Calling the disease detectives

Understanding the genes of viruses is key to beating them, explains Andy Extance

There is still so much we don't know about Covid-19, and governments all over the world are scrambling to get answers. Who is most at risk? How is it spreading? How many people are likely to get infected? How will healthcare systems cope? How effective will different ways to control the spread of the coronavirus be?

These are questions that only science can solve and when it comes to disease, epidemiologists are the most powerful scientific detectives. In the UK and across the world, they're busy looking at a wide range of clues, including the coronavirus's biochemical makeup. Ultimately, it is their graft that will enable us to understand and finally beat the virus.

Epidemiologists study the patterns and causes of disease, as well as the methods to control them, so that the burden on society can be reduced. '[It] often involves many different people with different skills, from mathematicians and statisticians, to biologists and medical doctors,' says Sema Nickbakhsh, an epidemiologist at the Medical Research Council and the University of Glasgow Centre for Virus Research. 'There are many roles that epidemiologists play in disease outbreaks to assist both public health agencies and governments directly.' At universities, she explains, they seek new ideas for how to study disease, and at national public health agencies they take the best of these ideas and put them into practice.

According to Nick Loman from the University of Birmingham, one of the main puzzles governments are looking to solve is the transmission of Covid-19 from person to person. 'What's the [infection] risk of sitting in a restaurant with another infected person, versus going to a football match or just walking down the street?' he asks. 'Do masks work? What will happen when international travel starts up again? Do we need to maintain social distancing for a long time? Epidemiology can help with these questions.'

Careful measures

In general, disease detection work starts by thinking carefully about the details people in hospitals and labs might record about people. Epidemiologists like Sema might design 'surveillance studies', where they try to understand how many people the coronavirus has already infected. They might also be trying to work out how we can tell if someone is infected. Sema notes that using mathematical models of how a disease spreads from an infected person to those at risk is getting more important. 'These models are used by governments to predict how the spread of disease may change if [and when] we implement interventions, such as social distancing,' she says.

Nick explains that an important example of modelling has been counting how many people have the disease, and drawing a graph called an epidemic curve. 'The shape of that curve can show you the rate of growth,' he says. Early curves using counts from China, Italy and the UK followed a very fast growth exponential curve. This meant that 'urgent measures were needed', Nick says. Another important measure is the reproduction number, known as R0, which estimates how many other people someone with Covid-19 will infect. 'The aim of public health measures in an outbreak is to bring this number down below one,' Nick says. 'It is now thought that social distancing and stay-at-home measures have done this.'

Sema adds that the models are very flexible and that even though the details about how many Covid-19 cases and deaths there have been are uncertain, epidemiologists can use the models to test for different scenarios. 'As biologists and medical researchers learn more about the details of who is at risk and how the virus behaves in different conditions, the models can be updated,' she says. Adding better health data further improves the model predictions. Then epidemiologists can bring the models together with molecular information about the virus to understand how it spreads.

Molecular information is important because viruses have evolved to invade and hijack cells. Once inside, the virus forces cells to make copies of virus genes and proteins, which then help the virus to infect even more cells. Infected cells respond to the genetic instructions given by the virus to help it spread. 'It is important to understand the genetic makeup of a new virus, so that we can determine what kind of virus it is, and where it came from,' Sema says. 'This can help us to predict its likely impact on the population, and find ways to protect the world from new viruses in the future.'

This is where scientists like Nick come in. Nick uses tools that can quickly read virus genes, collectively called its genome, using methods called sequencing. He is currently leading Birmingham's part in the Covid-19 genome sequencing consortium in the UK. The consortium studies coronavirus genes made of RNA molecules that encode instructions on how to assemble small amino acid molecules into larger proteins, like those in coronavirus' spiky shell.

'Sequencing is now a vital tool in detecting viruses,' Nick says. When doctors in China saw the first cases, they immediately thought of a different coronavirus: severe acute respiratory syndrome, or Sars, a coronavirus that first emerged in China in 2002. But when they sequenced all the RNA from a patient sample, they found the virus was related but not identical to Sars. That's why the current coronavirus got a new name, Sars-CoV-2. 'Having a genome sequence of a virus is incredibly important because it gives you a view of the genes, and therefore proteins, that the virus can produce,' Nick says. He notes the detection of Sars-CoV-2's 'spike' protein may help us defeat the virus because we can potentially design vaccines that tell our immune systems to go after anything looking like this spike protein. However, he adds that it will be useful to look at human DNA as well. This might help explain why some people seem to be more severely affected than others.

Family trees

The advantages of sequencing aren't limited to virus detection, however. 'The molecular characteristics also tell us about the evolution of the virus,' Sema says. 'This tells us about the likely origin, [and] in the case of Sars-CoV-2 it looks like it already existed in wildlife before it crossed over to people.' This makes sense because the very first cases have been linked to Huanan Seafood Market in Wuhan, China.

'The genetic codes of Sars-CoV-2 viruses that have been generated around the world are shared on a big database that all researchers have access to,' Sema says. 'These databases also contain the genetic codes of many other viruses, including other kinds of coronaviruses that are closely related to Sars-CoV-2.' The codes are put there by scientists everywhere, and include viruses that infect people and many kinds of animals. As such, researchers can compare how closely related different viruses are.

Scientists known as evolutionary biologists have found that Sars-CoV-2 probably came from a virus that infected bat cells. 'When we compare the genomes of Sars-CoV-2 to all the other virus genomes we've ever seen, the closest match is a coronavirus that has been detected in bats,' Nick explains. '[But] it's still not that close. They probably shared a common ancestor as much as 50 years ago, but it's a good clue that bats may be an important origin.'

But we still don't know if the first Covid-19 cases came directly from a bat, or through another animal. 'We know that Sars-CoV-2 can affect many different species of animal, including ones that were sold live in the [Huanan] Seafood Market,' Nick says. 'So, it is possible that animals were infected in the wild from a bat and then transmitted to humans at the market. Some have said that pangolins might be a possible host, but the evidence for this is less strong than for bats.'

Once the virus was detected and its evolutionary origin more or less mapped, sequencing also helped with tracking it in the infected population. Scientists do this by comparing virus genes between patients with Covid-19, which reveals mistakes that happen when infected cells copy viruses. 'Helpfully for epidemiology, these mistakes are observed at a fairly fixed rate, [known as] the "molecular clock", which ticks at around two mistakes a month,' Nick says. Therefore when cases have similar sequences, they are recently related. When they are more different, the cases are not linked. From this, scientists generate family trees; you can see them on the Nextstrain website.

These family trees can be used to track the virus, for example in a care home, Nick explains. If all patients have the same viral genome, then people living there probably caught it from each other. If they have different genomes, the virus was brought in more than once.

Sema concludes that doing many such studies provide powerful clues. 'Eventually we will learn … whether certain geographical conditions favour or inhibit it,' she says. 'We will learn whether it can coexist among influenza and the other common cold coronaviruses that it is related to.' Disease detectives like Sema and Nick will gradually solve these and other Covid-19 mysteries, and help the world find the answers we need so much.