Molecular pathology selected abstracts

Editors: Donna E. Hansel, MD, PhD, chair of pathology, Oregon Health and Science University, Portland; Richard D. Press, MD, PhD, professor and director of molecular pathology, OHSU; James Solomon, MD, PhD, assistant professor, Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York; Sounak Gupta, MBBS, PhD, senior associate consultant, Mayo Clinic, Rochester, Minn.; Fei Yang, MD, assistant professor, Department of Pathology, OHSU; Erica Reinig, MD, assistant professor and medical director of molecular diagnostics, University of Wisconsin-Madison; and Marcela Riveros Angel, MD, molecular genetic pathology fellow, Department of Pathology, OHSU.

Host genetic factors affecting COVID-19 disease susceptibility and severity

September 2021—More than a year into the COVID-19 pandemic, an essential question remains unanswered: Why do some people infected with SARS-CoV-2 develop severe life-threatening disease or die while others are asymptomatic or have only mild disease symptoms? Severity of COVID-19 has been shown to be negatively affected by "typical" host factors, such as increasing age, underlying medical conditions, male gender, higher body mass index, smoking, and lower socioeconomic status. Collectively, however, these traditional risk factors do not explain all of the variability in disease severity in the general population. Adding to the complexity of host factors, an international collaborative network of investigators, called the COVID-19 Host Genetics Initiative, has shown in a large genome-wide association study that many polymorphic loci across the human genome are highly correlated with COVID-19 disease susceptibility and severity. To better understand the role of genetics in SARS-CoV-2 infection, the network, a consortium of approximately 3,000 researchers and clinicians, pooled clinical and genetic data from 49,562 SARS-CoV2–infected patients in 46 studies across 19 countries and six ancestry groups. Two million control subjects were accrued from a variety of sources, including biobanks, other clinical studies, and direct-to-consumer genetic companies. This large number of study participants allowed the investigators to amass sufficient statistical power to address the role of human host genetic factors in disease severity. The latter is defined categorically as infection without hospitalization, hospitalization, or critical illness requiring respiratory support or causing death. By combining this phenotypic information with detailed genotype data, the investigators identified 13 human genomic loci that were associated with SARS-CoV-2 infection susceptibility (four loci) or disease severity (nine loci). Two of the loci were discovered only after including studies of people of East Asian ancestry in the meta-analysis, highlighting the value of including diverse populations in human genetic studies. In the genomic proximity of these 13 COVID-19 disease susceptibility loci were 40 candidate genes, many of which play a role in immune function or pulmonary pathophysiology, or both. One intriguing loci was near the *FOXP4* gene, which is linked to lung cancer. The FOXP4 variant associated with severe COVID-19 increases expression of the gene, suggesting that inhibiting the gene could be a potential therapeutic strategy. Other loci associated with severe COVID-19 included *DPP9*, a gene linked to lung cancer and pulmonary fibrosis, and *TYK2*, which is implicated in some autoimmune diseases. Results of the meta-analysis may inform future efforts to identify those at greatest risk of severe SARS-CoV-2 infection and identify novel therapies and vaccines to ameliorate poor outcomes.

COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature*. 2021. https://doi.org/10.1038/s41586-021-03767-x

Correspondence: Dr. Andrea Ganna at andrea.ganna@helsinki.fi

Predicting protein structures using artificial intelligence

It has long been theorized that the primary amino acid sequence of any protein should directly predict its active folded three-dimensional structure that largely dictates the protein's biological function. However, even though researchers have known for years the primary sequence of the approximately 20,000 proteins in the human proteome, only about one-third of those proteins have had their 3D structures determined experimentally.

Accurate computer models of protein structure based solely on primary sequence would be a scientific advance over laborious, resource-intensive experimental methods for determining protein structures, such as x-ray crystallography and cryo-electron microscopy. Understanding how a protein or protein complex is three-dimensionally oriented is a key step toward designing drugs that can modulate protein function and, therefore, treat a myriad of health issues, such as cancer, infections, and inflammatory conditions. Recent articles in *Nature* and *Science* described advanced computer modeling programs that can predict the 3D atomic structures of proteins given their primary sequence. One such artificial intelligence tool, called AlphaFold (DeepMind, London), has been shown to predict the structure of not only 98 percent of the proteins in the human proteome but also hundreds of thousands of nonhuman proteins from model organisms. For the human proteome, 58 percent of the software's predictions for the locations of individual amino acids were sufficiently accurate to inform the precise shape of the protein's folds. A subset of those predictions (36 percent) were potentially precise enough to detail atomic features useful for drug design, such as the active site of an enzyme. The approximately 350,000 predicted protein structures are more than twice as many as had been previously solved by experimental methods. The AlphaFold tool uses a novel machine-learning approach that incorporates physical and biological knowledge about protein structure, leveraging multisequence alignments, into the design of the deep-learning algorithm. Open-source code for the AlphaFold tool is accessible online (https://github.com/deepmind/alphafold), as is the database of its structural protein predictions (https://alphafold.ebi.ac.uk). Inspired by the AlphaFold tool, an academic team from the University of Washington has also created an artificial intelligence program for predicting protein structures, called RoseTTAFold. Open-source code for RoseTTAFold is accessible online at https://github.com/RosettaCommons/RoseTTAFold. The UW team has already used the tool to model more than 4,500 protein sequences submitted by other researchers. RoseTTAFold is nearly as accurate as AlphaFold and works on not only individual proteins but also complexes of proteins. For example, RoseTTAFold was used to create a structure database of hundreds of G-protein–coupled receptors, a class of common drug targets. Both programs use AI to spot folding patterns in vast databases of solved protein structures. The programs compute the most likely structure of unknown proteins by also considering basic physical and biological rules governing how the neighboring amino acids in a protein interact. These highly accurate in silico tools for predicting protein structure could initiate a fundamental paradigm shift in understanding how thousands of unknown proteins function. The practical applications for this new technology are immense and varied and include drug design and optimization, creation of novel enzymes for breaking down waste materials such as plastic, and development of crops that are resistant to viruses or extreme weather. The tools have already been used to better understand the novel viral proteins encoded by the SARS-CoV-2 virus.

Baek M, DiMaio F, Anishchenko I, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021. doi:10.1126/science.abj8754

Correspondence: Dr. David Baker at dabaker@uw.edu

Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021. https://doi.org/10.1038/s41586-021-03819-2

Correspondence: Dr. John Jumper at jumper@deepmind.com