

SPHERES

[Announcer] *This program is presented by the Centers for Disease Control and Prevention.*

[Susan Laird] Hi. Welcome to *Beyond the Data*, a podcast that gets to the heart of our monthly Public Health Grand Rounds and goes *beyond the data* to give you scientifically accurate, up-to-date, actionable information from CDC and our partners.

[Dr. John Iskander] I'm your host, Dr. John Iskander with CDC's Office of Science, and with me is Dr. Duncan MacCannell. He is the Chief Science Officer at CDC's Office of Advanced Molecular Detection. Welcome, Duncan.

[Dr. Duncan MacCannell] Thanks, John.

[Dr. John Iskander] CDC is leading a new research effort to determine how the coronavirus is spreading across the country. Can you tell us more?

[Dr. Duncan MacCannell] Today, I'd like to introduce a new national genomics consortium called SPHERES, which CDC recently put in place to help coordinate SARS-CoV-2 sequencing in laboratories across the United States.

[Dr. John Iskander] So SPHERES is an acronym, which stands for sequencing for public health emergency response, epidemiology, and surveillance. Before we talk about the significance of this sequencing effort, let's define the difference between SARS-CoV-2 and COVID-19 for our listeners.

[Dr. Duncan MacCannell] Sure, SARS-CoV-2 is the name of a particular type of coronavirus that causes the disease, COVID-19. The SARS-CoV-2 coronavirus is from the same genus as SARS and MERS, and was first identified in humans in Wuhan province, China in late 2019, likely as a spillover from wild bat populations.

[Dr. John Iskander] And could you briefly explain what sequencing is.

[Dr. Duncan MacCannell] Genomic sequencing is a method that reads the genetic code of any living thing – like a human, or a zebra fish, or a redwood – or a virus like the SARS-CoV-2 coronavirus. To sequence something, we usually break its genome into many shorter pieces of DNA, and we use special instruments to read the code that each of these smaller fragments contain. DNA is made up patterns of As, Ts, Gs and Cs — which is short for adenine, thymine, guanine, and cytosine. To figure out what it says we use bioinformatics and high-performance computing to reassemble fragments and examine sequence — or the order — in which the As, Ts, Gs, and Cs occur. Once we have the assembled sequence, we can compare it to other sequences we've already seen to determine how similar they are. That's how we can look at viruses and figure out what part of the world they might have come from, what animals they might have originated from, or how they might be changing over time. In public health, open data sharing has become a critically important part of this process. Public health laboratories

from all over the world share pathogen sequences in public databases, allowing us to rapidly compare data, respond to public health threats, and collaborate internationally. The SARS-CoV-2 coronavirus has a genome that is just under 30,000 As, Ts, Gs and Cs long, and it slowly accumulates changes to its genome as it reproduces and transmits through the world— about one change every two weeks, on average. By comparing these changes, scientists at the CDC and other public health agencies can follow the virus as it spreads, and use this information to find better strategies to prevent future transmissions. We can also monitor circulating viruses for any important changes that might affect how reliably we can detect and treat new cases, or that may affect how the virus behaves. The SPHERES consortium will enable us to sequence many, many thousands of SARS-CoV-2 genomes over the course of the pandemic by coordinating sequencing efforts in laboratories across the country.

[Dr. John Iskander] And, who is part of the consortium?

[Dr. Duncan MacCannell] Well, the consortium builds on investments in genomics and bioinformatics by my program, the advanced molecular detection program and others across the public health system. It includes almost 450 scientists from across the country, including 27 state public health departments and those range from New York and California to Alaska and Hawaii; it includes 17 county and local public health departments; and dozens of academic laboratories and bioinformatics groups; and a range of clinical and commercial laboratories, from very small startups up to big national diagnostic labs like LabCorp and Quest Diagnostic. We've also seen huge engagement from across the private sector, with software and systems manufacturers like Illumina, and Oxford Nanopore, Pacific BioSciences and ThermoFisher taking part. The scale of this public-private partnership is fairly new to public health, but it brings together world class expertise, institutions and innovators, with a common purpose and a shared set of goals.

[Dr. John Iskander] And briefly, what are the goals of SPHERES?

[Dr. Duncan MacCannell] Well, the goals of SPHERES are really quite simple. Our aim is simply to maximize the quality, quantity and usefulness of SARS-CoV-2 genomic sequence data from labs across the United States. And to ensure that these data are widely and openly accessible for public health and basic research. We're coordinating these efforts with other sequencing programs around the world – in the UK and Canada, for example – to make sure that the data we collect, and the information that goes along with it, is complete and consistent with global standards.

[Dr. John Iskander] Ultimately, what role will SPHERES play in monitoring the novel coronavirus?

[Dr. Duncan MacCannell] Well, large scale and rapid genomic sequencing of the virus that causes COVID-19 can help us in a lot of ways. For example, it will allow public health experts to do things like monitor changes in the virus as it continues to circulate and we can gain important insights to support things like contact tracing and infection control. We will provide crucial information about the stability of diagnostic and therapeutic targets. And perhaps most importantly, it will advance critical public health research in labs across the world in things like

transmission dynamics, host response, and evolution and many other domains. At the end of the day, the SPHERES consortium will help generate important new information about the virus that will strengthen our national COVID-19 response and mitigation strategy. And it will help build a strong new foundation for public health and basic research to better understand the virus and how to help keep people healthy and safe.

[Dr. John Iskander] It sounds like a very exciting project. Where can our listeners learn more about SPHERES?

[Dr. Duncan MacCannell] You can find it easily by searching for CDC SPHERES. S. P. H. E. R. E. S.

[Dr. John Iskander] Thank you, Duncan. As you know, CDC is a data-driven agency. We like to close our segment with CDC's numbers game. What's your number, Duncan?

[Dr. Duncan MacCannell] Today, my number is 29,903. That's the number of As, Ts, Gs and Cs that make up the SARS-CoV-2 reference genome.

[Dr. John Iskander] Thank you very much, Duncan. Please join us next time for our *Beyond the Data* podcast.

[Susan voiceover] Thanks for listening to *Beyond the Data*. Please join us next month. In the meantime, visit our Public Health Grand Rounds website and sign up for our mailing list. If you have questions, please email us at grandrounds@cdc.gov.

[Announcer] For the most accurate health information, visit cdc.gov or call 1-800-CDC-INFO