiden    |
| F2    | 11p11.2                 | 0        | 0.29     | rs1799963 | Susceptibility to thrombus formation |

es selected from Orphanet, 2 Available at Ensembl, 3 Available at OMIM, 4 Thrombopoietin receptor 4 Metallopeptidase bospodrin type 1 motif 13 6 Coagulation factor XIII B chain 7 Fibrinogen alpha chain 8 Coagulation factor IX 9 Thrombomodulin binding protein 2 11 Serpin family E member 1 12 Methylenetetrahydrofolate reductase 13 Fibrinogen alpha chain 15 Fibrinogen gamma chain 16 Coagulation factor V 17 Coagulation factor II, Thrombin. The other six are not included in this list are: Kininogen 1 (KNG1), Protein C, inactivator of coagulation factors Va and VIIIa (PROC), Pro
1), Serpin family C member 1 (SERPINC1), Vitamin K epoxide reductase complex subunit 1 (VKORC1), Janus kinase 2 (JAK2) and cytokine C2 (HCFC2). KNG1, PROC and JAK2 does not have any frequency available for their SNPs, the PROS1, SERPINC1 and HCF2 it, two SNPs with frequency available, but all of them have frequencies lower than 1%. The genes SERPINC1, PROC and PROS1 are orphic, having at least 8 mutations described.

We also sought to apply an indirect approach to detect the relevance of these SNPs in COVID-19 prognosis was applied, by correlating worldwide frequencies of these SNP with their mortality rates. Estimates of number of individuals that are infected, or died, by SARS-CoV-2 in over 200 countries were obtained from WHO Dashboard (11) in November, 02, 2020, as well as their respective inhabitant numbers.

These data were used to estimate for each country: (i) the Case Fatality Rate (CFR), defined by the number of deaths by COVID-19 divided by the number of confirmed cases and (ii) the Daily Death Rate (DDR), represented as the average number of deaths per day (since the first confirmed case) per ten million inhabitants. CFR and DDR estimates for all countries are presented along with genetic data in Supplementary Data.

For thirteen highly polymorphic SNPs Spearman Linear Correlation of their frequencies with CFR and DDR estimates was carried out. Five SNPs were polymorphic at only a fraction of the populations. Thus, CFR and DDR estimates were compared between two groups of populations, one composed by populations where the SNP was polymorphic and another where the SNP was monomorphic. These comparisons were made using Mann-Whitney test. Correction for multiple tests were applied accordingly. The Spearman Linear Correlation and Mann-Whitney were performed on the program BioEstat version 5.3 (12).

The results of the statistical analysis are presented in Table 2. We found significant correlation between frequencies and DDR in seven SNPs, remaining six significant after correction for multiple tests. No frequencies were correlated with CFR. Moreover, results from Mann-Whitney test suggested that two SNPs are associated to DDR, even after correction for multiple tests. No association was detected between SNP polymorphism and CFR.

92: Spearman Linear Correlation p-value and rs value for each mutation comparing DDR and CFR with the frequency of the mutation

Mann-Whitney Test p-value for each mutation comparing DDR and CFR with the frequency of the mutation
Our results suggest association of eight thrombophilia related SNPs with death rates attributable to COVID-19. Only two SNPs enlisted as probable mechanistic candidates were closely mapped to previously described positions associated with poor prognosis, namely the rs2301612 within ADAMST13 and rs5985 within F13A genes. We observe that the polymorphism rs2301612 is functionally related to thrombophilia and localized within the Chr:9q34.2 region, the same locus implicated by a recent COVID-19 GWAS study (6) and, which also, overlap with the ABO blood group locus at 9q34.2. Moreover, the polymorphism rs5985 is localized within the F13A gene were found to be significantly associated with thrombophilia. This SNP is approximately one million pb distant from RIOK1 gene, a candidate outside MHC region described previously by GWAS (4).

The remaining seven SNP linked with DDR mapped the genes FGA, FGG, F2, F5, F9, HABP2 and MTHFR, which were not clusters described previously by other authors, though they are considered well known polymorphisms associated with thrombophilia. FGA and FGG polymorphisms have been associated with D-Dimer levels that has been shown to be an important marker in COVID-19 (3). Both F2 and F9 SNPs were associated to susceptibility to thrombus formation as well as the F5 SNP known as Factor 5 Leiden, responsible by hypercoagulability and thrombosis (13). In the same context the SNP within MTHFR gene is also been linked with thrombosis susceptibility (14).
Considering three major ethnicities, Africans, Europeans and East Asians there are clearly remarkable differences in DDR, in which mortality rates were significantly higher within the European populations compared to others. Interestingly, the allele frequencies of the implicated markers we evaluated were found to be consistently higher amongst Europeans (Table 3), further strengthening our assessment strategy. In this context classical mutations related to thrombophilia, Factor V Leiden and MTHFR, that are more frequent among Europeans, are highlighted as potential biomarkers for thromboembolic risk in COVID-19.

SNPs average frequencies (%), DDR and Mann-Whitney test p-value for the average frequencies of SNP in three major ethnicities

| SNP (Frequencies) | Europeans | Africans | South-East Asians |
|-------------------|-----------|----------|-------------------|
| rs6050            | 30.75     | 44.1     | 37.85             |
| rs6048            | 27.87     | 13.68**  | 4.6**             |
| rs7080536         | 2.86      | 0**      | 0**               |
| rs1801133         | 38.94     | 7.5***   | 26.6**            |
| rs5985            | 24        | 16**     | 0**               |
| rs2066865*        | 22.4      | 34.48**  | 37.55**           |
| rs6025            | 5         | 0***     | 0.6***            |
| rs1799963         | 1.1       | 0**      | 0**               |
| DDR               | 9.83      | 1.07     | 1.63              |

Allele frequencies are presented in supplementary data. * SNP that are protective (negatively correlated with DDR); ** p-value ≤ 0.05 in means but not significant after correction for multiple tests; *** p-value ≤ 0.01 if compared with Europeans and remained significant for multiple tests.

For the majority of healthcare services worldwide, the molecular or serological testing for SARS-CoV-2 infection have been preferentially applied to severe/critical cases and to clear suspect death by COVID-19. Hence, differences in testing coverage would impact less in the number of deaths than in the number of cases. This global scenario favors epidemiological statistics like DDR that considers country's population as denominator and take in account the time. Hence, DDR, while an average daily incidence, seems to be more suitable to COVID-19 that is still an ongoing pandemic.

The present approach presents preliminary evidence that a significant proportion of deaths by COVID-19 must likely be the result of thrombophilia-related events that can be at least in part explained by difference in the genetic distribution of underpinning polymorphisms. Correlations tend to be stronger with DDR rather than CFR. Since CFR is obtained through the product of the number of deaths over the number of confirmed cases (as the denominator), it is automatically inferred this can be heavily biased by low testing coverage and the high frequency of both asymptomatic and non-reported mild cases, which altogether are reliant on socio-economic and political management factors. It is widely acknowledged that in the majority of countries, SARS-CoV2 testing have been preferentially offered to particular segments of the population, noticeably to severe/critical cases and suspect COVID-19 deaths. Hence merely meta-analytical, the strength of the correlation tests makes these results interesting to be considered for future evaluation and need to be corroborated in more structured case control or cohort studies. Although our direct meta-analytical approach endorsed the contributing role of the thrombophilia-associated genetic markers towards COVID-19 outcomes, further work is warranted to generate further experimental evidence. Additionally, the results suggest careful planning of sampling strategies in order to avoid stratification (15), because these polymorphisms have a wide range of variability associated with ethnicity. Finally, cumulative effects of multiple polymorphisms should be considered by evaluating the role of these SNPs in COVID-19.
Declarations

Ethical Approval and consent to participate

Not applicable

Consent for publication

Not applicable

Author’s contributions

Marcos Henrique Damasceno Cantanhede participated of all phases of data gathering, analysis and manuscript preparation. Kevin Matheus de Lima Sarges and Mauro de Meira Leite data gathering and participation on. Fabio Miyajima contribution to data analysis and manuscript writing. Eduardo José Melo dos Santos idealized the project and participated in all phases of the study and manuscript preparation.

Availability of data and materials

The datasets generated and analyzed in the current study are available in OMIM repository, https://www.omim.org/, Ensembl repository, https://www.ensembl.org/index.html, Orphanet, https://www.orpha.net/consor/cgi-bin/index.php, WHO dashboard, https://covid19.who.int/ and in published articles and its supplementary information files.

Competing Interests

The authors declare that they have no competing interests

Funding

Funding MSc fellowship to Marcos Henrique Damasceno Cantanhede. Financial support by CNPq grant MCTI/CNPq/MS/SCTIE/Decit Processo: 401235/2020-3

Acknowledgement

Not applicable

References

1. Cevik M, Kuppalli K, Kindrachuk J, Peiris M. Virology, transmission, and pathogenesis of SARS-CoV-2. BMJ. 2020;371:m3862.

2. Manne BK, Denorme F, Middleton EA, Portier I, Rowley JW, Stubben C, et al. Platelet gene expression and function in patients with COVID-19. Blood. 2020;136(11):1317–29.

3. Sakka M, Connors JM, Hékimian G, Martin-Toutain I, Crichi B, Colmegna I, et al. Association between D-Dimer levels and mortality in patients with coronavirus disease 2019 (COVID-19): a systematic review and pooled analysis. Journal Médecine Vasc. 2020;45:268–74.

4. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick A, Rawlik K, Pasko D, et al. Genetic mechanisms of critical illness in Covid-19. medRxiv. 2020;17(8):25.
5. Wang W, Zhang W, Zhang J, He J, Zhu F. Distribution of HLA allele frequencies in 82 Chinese individuals with coronavirus disease-2019 (COVID-19). Hla. 2020;96(2):194–6.

6. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albiolos A, Invernizzi P, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N Engl J Med [Internet]. 2020;NEJMoa2020283. Available from: http://www.nejm.org/doi/10.1056/NEJMoa2020283.

7. Zhang Q, Bastard P, Liu Z, Le Pen J, Moncada-Velez M, Chen J, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science (80-). 2020;4570:eabd4570.

8. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005;33:514–7. (DATABASE ISS ).

9. Orphanet. an online database of rare diseases and orphan drugs. INSERM, http://www.orpha.net Accessed in 4 March 2020.

10. Hunt SE, McLaren W, Gil L, et al. Ensembl variation resources Database (Oxford). 2018;2018(8):1–12.

11. WHO. Coronavirus Disease (COVID-19) Dashboard, https://covid19.who.int Accessed in 2 November 2020.

12. Ayres M, Ayres M Jr, Ayres DL, Santos AAS. Bioestat 5.0 aplicações estatísticas nas áreas das ciências biológicas e médicas. 5th ed. Pará-Brazil: Belém; 2007.

13. Smith NL, Huffman JE, Strachan DP, et al. Genetic predictors of fibrin d-dimer levels in healthy adults. Circulation. 2011;123(17):1864–72.

14. Masud R, Qureshi IZ. Tetra primer ARMS-PCR relates folate/homocysteine pathway genes and ACE gene polymorphism with coronary artery disease. Mol Cell Biochem. 2011;355(1–2):289–97.

15. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population Stratification in Genetic Association Studies. Curr Protoc Hum Genet. 2017.;95(1):1.22.1–1.22.23.