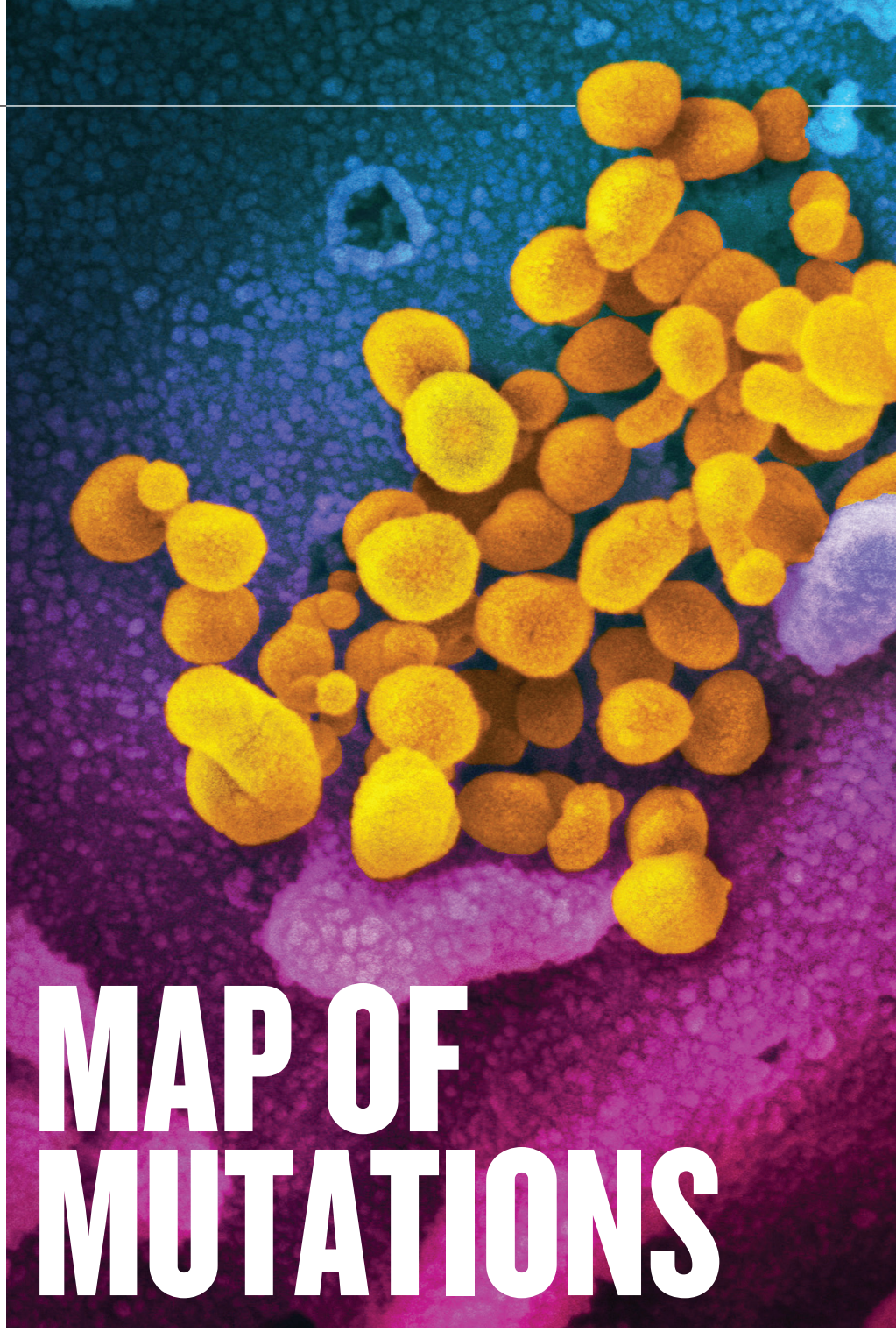

MAP OF MUTATIONS

Research from UCL has identified almost 200 changes to the COVID-19 genome, each one a clue to revealing the history of the virus and possibly its future.

As the COVID-19 story continues to unfold around us, research teams across the world have been toiling to unlock the secrets of the virus and provide vital intelligence in the race to develop an effective vaccine and other therapies. During the first week in May one of those teams, from UCL, published a paper called "Emergence of genomic diversity and recurrent mutations in SARS-CoV-2" in *Infection, Genetics and Evolution*. Having analysed some 7,600 complete samples taken from patients, they found that 198 sites in the SARS-CoV-2 genome have already undergone recurrent, independent mutations. Such mutations were always expected – it's what viruses do – but the team's results have also thrown a light on the possible genesis of the pandemic, and provide hints about the next chapter of the story.

One of the lead authors is Dr Lucy van Dorp, Research Associate at UCL Genetics Institute. She says the project began the moment they started comparing SARS-CoV-2 genomes shared by the global research community. "We were monitoring the virus sequences daily and looking at how the genomes have changed over the course of the pandemic, and we started to see some interesting things, which prompted us to ask a few key questions."

The first of those questions focused on the origin of the virus. "We saw genetic diversity in the virus, so could we use these patterns to predict its origin?" says van Dorp. "We could use the evolutionary rate of the virus to infer when it jumped from animals to humans." Their results suggest it could have occurred as early as October last year. "Some people think that's a bit early, but we think it's consistent with the case reports. It still highlights the fact that the virus hasn't been circulating in humans for long and was picked up quite quickly."

Patient zero?

The second question wondered if viruses sequenced from patients in the same country were similar to each other? "The answer to that was often not at all, surprisingly. Generally when you study the genetic data of pathogens infecting patients in a specific place, the genomes are more similar compared to any others circulating around the world. But not so much with SARS-CoV-2. For infections in many countries, any SARS-CoV-2 genome can recapitulate a large fraction of the global diversity, which fits with the many imported cases we've seen. Those many

"We could use the evolutionary rate of the virus to infer when it jumped from animals to humans"

IMAGES: SCIENCE PHOTO LIBRARY

introductions have gone on to seed local transmission events. It's a quirk in terms of understanding the genomes statistically, in that we have quite a random sample even if we look at infections at one location. It also means that so many introductions have happened at so many different times that the concept of a patient zero or index case in several countries doesn't make any sense, because there's hundreds or thousands of them."

The study's final question was whether genetic mutations were equally distributed across the whole genome, or do some parts vary more than others? "In this case we looked along the genome for mutations across thousands of SARS-CoV-2 sequences," says van Dorp. "These changes occur naturally when a virus replicates and are signposts of its evolution. We identified many mutations, though that wasn't a surprise. The virus itself is mutating in a similar way to other coronaviruses, but we did hone in on a small subset that appear to be both independent and recurrent."

"The idea is that if you see a genetic change, or mutation, occurring in multiple samples, arising independently of each other given the evolutionary history of the virus, these are most likely to be biologically relevant, pointing to how the virus is adapting to infecting humans or how its transmissibility might be changing. Right now we have identified 198 of these potential sites that we are continuing to monitor.

"The next stage of our research is to follow these up and to test if they have functional consequences."

Genetic puzzle

The analysis of further samples will demand more time and computational power, as the number of viral assemblies available at the time of writing had surpassed 17,000, with hundreds more appearing every day. "When we started in March, everything could be done on my laptop. It could cope with a few hundred virus genomes," says van Dorp. "But now we're into tens of thousands we need the high-performance computing clusters at UCL, which allow us to run not only high-intensity analysis but also analyses in parallel."

Despite the mounting intensity and importance of this work, it remains just one piece in the SARS-CoV-2 genetic puzzle. Plenty of other research is looking at the human genome – and van Dorp and the team are keeping a close eye on this, waiting for news of any leap in the knowledge of how the virus affects us during an infection. "Do we see mutations in the virus at those regions that interact with the human genome? If so, we want to be aware of them because they might be the exact interactions that will be targeted by drugs and vaccines."

It's a new virus, so in absolute terms it hasn't had much time to evolve significantly. Regardless, says van Dorp, closely monitoring any changes or signs of mutations is vital. "For example, a lot of vaccine design focuses on the spike membrane protein, which the virus uses to invade a human cell, so we want to be aware of any significant changes there as well as in other key structural regions of the virus."

Eventually, one of the many vaccines currently in development will make a breakthrough, and the work carried out by UCL and others in mapping the mutations will have played its part. "I think that what we've added is a tractable list of candidate adaptive mutations for this stage of the pandemic. This has been a massive, collaborative endeavour. I'm hopeful that a vaccine will appear more quickly than usual, though, of course, the day that happens does not spell the end." BMS

DR LUCY VAN DORP

- ✔ **2011:** BSc, Royal Holloway
- ✔ **2012:** MSc, Imperial
- ✔ **2013:** MRes, UCL
- ✔ **2017:** PhD, UCL Genetics Institute, UCL
- ✔ Her research applies computational methods to big genomic data to identify the genetic and ecological drivers of disease emergence, transmission and spread.