doi: 10.4183/aeb.2020.346

REVIEW

HOW GENETICISTS CONTRIBUTE TO UNDERSTANDING OF COVID-19 DISEASE PATHOGENICITY

F. Grigorescu[1,2,*], C. Lautier[3]

[1]Direction of Clinical Research and Innovation (DCRI), Montpellier Cancer Institute, University of Montpellier, [2]Institut Convergences Migrations, Collège de France, Paris, [3]Nutrition & Genome, UMR204 NUTRIPASS (IRD, UM, SupAgro), University of Montpellier, Montpellier, France

Abstract

Human populations are faced to the COVID-19 pandemic due to the emerging SARS-CoV-2 coronavirus originating from Wuhan (China) and with dramatic Public Health consequences. Despite periods of panic, the scientific community demonstrated an incredible innovation potential and energy ending up in one year with new vaccines to be used in population. Researchers are interrogating on how individual genetic differences contribute to the diversity of clinical manifestations or ethnic and geographic disparities of COVID-19. While efforts were spent to understand mechanistically the infectious potential of the virus, recent progresses in molecular genetics and bioinformatics allowed the characterization of viral sequence and construction of phylogeographical maps of viral dispersion worldwide. These data will help understanding epidemiological disparities among continents and ethnic populations. Much effort was also spent in analyzing host genetics by studying individual genes involved in innate and immune responses or explaining pathogenesis of comorbidities that complicate the fate of elderly patients. Several international consortia launched already *Genome wide Association Studies* (GWAS) and whole genome sequencing strategies to identify genetic markers with immediate application in patients at risk of respiratory failure. These new genetic data are important not only for understanding susceptibility factors for COVID-19 but they also contain an important message of hope for mankind warranting our survival and health.

Keywords: COVID-19, SARS-CoV2, single nucleotide variation (SNV), genome wide association study (GWAS).

INTRODUCTION

In 2020, humanity experienced a dramatic COVID-19 pandemic due to SARS-CoV-2 coronavirus with serious consequences in Public Health as well as significant socio-economical and political consequences (https://www.who.int/). By November 2020, there were 19,362,465 reported cases of COVID-19, among which 106,142 were serious or critical cases. The outbreak continues to spread today, while 1,551,760 deaths are recorded around the world, including both developed and low- and middle-income (LMI) countries (https://www.worldometers.info/coronavirus/). This pandemic is the largest viral outbreak in the last 100 years since *Influenza* H1N1 pandemic in 1918, which killed more than 50 million people (1, 2). The actual COVID-19 is due to an emerging coronavirus occurred in Wuhan (China), which rapidly spread widely in all countries with some unusual manifestations. Although coronaviruses were known by virologists, this family of viruses - unlike *Influenza* – was not in the main focus of geneticists because the zoonosis they produced accounted for common colds in children and were geographically limited. In the last two decades, attention was however drawn by SARS (severe acute respiratory syndrome) due to SARS-CoV-1 and MERS (Middle East Respiratory Syndrome) due to MERS-CoV, responsible for 8000 and 3000 deaths, respectively (3). The striking feature of COVID-19 pandemic - differently from *Influenza* that affects young people - is that the great majority of infected people have none or only minor symptoms. Indeed, only 20% of the population of older age shows a disproportionate severity of pulmonary failure leading to acute respiratory distress syndrome (ARDS) or death in the absence of an appropriate care. Elderly patients at high risk are those with co-morbidities such as diabetes mellitus, high blood pressure (HBP), cardiovascular (CV) diseases and obesity, all components of the metabolic syndrome (MetS). In a short time-period, the high number of severely affected patients overcrowded intensive care units and almost all governments were constrained to impose

*Correspondence to: Florin Grigorescu MD, PhD, Scientific Editor, Institut du Cancer de Montpellier (ICM), Department of Clinical Research and Innovation (DCRI), E-mail: florin.grigorescu@icm.unicancer.fr

strong sanitary measures culminating with lockdowns. COVID-19 pandemic mobilized the entire medical community from practitioners to researchers with an incredible enthusiasm of scientific collaboration. In one year, the viral genome was sequenced and analyzed for different strains and numerous vaccines were developed (4). By December 8th 2020 the UK already started vaccination of the population. All these events translate an unbelievable innovation potential of research despite periods of fear and even panic among people. Never, in our history the scientific approach was so well placed and so rapidly translated into medical progresses. Altogether these events represent a deep message of hope for mankind.

During the XXIIIrd Symposium of the Romanian Psychoneouroendocrine Society, held in Bucharest on 5-7 November 2020 - during the lockdown period - internists and endocrinologists raised the question of how geneticists could explain clinical features of COVID-19, susceptibility to severe complications and disparities among ethnic and geographically dispersed populations. The influential role of genetic factors in infectious diseases is not new. For *Influenza* pandemic, there are compelling evidences suggesting that individual genetic differences contribute to the severity of the disease. Thus, studies during *Influenza* outbreaks (e.g. "swine flue" pandemic in 2009) revealed several genetic markers as SNV (single nucleotide variations) in the genome, associated with pneumonia and located in genes such as interferon response factor 7, Fc fragment of immunoglobulin G or complement components (1, 5).

In the context of enormous progress in molecular genetics and bioinformatics, geneticists intended to explain variability among populations and how they contribute to the understanding of COVID-19 susceptibility. Indeed, research laboratories are now able to better evaluate genetic components using Genome wide Association Studies (GAS), whole genome sequencing or phylogenetic analysis. Moreover, together with anthropologists, they are even able using ancient (a)DNA to go thousands of years back in our history to trace susceptibility markers. In this paper we will focus on two major aspects: one related to virus genetics and another on global strategies using whole genome analysis.

Genetics of SARS-CoV-2 virus

A first source of variation in COVID-19 pandemic is coming directly from genetics of the virus. The International Committee on Taxonomy of Viruses (ICTV) named the emerging coronavirus 2019-nCoV as SARS-CoV-2. Coronaviruses are classified as α (HCoV-NL63 and HCoV-229) and β (HCoV-OC43 and HCoV-HKU1) coronaviruses. SARS-CoV2 is a single stranded RNA virus, which compared to HIV, hepatitis C or *Influenza* possesses a large genome (> 30 kb). Variations occur due to the high rate of recombination, which represents the common source for cross-species transmission. The virus also displays a high rate of mutation (10^{-10}) but lower than other viruses because it possesses a proofreading mechanism, thus keeping the virus to accumulate mutations (7). The viral genome was sequenced in the beginning of year 2020 (6). Based on sequences, the virus likely originated from bats (*Rhinolophus affinis*), which represent the natural reservoir. By spontaneous mutations/ recombination the virus acquired the pathogenicity and jumped in humans, spreading then from human-to-human (8). The transmission vector remains unknown but several live-animals sold on Huanan market in Wuhan (e.g. Pangolin) were suspected. The role of Pangolins (*Manis javanica*) was excluded since the virus is harmful for these animals. Moreover, SARS-CoV-2 has 96% homology with bat version (RaTG13) but much less with Pangolin virus. Two thirds of the viral genome contains ORF1a and ORF1b regions and another terminal domain encodes for spike (S) protein, envelope (E), membrane (M), nucleocapsid and other helper proteins. The spike protein drew much attention because through this molecule the virus expresses its cellular tropism by binding to ACE2 (angiotensin-converting enzyme 2 receptor) on the cell surface. The S protein is composed of a N-glycosylated ectodomain, a transmembrane domain and a cytoplasmic tail and is further divided in S1 and S2 subunits. C-terminal of the S1 domain contains the specific receptor binding domain (RBS) for ACE2 (9). During viral infection, a crucial step is the cleavage of S protein (called priming) at specific cleavage sites by cellular proteases such as furin, trypsin, TMPRSS (serine protease transmembrane protease serine 2) and cathepsin L (10). Digestion by furin involves a specific motif (R-X-X-R) and occurs in the Golgi compartment while other proteases operate on the cell surface.

Sequence analysis and biochemical studies indicated that compared to the bat progenitor (RaTG13), several mutations occurred with particular functions. Following mutations, two main features were gained by SARS-CoV-2. Firstly, 5 mutations occurred in RBD of the S protein increasing the ability to bind with high affinity to ACE2 receptor. Secondly, another feature

was dictated by occurrence of a polybasic cleavage site between S1 and S2 subdomains, thus allowing the cleavage by furin (11). This last modification is supposed to increase infectivity. Close related to cleavage site at S1-S2 junction, the addition of a proline residue resulted in the ability to bind O-linked glycans, which flanks the cleavage site. The function of these O-linked glycans remains poorly understood, but they are supposed to create a mucin-like domain, used by viruses in immunoevasion. It is not clearly established whether these mutations occurred in bat and then jumped in humans or they appeared during a still undetected human-to-human transmission (11).

Numerous sequences of the viral genome were compiled by the Global Initiative on Sharing Avian *Influenza* Data (GISAID) website and are publicly available (https://www.gisaid.org/). Updates are also kept by China National Center of Bioinformatics ((https://bigd.big.ac.cn/ncov/tool/annotation?lang=en) and NEXTSRAIN portal (https://nextstrain.org). Accumulation of such mutational data allowed the phylogenetic network analysis, methods previously used by researchers to reconstruct prehistoric population movements of *Homo sapiens* or language prehistory (12). These procedures are powerful and able to identify the virus origin and how it spread in timeline. The first report by Forster *et al.* 2020 contained analysis of 160 viral genome sequences and revealed A, B and C subclades. The clade A is the ancestral type. Type A and C spread in Europe and America while the B form in East Asia. Moreover, this analysis was able to trace infection sources at individual level. For instance, the Mexican strain was detected in an individual traveler from Italy, clearly explaining how the virus was spread. In another recent report, Gómez-Carballa *et al.* (13) by studying more than 4700 SARS-CoV-2 genomes of high quality, reconstructed virus variations by phylogeographic large-scale analysis and was able to investigate dynamics worldwide. By analyzing numerous mutations and considering the timing of samples, several macro-haplogroups were detected (A and B) both of Asian origin as well as more than 160 sub-branches. The haplo group A originated from Asia and had the highest frequency in Europe and Africa (97 and 93%) but remaining at a prevalence of 77% in Asia and only 53% in America. The sub-clade A2 originating in Europe from an Asian ancestor is by far the most frequent worldwide. Multiple founder episodes were identified. For instance, the first A2 genome was detected in a Bavarian (German) patient and another sub-clade A3 was the most common in Asia. Another subclade A2a2a was detected in North America, but which was probably of European origin (France); the clade A2a4 was found with the highest frequency in Switzerland. The second major haplogroup B was very frequent in North America (47%), South-America (32%) and Asia (23%). Finally, the clade B3a emerged in Europe (high prevalence in Spain) and from where went to South America (71% of all B genomes). These data are compatible with the idea that the virus occurred in Asia and infected one person (index case) in Hubei (China) in a silent mode up to November 2019 when the outbreak was declared. Analysis also suggested that some mutations occurred in interval of days, compatible with super-spreaders and founder effect in geographical areas, as for instance, the case of haplotype H4 in the haplogroup A. Conversely, there are data indicating that some strains were at high frequency in several restricted geographical areas such as the haplogroup A3 in Japan, or haplotype H8 (haplogroup A2a5a) in Iceland. Phylogeography analysis may also help the understanding the potential effects of other newly discovered mutations. For instance, Korber *et al.* detected a new mutation D614G, which is supposed to enhance the infectivity of the virus (14). These data are still debated, but from phylogeographic analysis it appears that D614G mutation is located on haplogroup A2, containing by itself another A23403T mutation that better correlates to high impact of COVID-19 in Europe and particularly in Spain. At present, there are no correlations found between these subclades and clinical variability of COVID-19, either with age or sex. However, these studies are crucial in understanding the variability of infectivity among continents and between ethnic populations.

Another aspect that drew the attention of geneticists was the ethnic variability of COVID-19 and geographical disparities. Variability among ethnic groups is very complex involving multiple factors. A recent report reviews these aspects and indicates that 34% of COVID-19 affected individuals in the UK are racial minority groups (15). Similarly in the USA, racial minorities groups are disproportionately harmed by the pandemic. In several states, African-Americans registered 52% of deaths although they represent only 30% of the population (15). There is no doubt that these disparities may be also explained by socio-economic or environmental factors. Inequalities in the place of work and high risk of exposure, crowded houses and health-care disparities, all could explain the ethnic variation. Although males are more affected than females in all populations, a recent report of the Office for National

Statistics of COVID-19 in the UK indicated that the African man and women are 4.2 and 4.3 more likely to have a COVID-19 death than White population (16). Many other chronic conditions are more prevalent in minorities groups in Europe and further explained by inequalities, adherence to medication and socioeconomic disadvantages. However, it is not excluded that genetic factors of susceptibility also contribute to these disparities. In Africa, the first confirmed case was reported in Egypt in February 2020 (17). In the whole continent there are 1.3 million confirmed cases with 30,370 deaths, relatively lower than in Europe. This may be explained by numerous factors (18). Africans are characterized by high genetic diversity that might influence resistance to infections. Such situations are documented for HIV, malaria or trypanosomiasis. Moreover, the African continent experiences numerous infection diseases and outbreaks in front of which governments take drastic measures at an early stage. The mean age of the population (19.7 years) and a small proportion of old individuals (> 60 years) would be another factor. It is not excluded that previous coronaviruses infections prepared (immunologically) the population. All these factors should be balanced to genetic factors from the virus itself or from the host population.

Host genetics

Another major source of variation in clinical and epidemiological aspects of COVID-19 is the genetics of the host. This domain implies many aspects, including immunologic features of innate or adaptive response as well as physiopathological mechanisms (19). Clinically, the COVID-19 disease is characterized by occurrence of mild symptoms such as fever, fatigue, cough, myalgia, headache or diarrhea. In high proportion the virus infection remains without symptoms but some patients present pulmonary complication, with typical CT scans followed by respiratory failure necessitating mechanical ventilation and oxygen supply. The respiratory failure could lead to the distress syndrome (ARDS) or death. The critical observation was the occurrence of serious cases in elderly patients with co-morbidities (20). Patients with respiratory failure display for 1/3 diabetes mellitus or CV-diseases and almost half have HBP or obesity (21). This clinical picture complicates the work of geneticists since case-control studies cannot be performed in asymptomatic patients and because many of these features are rather explained by immunological alterations or physiopathological mechanisms (22, 23). However, novel strategies developed in genetics, were able, even under these circumstances, to detect genetic components. One classical model used by geneticists is the twin study. Indeed, a recent report from the UK investigated the clinical manifestations of COVID-19 in 2633 twin volunteers using the "Covid Symptoms Tracker" application (https://doi.org/10.1101/2020.04.02.20051334). Investigators predicted the heritability of COVID-19 by 50% (29-70%), including 41% heritability for fever, 47% for anosmia and 49% for delirium (24). Another good example of genetic approach was offered by the study of obesity. Obesity is a common component of COVID-19. CDC in the USA estimated the prevalence of 28-30% (21). The role of obesity was explained by multiple factors such as co-existence of hypoventilation, obstructive sleep apnea, cellular stress and chronic state of systemic inflammation with tissue hypoxia. All these lead to a pre-inflammatory state with increased C-reactive protein and high levels of IL-1, IL-6 and TNFα. Of particular importance is the presence of visceral adiposity with high expression of ACE2 (24). An interesting study was performed in 489,769 individuals (with 28% obesity) registered in the UK Biobank, among whom 641 patients displayed severe COVID-19 forms (25). Central obesity was associated with significantly higher risk of COVID-19. Moreover, a Polygenic Risk Score (PRS) using genome wide data from the GIANT consortium (Genetic investigation of Anthropometric Traits) showed an association with higher risk of outcomes ($P < 0.004$ and Z-score of 1.14). These data indicated that central obesity and the genetic predisposition to obesity represent a higher risk of developing severe forms of COVID-19.

In the hypothesis of a genetic susceptibility for COVID-19 disease, geneticists are faced to complex situations because the risk of exposure in the general population is pending on many factors such as transport of people from one geographical region to another. For instance, Iceland is a quite particular country where genetic studies performed by DECODE genetics indicated a homogeneous population and ideal genetic studies. A recent report from this country indicated that the first case of COVID-19 was a person coming from Italy (26). And, there are 7 million of travellers per year through the unique airport of Iceland. In this population it was observed a lower incidence of COVID-19 in children and females compared to adults and males. While Iceland may represent an excellent model to understand genetics of COVID-19, the Mediterranean populations are by contrast more complicated since

Mediterranean borders were places of admixed populations for millennia, including the admixture with North African populations (27-29). It is worth saying that analysis of genetic data triggering the MetS and its components from the MEDIGENE program ((https://cordis.europa.eu/result/rcn/195753_fr.html) in the Mediterranean area would be of crucial importance.

A recent study in bioinformatics reviewed candidate genes that would be involved in COVID-19 pathogenicity (30). Many of these genes were known from *Herpex simplex* encephalitis, ARDS due to *Influenza* virus, infections by Varicella Zoster viruses and some other immunodeficient syndromes. A number of 40 genes were proposed as candidates and involved in Toll like Receptor (TLR), lectin and inflammasome pathways. We scrutinized these 40 genes among the 1511 genes for insulin resistance or 175 genes for MetS in the HuGE navigator (https://phgkb.cdc.gov/PHGKB/hNHome.action) and found only 6 genes in common with insulin resistance - HLA-B of major histocompatibility complex, CCL5 (Chr17q12), ABO (Chr9q34.2), MBL2 (Chr10q21.1) and TRAF3 (Chr14q32.32) and only one CCL2 (Chr 17q12) with MetS. Since many of these genes encode for cytokines, one might speculate on their potential role in severe forms of COVID-19, although more detailed studies would be necessary to reveal culprit genes.

The host genetics is a large field of research and concerns besides immunological aspects individual genes involved in the viral infection or various steps of metabolism. The most studied genes are ACE2 and TMPRSS2 (31). ACE2 is located on Chr Xp22.2 and encodes for angiotensin-converting enzyme 2 receptor catalyzing the conversion of angiotensin II to angiotensin 1-7. The interest in this gene was dictated by the fact that ACE2 protein is the binding receptor for the spike protein and because renin angiotensin system is involved in regulation of the blood pressure. Its dysfunction generates CV and kidney disorders. ACE2 is widely expressed in tissues, including the lung. Tissue distribution of ACE2 might explain manifestations at the level of the gastro-intestinal tract. TMPRSS2 is a serine protease involved in the priming of spike protein. The gene is located on Chr21q22.3 and its expression is high in ciliated cells and alveolar epithelial cells of the lung. Moreover, an androgen-responsive enhancer upstream of TMPRSS2 suggests a potential explanation of male to female disparities in the severity of COVID-19. These aspects of host genetics were largely reviewed in the literature (19, 22, 23, 32, 33). Here we would like to focus on global genetic approaches using GWAS or whole genome sequencing. Of note, at the international level, efforts were spent to understand host genetics. Thus, several consortia were created: COVID-19 Host Genetics Initiative, COVID Human Genetic Effort and Genetics of Mortality in Critical Care (GenOMICC). For the reader we indicate that a curated hub (LitCOVID) from the National Library of Medicine tracks all publications on COVID-19 and corresponding genes involved (34).

A recent study was offered by COVID-19 Host Genetics Consortium and published in New Engl. J. Med (35), which performed a case-control association study in two populations from epicenters of COVID-19 (Italy and Spain). A number of 1980 participants were included presenting positive PCR tests and pulmonary failure requiring oxygen only or additional mechanical ventilation. The patients were compared to 2381 controls from both countries. The GWAS revealed two genomic regions: one at the locus Chr3p21.31 and another one on Chr9q34.2. The region on Chr3 (centered by rs11385942) indicated potential implications of six genes: SLC6A20, LZTFL1, CCR9, FYCO1, CXCR6 and XCR1. The association of the leader SNV was significant with $P < 1.15x10-10$, OR 1.77. The precise gene is not identified yet. However, genes in this candidate region are either largely expressed in the lung (LZTFL1) or functionally related to ACE2 such as SLC6A20, which encodes the sodium–aminoacid (proline) transporter 1 (SIT1). The marker on Chr 9 (rs657152) corresponds to ABO blood groups, confirming some other previous results. Indeed, the distribution of ABO blood groups among 1775 infected in Wuhan (China) and 23,386 people from Shenzhen City indicated that blood group A had a significant higher risk for making the disease while the group O would be protective (36). More recently, a Mendelian Randomization (MR) study identified IFNAR2 gene involved in susceptibility or prognosis of COVID-19 (37). Gene Ontology (GO) enrichment analysis indicated multiple genes and tagged by the top 5 probes, including IFNAR2, TRIM5, NLRC5, MCTP1, PTPN1, FCER1G, ATF4 and PON2. Further, the deep sequencing and GWAS performed in Chinese individuals detected two other potential genes (TMEM189 and UBE2V1) on Chr20q13.13 in severe and mild forms of COVID-19. The candidate region was centered by rs6020298. The UBE2V1 gene encode for ubiquitin-conjugating enzyme E2 variant 1, which together with TMEM189 and TRIM5 is involved in interleukin-1 (IL-1) signaling and immune response.

Genetic anthropologists with the goal to trace

genetic markers in the human history then studied data from GWAS on COVID patients. A recent study from the group of S. Pääbo in Germany scrutinized the SNV the marker on Chr3 and previously found by GWAS in the Neanderthal genome. The SNV in this region of Chr 3 are in high linkage disequilibrium (LD) and restricted to a "core haplotype" of around 49 kb. This core haplotype extends to a weaker LD along a much longer haplotype (around 333 kb), which could enter by introgression from Neanderthals. The study indicated the leader SNV was homozygous in several samples of Neanderthal but none from Denisovans. By studying the clade of Neanderthal genome in the present day populations (1k genome project) it was observed that the core haplotype occurs at high frequency in South Asia (30%), much less in Europe (8%) or admixed Americans (4%) and virtually absent in Africans. The highest frequency was detected in Bangladesh.

In conclusion, these genetic studies are crucial in the understanding of the variability of the virus infection and manifestations of the COVID-19 in ethnic populations. One can speculate from these whole genome data that identification of more robust genetic markers for COVID-19 susceptibility may be helpful in order to prioritize individuals for vaccination, for instance. Genetics laboratories have shown in the last year an incredible capacity of innovation and intensive work should be now oriented towards identification of genetic factors of susceptibility for COVID-19. In these efforts both developed and LMI countries should be involved bearing in mind that understanding genetic diversity represents a virtue of mankind and a source of evolutionary resilience warranting our survival and health.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Short KR, Kedzierska K, van de Sandt CE. Back to the Future: Lessons Learned From the 1918 *Influenza* Pandemic. Front Cell Infect Microbiol. 2018; 8: 1-19.
2. Johnson NPAS, Mueller J. Updating the accounts: global mortality of the 1918-1920 'Spanish' *Influenza* pandemic. Bull. Hist. Med. 2002; 76: 105-115.
3. Sun J, He WT, Wang L, Lai A, Ji X, Zhai X, Li G, Suchard MA, Tian J, Zhou J, Veit M, Su S. COVID-19: Epidemiology, Evolution, and Cross-Disciplinary Perspectives. Trends Mol Med. 2020; 26(5): 483–495.
4. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020; 395: 565-574.
5. Ciancanelli MJ, Huang SX, Luthra P, Garner H, Itan Y, Volpi S, Lafaille FG, Trouillet C, Schmolke M, Albrecht RA, Israelsson E, Lim HK, Casadio M, Hermesh T, Lorenzo L, Leung LW, Pedergnana V, Boisson B, Okada S, Picard C, Ringuier B, Troussier F, Chaussabel D, Abel L, Pellier I, Notarangelo LD, García-Sastre A, Basler CF, Geissmann F, Zhang SY, Snoeck HW, Casanova JL. Infectious disease Life-threatening *Influenza* and impaired interferon amplification in human IRF7 deficiency. Science. 2015; 348: 448-453.
6. Hu D, Zhu C, Ai L, He T, Wang Y, Ye F, Yang L, Ding C, Zhu X, Lv R, Zhu J, Hassan B, Feng Y, Tan W, Wan C. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. Emerg Microbes Infect. 2018; 7: 154.
7. van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, Wertheim-van Dillen PM, Kaandorp J, Spaargaren J, Berkhout B. Identification of a new human coronavirus, Nat. Med. 2004; 10 (4): 368–373.
8. Sallard E, Halloy J, Casane D, van Helden J, Decroly E. Retrouver les origines du SARS-CoV-2 dans les phylogénies de coronavirus. médecine/sciences 2020; 36: 783-96.
9. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function and antigenicity of the SARS-CoV-2 spike glycoprotein. 2020; Cell 180: 281-292.
10. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C, Pöhlmann S. SARS-CoV-2 Cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. 2020; Cell 181: 271–280.
11. Kristian G, Andersen AR, Lipkin I, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020; 1-3.
12. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc Natl Acad Sci USA. 2020; 117(17): 9241-9243.
13. Gómez-Carballa A, Pardo-Seco XBJ, Martinón-Torres F, and Salas A. Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. Genome Reseaerch. 2020; 30:1434-1448.
14. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DJ, Evans CM, Freeman TM, de Silva TI, Sheffield, COVID-19 Genomics Group, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 2020; 182(4): 812–827.e19.
15. Bhala N, Curry G, Martineau AR, Agyemang C, Bhopal R. Sharpening the global focus on ethnicity and race in the time of COVID-19. Lancet. 2020; 395(10238): 1673-1676.
16. Office for National Statistics. Coronavirus (COVID-19) related deaths by ethnic group, England and Wales. May 7, 2020 (https://www.ons.gov.uk).
17. Rosenthal PJ, Breman JG, Djimde AA, John CC, Kamya MR, Leke RGF, Moeti MR, Nkengasong J, Bausch DG. COVID-19: Shining the Light on Africa. Am J Trop Med Hyg. 2020;102(6):1145-1148.
18. Musa HH, Musa TH, Musa IH, Musa IH, Ranciaro A, Campbell MC. Addressing Africa's pandemic puzzle: perspectives on COVID-19 infection and mortality in sub-Saharan Africa. Int J Infect Dis, 2020.
19. Jacob CO. On the genetics and immunopathogenesis of COVID-19. Clinical Immunology. 2020; 220: 1-16.
20. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R,

Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020; 395(10223): 497-506.
21. Shekhar S, Wurth R, Kamilaris CDC, Eisenhofer G, Barrera FJ, Hajdenberg M, Tonleu J, Hall JE, Schiffrin EL, Porter F, Stratakis CA, Hannah-Shmouni F. Endocrine Conditions and COVID-19. Horm Metab Res. 2020; 52: 471-484.
22. Elhabyan A, Elyaacoub S, Sanad E, Abukhadra A, Elhabyan A, Dinu V. The role of host genetics in susceptibility to severe viral infections in humans and insights into host genetics of severe COVID-19: A systematic review. Virus Res. 2020; 289: 198163.
23. Ovsyannikova IG, Haralambieva IH, Crooke SN, Poland GA, Kennedy RB. The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity. Immunol Rev. 2020; 296(1): 205-219.
24. Williams, FM, Freidin MB, Mangino M, Couvreur S, Visconti A, Bowyer RCE, Le Roy CI, Falchi M, Sudre C, Davies R, Hammond C, Menni C, Steves CJ, Spector TD. Self-reported symptoms of covid-19 including symptoms most predictive of SARS-CoV-2 infection, are heritable. medRxiv. 2020.
24. Ritter A, Kreis NN, Louwen F, Yuan J. Obesity and COVID-19: Molecular Mechanisms Linking Both Pandemics. Int J Mol Sci. 2020; 21(16): 5793.
25. Zhu Z, Hasegawa K, Ma B, Fujiogi M, Camargo CA, Liang L. Association of obesity and its genetic predisposition with the risk of severe COVID-19: Analysis of population-based cohort data. Metabolism. 2020; 112: 1-7.
26. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl JL, Saemundsdottir J, SigurdssonA, Sulem P, Agustsdottir AB, Eiriksdottir B, Fridriksdottir R, Gardarsdottir EE, Georgsson G, Gretarsdottir OS, Gudmundsson KR, Gunnarsdottir TR, Gylfason A, Holm H. Jensson BO, Jonasdottir A, Jonsson F, Josefsdottir KS, Kristjansson T, Magnusdottir DN, le Roux L, Sigmundsdottir G, Sveinbjornsson G, Sveinsdottir KE, Sveinsdottir M, Thorarensen EA, Thorbjornsson B, Löve A, Masson G, Jonsdottir I, Möller AD, Gudnason T, Kristinsson KG, Thorsteinsdottir T, Stefansson K. Spread of SARS-CoV-2 in the Icelandic Population. N Engl J Med. 2020; NEJMoa2006100.
27. Haydar S, Paillot T, Fagot C, Cogne Y, Fountas A, Tutuncu Y, Vintila M, Tsatsoulis A, Chi PT, Garandeau P, Chetea D, Badiu C, Gheorghiu M, Ylli D, Lautier C, Jarec M, Monnier L, Normand C, Šarac J, Barakat A, Missoni S, Pugeat M, Poucheret P, Hanzu F, Gomis R, Macias JM, Litvinov S, Khusnutdinova E, Poiana C, Pasquali R, Lauro D, Sesti G, Trischitta V, Abdelhak S, Zenati A, Ylli A, Satman I, Kanninen T, RinatoY, Grigorescu F. Branched-Chain Amino Acid Database Integrated in MEDIPAD Software as a Tool for Nutritional Investigation of Mediterranean Populations. Nutrients. 2018; 10(10): 1392.
28. Grigorescu F. Medigene Program Rewarded with the European Stars Award 2017. Acta Endocrinol (Buchar). 2017; 13(4):512-513.
29. Z. Douma, C. Lautier, S. Haydar, T. Mahjoub, F. Grigorescu. Portability of GWAS results between ethnic populations: genetic markers for polycystic ovary syndrome (PCOS) in Mediterranean area. Acta Endocrinol (Buchar). 2019; 15(3): 364-371.
30. Elhabyan A, Elyaacoub S, Sanad E, Abukhadra A, Elhabyan A, Dinu V. The role of host genetics in susceptibility to severe viral infections in humans and insights into host genetics of severe COVID-19: A systematic review. Virus Res. 2020; 289: 198163.
31. Hou Y, Zhao J, Martin W, Kallianpur A, Chung MK, Jehi L, Sharifi N, Erzurum S, Eng C, Cheng F. New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. BMC Med. 2020; 18: 216.
32. Carter-Timofte ME, Jørgensen SE, Freytag MR, Thomsen MM, Andersen NSB, Al-Mousawi A, Hait AS, Mogensen TH. Deciphering the Role of Host Genetics in Susceptibility to Severe COVID-19. Front Immunol. 2020; 11: 1606.
33. Debnath M, Banerjee M, Berk M. Genetic gateways to COVID-19 infection: Implications for risk, severity, and outcomes FASEB J. 2020; 10.1096/fj.202001115R.
34. Stoeger T, Amaral LAN. COVID-19 research risks ignoring important host genes due to pre-established research patterns. eLife. 2020; 9: e61981.
35. Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, Fernández J, Prati D, Baselli G, Asselta R, Grimsrud MM, Milani C, Aziz F, Kässens J, May S, Wendorff M, Wienbrandt L, Uellendahl-Werth F, Zheng T, Yi X, de Pablo R, Chercoles AG, Palom A, Garcia-Fernandez AE, Rodriguez-Frias F, Zanella A, Bandera A, Protti A, Aghemo A, Lleo A, Biondi A, Caballero-Garralda A, Gori A, Tanck A, Carreras Nolla A, Latiano A, Fracanzani AL, Peschuck A, Julià A, Pesenti A, Voza A, Jiménez D, Mateos B, Nafria Jimenez B, Quereda C, Paccapelo C, Gassner C, Angelini C, Cea C, Solier A, Pestaña D, Muñiz-Diaz E, Sandoval E, Paraboschi EM, Navas E, García Sánchez F, Ceriotti F, Martinelli-Boneschi F, Peyvandi F, Blasi F, Téllez L, Blanco-Grau A, Hemmrich-Stanisak G, Grasselli G, Costantino G, Cardamone G, Foti G, Aneli S, Kurihara H, ElAbd H, My I, Galván-Femenia I, Martín J, Erdmann J, Ferrusquía-Acosta J, Garcia-Etxebarria K, Izquierdo-Sanchez L, Bettini LR, Sumoy L, Terranova L, Moreira L, Santoro L, Scudeller L, Mesonero F, Roade L, Rühlemann MC, Schaefer M, Carrabba M, Riveiro-Barciela M, Figuera Basso ME, Valsecchi MG, Hernandez-Tejero M, Acosta-Herrera M, D'Angiò M, Baldini M, Cazzaniga M, Schulzky M, Cecconi M, Wittig M, Ciccarelli M, Rodríguez-Gandía M, Bocciolone M, Miozzo M, Montano N, Braun N, Sacchi N, Martínez N, Özer O, Palmieri O, Faverio P, Preatoni P, Bonfanti P, Omodei P, Tentorio P, Castro P, Rodrigues PM, Blandino Ortiz A, de Cid R, Ferrer R, Gualtierotti R, Nieto R, Goerg S, Badalamenti S, Marsal S, Matullo G, Pelusi S, Juzenas S, Aliberti S, Monzani V, Moreno V, Wesse T, Lenz TL, Pumarola T, Rimoldi V, Bosari S, Albrecht W, Peter W, Romero-Gómez M, D'Amato M, Duga S, Banales JM, Hov JR, Folseraas T, Valenti L, Franke A, Karlsen TH. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N Engl J Med. 2020;383:1522-34.
36. Wu BB, Gu DZ, Yu JN, Yang J, Shen W. Association between ABO blood groups and COVID-19 infection, severity and demise: A systematic review and meta-analysis. Infect Genet Evol. 2020; 84: 104485.
37. Liu D, Yang J, Feng B, Lu W, Zhao C, Li L. Mendelian randomization analysis identified genes pleiotropically associated with the risk and prognosis of COVID-19. Journal of Infection. 2020.
38. Wang F, Huang S, Gao R, Zhou Y, Lai C, Li Z, Xian W, Qian X, Li Z, Huang Y, Tang Q, Liu P, Chen R, Liu R, Li X, Tong X, Zhou X, Bai Y, Duan G, Zhang T, Xu X, Wang J, Yang H, Liu S, Jin QJ, and Liu L. COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. EJHG. 2020; 28(6):715-718.
39. Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. Nature. 2020.