**ARTIFICIAL INTELLIGENCE**

**The scientists who swab subways for coronavirus**
Ride along on a subway-swabbing mission and meet scientists racing to find an existing drug that might treat the disease.
August 26, 2020

What weird bugs did you pick up last time you rode a subway train? Just as the covid-19 pandemic was taking off, a global network of scientists began mapping the DNA of urban microbes and using AI to look for patterns. Join host Jennifer Strong as she rides along on a subway-swabbing mission and talks to scientists racing to find an existing drug that might treat the disease.

**We meet:**
- Christopher Mason, Weill Cornell Medicine
- David Danko, Weill Cornell Medicine
- Baroness Joanna Shields, BenevolentAI CEO

# Credits:

**Full episode transcript:**

**Jennifer Strong**: You don't need me to tell you Covid-19 changed everything. It's changed the way we work. It's changed the way we school our kids. It's also changed the way we move around. It's no longer a simple decision to hop a flight or jump on public transit. It's forcing us to get creative with technology and, in some cases, speeding up innovation. People are trying to use artificial intelligence in all sorts of ways to help make sense of this new world and hopefully restore some normalcy. I'm Jennifer Strong and this episode we're heading down into the New York Subway because even as life as we knew it ground to a halt, researchers in over a hundred cities came together. In train cars and sewers, on hospital walls and buttons of ATMs, they're collecting data for AI to help them hunt for Covid-19 and other pathogens.

**Christopher Mason**: So there's this continual genetic mapping generated and it's in snapshots. But if we did it literally daily or weekly, we would be able to see the

emergence of new pathogens as they're happening and track them. And so I think we will use this as a warning system where we can be prepared.

**Jennifer Strong**: By taking samples and sequencing the genetics researchers can watch pathogens spread, like what happened after a Biotech conference in Boston that sent the virus across the state, across the country, across the world. And we know that because of the genetics. This work entails mountains of data which A-I can help make more useful by finding patterns, making predictions, and just maybe, spotting future pandemics and ways to treat them… earlier.

**Jennifer Strong**: In October of 1918, a flu pandemic that would be the most severe in modern history was starting to break out in New York. The city's Board of Health was trying to figure out how to slow the spread and keep businesses open. Without the internet, and promises of remote work, the constraints were a bit different... and a heated debate followed. Crowded subways were a potential disaster in the making and about a billion and a half trips were taken on the subway that year which is about the same as more recent years. They settled on an unusual decision. To stagger the operating hours of different types of jobs and businesses in order to relieve crowding on the subway. Textile manufacturing would open at 9. Jewelers at 7:45. Wholesalers at 8:15. But the subway kept running. People kept riding. And it appeared to work. Ultimately, the city's death toll per capita was much lower than in Philadelphia or Boston

**Jennifer Strong**: Fast forward to the present. As the coronavirus pandemic rages across the world, the debate over how to best manage the movement of people and goods does too along with an assumption that public transit systems are petri dishes for infection, and maybe largely responsible for the spread of Covid-19 in New York City. And if you're wondering what all this has to do with artificial intelligence? It's called Microbial Surveillance, or the tracking of populations of microbes, and other things we can't see.

**Christopher Mason**: How predictive can we be looking at fragments of DNA left on the soles of someone's shoes? You know you think about like curious scientists in masks and gloves and swabs, what does that have to do with machine learning, artificial intelligence, mathematical modeling. But it actually is the raw substrate of the data that we've been using to track where a virus is going, what are they becoming resistant to?

**Jennifer Strong**: Christopher Mason is a professor of genetics and computational biology at Weill Cornell Medicine in New York City.

**Christopher Mason**: And a very curious person looking at DNA and RNA of all kinds.

**Jennifer Strong**: This type of science uses biological data to train machine-learning algorithms and develop models. It requires significant computing power to help make sense out of huge amounts of information. So, what does it take to surveil a pathogen? Well, there's the tracking of changes in microscopic populations. Detecting mutations, and if a treatment exists, figuring out whether a strain is resistant to it. Plus, investigating outbreaks.

And in the case of coronavirus, researchers are also hurrying to identify surfaces where it may be active, like a handrail on public transit. Even before this new reality… maybe you've wondered what's on those surfaces? Christopher Mason certainly did. He started looking for answers by swabbing subways long before this pandemic.

**Christopher Mason**: You want to know what's there so there's an innate curiosity that was a part of it. I've also been living in New York now for almost 15 years and sometimes you touch a railing and it's weirdly moist when you thought it would be dry or it's sticky or you know, there's always things you find so there's that curiosity. I also, when my daughter got old enough to ride the subway one day she was kind of licking the pole when she was very young. So I really, um, I had this moment of complete parental terror of like, you know, what's happened, something, something has transpired, some microbial transmission has certainly just happened. But I just wanted to know what it was.

**Jennifer Strong**: Back then there was almost no information about what was on these surfaces. He founded a community with other like-minded researchers and together they created genetic maps of city transit systems. The group's called Metasub.

**Christopher Mason**: So Metasub stands for the metagenomics of subways and urban biomes. And the concept is, is to build this genetic map of the world around us that has previously really just been invisible. It started with just a handful of cities now it's 106 cities around the world, all profiling and swabbing, cataloging, mapping, and then modeling the data we find.

**Jennifer Strong**: He says this work could help cities better understand the spread and eventually, the recovery. So he mobilized researchers from the metasub group, already on the hunt for pathogens, to start looking for the virus.

**Christopher Mason**: We realized it was a unique opportunity to leverage an existing network of scientists, clinicians, city planners, epidemiologists, and other researchers

to say, okay, well let's set out a uniform protocol. And as the outbreak is ramping up, sampling in South Korea, looking in Poland, looking in the U-K, across the United States, looking in Brazil, out in Nigeria, looking anywhere we could that anyone who was ready to go and could still get out of their labs and their houses.

**Jennifer Strong**: I'm based in New York City, where those first samples were collected, and on March 17th, five days after Broadway turned off the lights I went hunting for the Coronavirus with researchers in Times Square. Before the pandemic, I passed through this station several times a week. That sound you hear is a man with a tambourine and a harmonica, and a regular part of the ambiance. I'm here with David Danko, a PhD student in Christopher Mason's lab, and we're making our way from the platform to a spot in the station just off a major walkway.

**David Danko**: Purely a practical concern, right? Ideally we would want to sample the busiest area, but you can't just sit there and swab and block people so we're right off of that. And we sample three sites here. We sample a floor sample, a hand railing sample, and a column sample. And the idea is to map sort of as you go higher up in the station are we going to find more things on the ground or we're gonna find more things on the handrail or we're gonna find more things on the column. We expect to pick up slightly different things.

**Jennifer Strong**: That scratching sound is him swabbing. At the time of this taping researchers were only sampling Times Square and Grand Central Station. Penn Station beneath Madison Square Garden came later and now testing is taking place in ten different stations, with researchers swabbing the turnstiles and kiosks too. Just to kind of describe the scene, he's swabbing this really filthy floor.

**David Danko**: We want to emphasize that what we pick up with microbes is not the same as what most people would consider to be dirty. What we consider dust as humans is not, is not really alive for the most part.

**Jennifer Strong**: It's not a very technical set up. Long swabs that resemble giant q tips. A wooden block with some strategically placed nails to help prop them up. Gloves to protect the samples from anything living on his hands. A duffle bag to hold it all. And patience.

**David Danko**: To get anything at all off of a subway surface. We have to really just do this for a really long time and cover a pretty wide surface area. And even then we don't get stuff all the time.

**Jennifer Strong**: Machine learning is used to more accurately identify the genetics. It can help pick out coronavirus from other viruses and bacteria picked up by the swab

from human hands, rodents—or whatever else you would expect to find on the surfaces of a New York subway.

AI can also be used to figure out mutations, and where geographically a pathogen came from. So, what did they find on those samples we took? We find out right after the break.

**Christopher Mason**: Basically our first batch of several dozen samples didn't show any SARS cov2 virus or the covid 19 virus.

**Jennifer Strong**: Once again, Christopher Mason from Weill Cornell Medicine.

**Christopher Mason**: And so I can't yet say which is the safest route in terms of transit except to say that any mode of transit where you can keep your distance is the best we've seen so far. We do, we do see flu, we see other variations of things like rhinoviruses. We do see other winter related infectious respiratory viruses. So we can see that things are there. But the Coronavirus is a relatively wimpy virus once it lands on the surface, you know there's been these reports that the virus can survive for many days. In some cases you have one or two days as much as seven or eight days on different kinds of surfaces. But in our testing we do what's called infectivity studies.

**Jennifer Strong**: In other words, his team looks for microbes but then also tests whether what they find can actually make somebody sick.

**Christopher Mason**: You need to have not only the virus which has the genetic information but wrapped up inside the nucleocapsid. So it has all the components of how it can get into your cells, like the spike protein people have heard about, all the functional essentially proteins that comprise this infectious virus particle. And if it doesn't have everything there, it can't infect you actually, which is a bit of good news. And also the MTA has been really ramping up the cleaning, people have been wearing masks more. So all these things help from what we've seen so far.

**Jennifer Strong**: There's no easy way to do this kind of testing. You take a sample, then see how fast it grows and infects cells, and this has to be done in a particular kind of lab that's safe for this work. Otherwise, you run the risk of making a whole bunch of infectious viruses. They're also mapping out microscopic life in medical environments. And that, perhaps not surprisingly, is where his group is finding this virus - in hospitals.

**Christopher Mason**: And what's interesting is there, we do see, in some cases, 50 - 60 percent of those areas when you sample them, you can actually see it there. And that's not surprising because it's obviously the medical environment where patients are sick

and where they're coming in, getting testing, obviously ill, and there has been, you know, really striking, really distinct cause that you can see it in some cases all over the room.

**Jennifer Strong**: There've been other surprises with this work too. Like that cities share a core urban microbiome, or a set of several dozen species of microbes that appear in 95-percent of all the samples taken anywhere in the world.

**Christopher Mason**: We found that about on average, half of the DNA doesn't match any known species, never been seen before until we sequenced that DNA. And that was striking because we thought maybe it'd be 5% maybe 20% but it really, depending on where you're looking, anywhere from half to even 80 or 90% of the DNA is from some things we've never seen before. Even though they're literally under our fingertips, we've never actually cataloged them to see what's there.

**Jennifer Strong**: Until recently he says this kind of surveillance was viewed as an expensive and labor intensive way to look for things like bioterrorism and antibiotic resistance.

**Christopher Mason**: But in the future, I think everyone's appetite for continual surveillance as a means of public safety will probably become much more appreciated and even standard. We all take for granted that there is a continual mapping and monitoring of any storm that's rising up in the Atlantic ocean, right? Because we want to know what's coming. We want to know if there's any risks that we should prepare for. We no longer have to be subject to the fancies of the universe. We can actually be predictive. And so I think a really simple analogy would be we've started doing sampling of sewage as well with all the metasub cities. We're doing sewage, we're doing the air, we're doing the cities, doing the hospitals. So there's this continual genetic mapping generated and it's in snapshots. But if we did it literally daily or weekly, we would be able to see the emergence of new pathogens as they're happening and track them. And so I think we will use this as a warning system where we can be prepared.

**Jennifer Strong**: Because, like with that hurricane example, proper planning and tracking could save lives by helping us be better prepared for future pandemics, rather than just react to outbreaks. We're also attempting to use A-I in a whole host of other ways in this pandemic… we're trying to have it do things like listen to coughs and diagnose patients... let people know when they've been exposed to covid… and supercharge the hunt for drugs that might prevent or treat it.

**Joanna Shields**: Someone needed to become the Google of biomedical information.

**Jennifer Strong**: Baroness Joanna Shields is the CEO of Benevolent AI. The company wants to change how medicines are discovered and brought to market. She worked at Google in the early days when they were creating algorithms to help people search the internet. She also ran Facebook in Europe.

**Joanna Shields**: This is about how do you curate all the world's relevant information to create an environment for scientists to innovate and to come up with new discoveries.

**Jennifer Strong**: So far this year, more than 50-thousand papers have been published about Covid-19 alone.

**Joanna Shields**: No scientist could possibly read the thousands of journals that are produced, every day, if not you know tens of thousands. And you can't keep up with any field, so what we aim to do is by developing natural language processing algorithm to give them the tools to enable them to identify things that wouldn't wouldn't necessarily be obvious to the human brain.

**Jennifer Strong**: BenevolentAI uses machine learning to sort through vast amounts of medical literature, and find patterns that doctors and researchers would likely miss. It then maps those relationships into a knowledge graph. Imagine the kind of graph you'd draw to map out your relationships on Facebook… but for the connection between viruses, drugs, and proteins.

**Joanna Shields**: In the knowledge graph, we have 24 different biomedical entities that we pull together, enable scientists to visually see how those entities interact, you know, genes and proteins and how they interact. What gene is up-regulated in that disease and what process is dysregulated in the body that causes that disease to happen and understanding the underlying cause of disease then enables the scientists to develop a treatment that will work and will be much more effective.

**Jennifer Strong**: So when covid-19 hit the buzzy, London-based company put its tools to use.

**Joanna Shields**: So the challenge that our scientists took on is how do we look at the existing treatments that are out there? And if there is any way we can make an impact or lessen the severity with a patient. And I called one of our leading scientists and I said, Peter, what, what can we do? And our graph is not optimized for infectious diseases. You know, infectious diseases is a completely different discipline. So I wasn't, I didn't have extremely high hopes that he would be able to find something immediately. And he said, I've been working on that all weekend.

**Jennifer Strong**: They combed through scientific papers on the virus, and added what's known about how it hijacks the body's proteins. Then, they sifted through existing drugs to see if any of them might stop that hijacking process and help reduce the severity of covid symptoms. Those drugs still have to go through clinical trials to prove they actually can address Covid but they can pass through the safety phases much faster than brand new drugs.

**Joanna Shields**: The team were able to identify a drug that that is currently being used for rheumatoid arthritis we think that could potentially be a treatment, we've published in the Lancet that our research has led us to this, but obviously with the caveat that we need to do the testing.

**Jennifer Strong**: The drug she's talking about is Baricitinib from Eli Lily. It's now in phase three trials. BenevolentAI isn't the only company working on this problem. Even before the pandemic, researchers started applying machine learning to the drug discovery process in hopes of speeding it up. Covid just gave it more urgency.

**Joanna Shields**: There's about 10,000 diseases that don't have treatments. So … there's so much work to be done. It's almost when you think of competition in this area, you think, excellent, please. I hope you're having great success because you know, there's so many people suffering from disease and there's over 300-million people suffering from rare disease that unless we change dramatically the economic model, or we can increase the efficacy, or we can reduce the time it takes to get drugs to market. We're not going to be able to address this. So you need technology to make a big impact along the way so that we can work on diseases that may only have a few thousand patients.

**Jennifer Strong**: Next episode, how should a self-driving car tell you it isn't driving anymore?

**Richard Corey**: The notification to let you know that Tesla no longer has control? Is four beeps. And we thought that doesn't seem like enough information (laugh) considering the neural processing and stuff we go through when we're driving a car and we started to look at what does this mean? How do you trust what's in front of you? How can we help the humans in the car?

**Jennifer Strong**: We'll also find out how Google's self driving unit, Waymo, composes the sounds for its cars in a particular key, hoping to help us relax.

This episode was reported and produced by me, Tate Ryan-Mosely, Emma Cillekens and Karen Hao. We had help from Benji Rosen. We're edited by Michael Reilly and

Gideon Lichfield. Our technical director is Jacob Gorski. Thanks for listening, I'm Jennifer Strong.