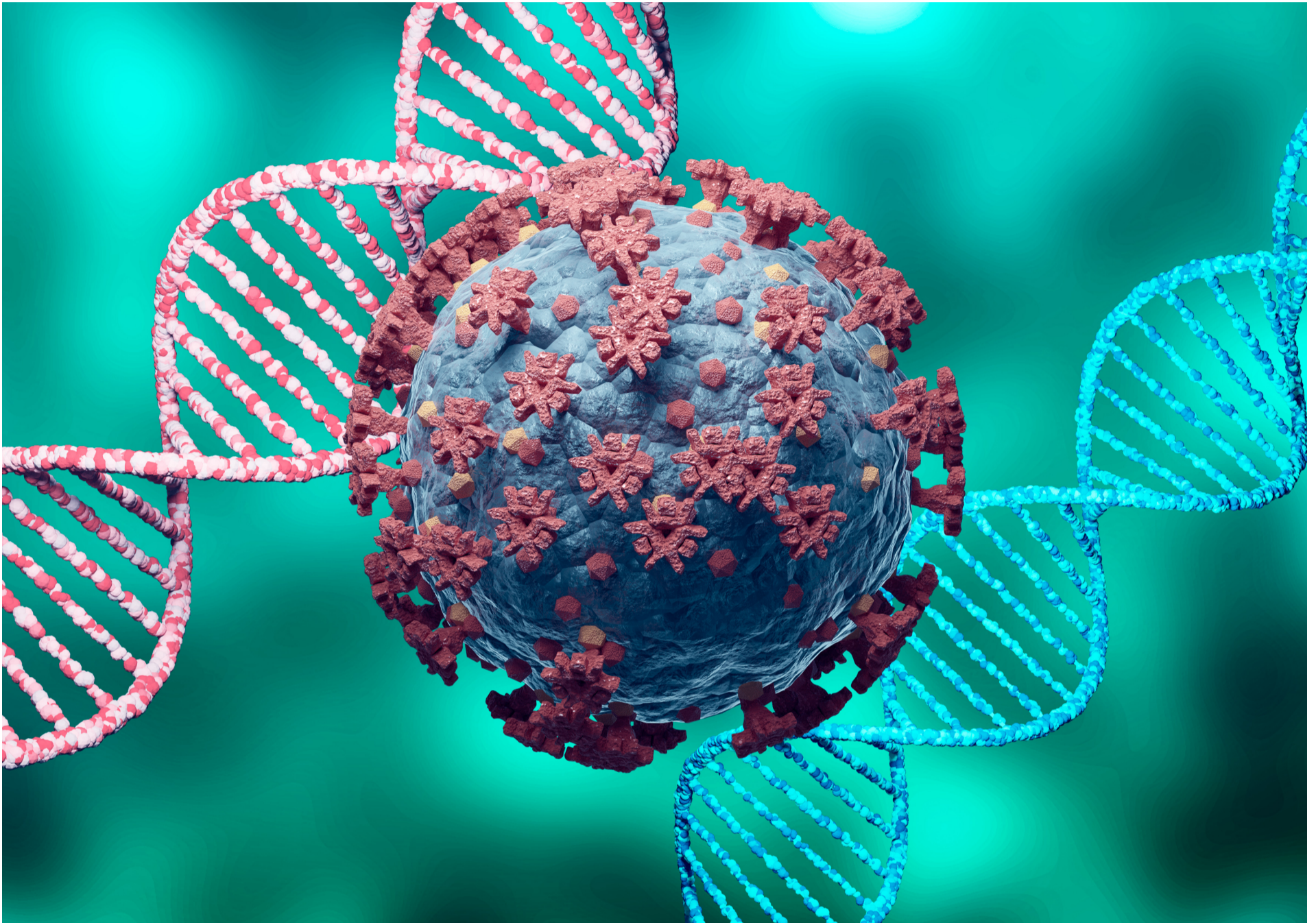


Genomic differences selected through evolution may offer clues as to why COVID-19 outcomes vary widely

A team from the University of Pennsylvania analyzed genomic data from global populations, including thousands of ethnically diverse Africans, to identify genetic variants that may be associated with clinical COVID-19 outcomes.



COVID-19's hard-to-predict effects likely owe in part to genetic differences. A Penn-led study analyzing the genomes of a diverse set of populations globally points to genetic variants that may help explain some of the variability in disease severity.

Though underlying medical conditions play an important role, many aspects of why COVID-19 severity can differ vastly from one person to another have remained unclear. A new study identifies dozens of genomic variants that may drive these hard-to-predict differences in clinical outcomes. According to work led by University of Pennsylvania scientists, genomic variants in four genes that are critical to SARS-CoV-2 infection, including the ACE2 gene, were targets of natural selection and associated with health conditions seen in COVID-19 patients.

The investigation, which used genomic data from diverse global populations, suggests that these variants may have evolved in response to past encounters with viruses similar to SARS-CoV-2. The team published the findings in the journal *Proceedings of the National Academy of Sciences* (<https://doi.org/10.1073/pnas.2123000119>).

“This study exemplifies my lab’s approach to genomic studies: We use what happens in nature and signatures of natural selection to identify functionally important variants that impact health and disease,” says Sarah Tishkoff, (<https://www.med.upenn.edu/tishkoff/Lab/Tishkoff/Tishkoff.html>) a co-corresponding author on the work and a Penn Integrates Knowledge University Professor (<https://pikprofessors.upenn.edu/>) with appointments in the Perelman School of Medicine (<https://www.med.upenn.edu/>) and the School of Arts & Sciences (<https://www.sas.upenn.edu/>). “Nature

has already done a lot of the screening and can give us clues as to what parts of a gene like ACE2 are important for infection.”

While other groups have conducted genome-wide association studies to identify genetic variants associated with COVID-19 severity, this is the first to include ethnically diverse Africans and a highly diverse dataset from the [Penn Medicine BioBank \(https://pmbb.med.upenn.edu/\)](https://pmbb.med.upenn.edu/). Including these often-overlooked groups revealed new variants that may be clinically significant.

Signals of selection

Before COVID-19 was even declared a pandemic, [Giorgio Sirugo \(https://www.med.upenn.edu/apps/faculty/index.php/g5455356/p9153923\)](https://www.med.upenn.edu/apps/faculty/index.php/g5455356/p9153923) of the Perelman School of Medicine hypothesized that there was a genetic basis for susceptibility to, or protection from, more severe outcomes.

“The idea is really a classic one, that infectious diseases have a host genetic component,” says Sirugo, a co-corresponding author on the paper. He reached out to Tishkoff and other colleagues to begin to address the question with a population genetics approach.

The researchers focused on a handful of genes known to play a role in how SARS-CoV-2 enters cells: ACE2, TMPRSS2, DPP4, and LY6E. They used genomic data from 2,012 ethnically diverse Africans, including people who practice traditional hunter-gatherer, pastoralist, and agriculturalist lifestyles, as well as 15,977 people of European and African heritage from the Penn Medicine BioBank, all of whom had associated electronic health record data available.

Looking for variations in these genes that showed evidence of being selected through evolutionary time, the researchers found 41 variants in the ACE2 gene that affected the amino acid sequence of the protein. Although these variants were rare when the team looked at the pooled global population, among a population of Central African hunter-gatherers, three variants were common.

“This really stood out to us,” says Tishkoff. “This is a group that lives in a tropical environment and continues to forage for bush meat, spending a lot of time in the forest. They’re likely exposed to all kinds of viruses introduced from animals. And, of course, SARS-CoV-2 is believed to have jumped from an animal to humans. So even though this population wouldn’t have been exposed to this exact virus in the past, they could have been exposed to similar types of viruses.”

These variants, in other words, might have evolved because they offered a protective effect against viruses with similarities to SARS-CoV-2. These variants showed signs of being positively selected, more evidence that they confer a fitness advantage.

Signs of natural selection were not only present in the parts of the genome that code for ACE2 and other genes, but also in what are known as regulatory regions, which affect how and where those genes are expressed. Many of these variants appeared to have been subject to what’s known as purifying selection, which occurs when evolutionary forces select for the removal of variants with negative impacts on fitness.

“We saw significant signals of natural selection in the regulatory regions of ACE2,” says Chao Zhang, a postdoc in Tishkoff’s lab and co-lead author. “I personally think that is going to be really important in thinking about clinical outcomes.”

“From an African and specifically Central African perspective, the discovery of three non-synonymous variants at ACE2 in Cameroonian indigenous populations is significant,” says Alfred K. Njamnshi, a coauthor and professor of neurology and neuroscience at Cameroon’s University of Yaoundé I. “The regulatory variants found at ACE2 do suggest targets of recent natural selection in some African populations, and this may have important disease risk or resistance implications that warrant further investigation.”

Rare variants are also likely playing a role in health outcomes, the team notes, accounting for individual-to-individual variation in disease severity. In East Asian populations, they found variations in the ACE2 regulatory region that may increase ACE2 expression, which could influence the degree to which SARS-CoV-2 infects host cells.

“To know for sure, we need to test the function of this variant and see whether we can get some indication that changes in this region are related to COVID infection susceptibility and severity,” says Yuanqing Feng, another Tishkoff lab postdoc who shared first authorship on the paper.

These variations in noncoding regions of the genome could also influence which organs the genes are expressed in, a relevant characteristic given COVID-19’s known effects on the heart, brain, lung, kidney, and other organs. In addition, the ACE2 receptor does not only play a role in binding to the SARS-CoV-2 spike protein; it is also involved in blood pressure regulation, and thus variants may affect health outside of just COVID infection.

Beyond ACE2, signals of natural selection were also apparent in the coding and regulatory regions of the TMPRSS2 gene, including variation that appears to have evolved after early human populations diverged from other great apes. “There are a lot of human-specific substitutions in that protein, which is really intriguing,” says Tishkoff, a suggestion that natural selection acted on these sites during human evolutionary history after the divergence from the ancestor of chimpanzees more than 5 million years ago. The team identified dozens more variants in the DPP4 and LY6E genes as well.

Genome-health connections

To get at the clinical relevance of these variants, the researchers made use of Penn Medicine BioBank data. The analysis was conducted in large part before the pandemic swept through the United States, and thus outcomes of COVID-19 disease were not part of patient medical records at the time. But because the biobank data contain genetic-sequencing information, the researchers were able to look at the genetic variants they had just identified and see if there were any links with medical conditions that were considered relevant to COVID-19 infection.

“With our data, we can look at the variants that were identified by Sarah’s team and link those with clinical data,” says [Anurag Verma](https://ritchielab.org/people/staff/verma-anurag) of Penn’s Perelman School of Medicine, a co-first author on the paper.

The team found certain variants of the coding regions they had identified were indeed associated with conditions with connections to or overlap with COVID-19, including respiratory disorders, infection with respiratory syncytial virus, and liver disease.

Building on these initial findings, the researchers say further exploration of key genetic variants could reveal a lot about how proteins function in the context of COVID-19 or other diseases.

“From a medical point of view, you could identify novel therapeutic targets, or even provide some personalized medicine depending on which variants a person had,” Sirugo says.

The team underscores the importance of looking in diverse populations for genome studies, as some of the newly identified variants that could be clinically significant were only identified in African populations that had not been investigated in this way previously.

“That is a deeply important and unique aspect of this study,” Tishkoff says.

[Chao Zhang](https://www.med.upenn.edu/tishkoff/Lab/Zhang/zhang.html) is a postdoctoral fellow in the Tishkoff lab at the [University of Pennsylvania](https://www.upenn.edu/).

[Anurag Verma](https://ritchielab.org/people/staff/verma-anurag) is an instructor in translational medicine and human genetics in the Perelman School of Medicine at Penn and associate director of clinical informatics and genomics, Penn Medicine BioBank.

[Yuanqing Feng](https://www.med.upenn.edu/tishkoff/Lab/Feng/feng.html) is a postdoctoral fellow in the Tishkoff lab at Penn.

[Giorgio Sirugo](https://www.med.upenn.edu/apps/faculty/index.php/g5455356/p9153923) is a clinician scientist and senior research investigator in Penn’s Perelman School of Medicine.

[Sarah Tishkoff](https://www.med.upenn.edu/tishkoff/Lab/Tishkoff/Tishkoff.html) is the David and Lyn Silfen University Professor in the Department of Genetics in the Perelman School of Medicine and the

Department of Biology in the School of Arts & Sciences at Penn and director of the Penn Center for Global Genomics & Health Equity.

Zhang, Verma, Feng, Njamnshi, Sirugo, and Tishkoff coauthored the study with the Perelman School of Medicine's Michael McQuillan, Matthew Hansen, Anastasia Lucas, Joseph Park, Alessia Ranciaro, Simon Thompson, Meagan A. Rubel, William Beggs, Daniel Rader, and Marylyn D. Ritchie; Marcelo C. R. Melo and Cesar de la Fuente of Penn Medicine and Penn Engineering's Machine Biology Group; Michael C. Campbell of the University of Southern California; Jibril Hirbo of Vanderbilt University; Sununguko Wata Mpoloka and Gaonyadiwe George Mokone of the University of Botswana; Thomas Nyambo of Kampala International University in Tanzania; Dawit Wolde Meskel and Guria Belay of Addis Ababa University; Charles Fokunang and Alfred K. Njamnshi of the University of Yaoundé I in Cameroon; Sarah A. Omar of the Kenya Medical Research Institute; Scott M. Williams of Case Western Reserve University; and the Regeneron Genetic Center.

Sirugo and Tishkoff are co-corresponding authors, and Zhang, Feng, and Verma are co-first authors.

The study was supported in part by the National Institutes of Health (grants X01HL139409, 1R35GM134957, R01GM113657, R01DK104339, R01AR076241, and R01LM010098) and the American Diabetes Association (Grant ADA 1-19-VSN-02).

CREDITS **Katherine Unger Baillie**
Writer

DATE **May 19, 2022**

SUBTOPICS **Genomics, Genetics, Coronavirus Research, Biology, Medicine, Research, Penn Integrates Knowledge Professors**

SCHOOLS **Perelman School of Medicine, School of Arts & Sciences**