Twitter Thread by Dr. Rachel Glover ■■■

Dr. Rachel Glover ■■■
@rachomics

Okey dokey, let’s get this thread on the road. We’re going to sequence the viral genomes of a bunch of COVID-19 positive controls (ie not infectious, just inactivated viral RNA). This is aimed at the general public rather than scientists. 1/n

The first thing we have to do is to create DNA from the viral RNA. This is so the next step (PCR) will work. This is called reverse transcription and uses an enzyme obviously called reverse transcriptase. 2/n

In order to create the DNA from the viral RNA (it’s called cDNA) we have to add the reverse transcriptase and a buffer to the same and then incubate it for 10 minutes at 55C and then 1min at 95C to kill the enzyme. It’s just like cooking, but smaller. 3/n

While we wait for the cDNA incubation to finish, meet Bob. Bob is one of the @taxagenomics pipetting robots and until recently was the powerhouse of COVID-19 testing on the Isle of Man, isolating RNA from patient samples. 4/n

Now that we’ve created some cDNA we need to set up the PCR reactions. PCR is how we amplify the tiny amount of cDNA we have and make enough of it to sequence. It’s like a molecular photocopier. We also need to mix small volumes so here’s the whirly mixer (technical term) 5/n

We mix the right volumes and concentrations of these reagents to put all the nuts and bolts of the PCR reaction together. It doesn’t look like much once the samples are added. Now for more cooking. 6/n

The PCR reactions for COVID sequencing are “multiplex” which means there are many individual PCR reactions happening in the same tube. For this protocol it takes about 5 hours, so we leave it to work overnight because who has time to watch a boiling kettle? More tomorrow! 7/n

OK! PCR has finished overnight and now we’re just waiting for the reagents (chemicals, enzymes etc) for the next steps to defrost. Which means one thing: coffee time ■

Difficult to see but we now have four rows (B-E) with liquid in. Row E is the one with our PCR/enzyme mixture. This enzyme prepares the ends of the DNA created by PCR so we can attach other bits of DNA to them. Incubation time 30mins. 10/n

For demo purposes I’m doing the moving of liquids by hand so I can take pics for you. But no sane scientist would bother doing it by hand these days. Pipetting robots (like Bob, and his twin Hal) mean these processes are automated and hands-free. 10/n

So we've just added short pieces of unique DNA - called barcodes - to each DNA strand in each sample, which means we can identify which sample a particular piece of DNA came from after we've sequenced it. It'll now incubate for another 30 minutes. More coffee and admin! 11/n

If you haven't figured it out yet, molecular biology is really just pipetting lots of small volumes of liquid from one place to another. Then defrosting more small volumes of liquid and cooking them at particular temperatures for a while. 12/n

We do this by binding the DNA we want to tiny magnetic beads. Once the DNA is bound we put the tube on a cheap-tat magnet (@samtomindustrys) so the beads and DNA move to one side. Then we can remove the liquid we don’t want. Timelapse time. 14/n

These washing and rinsing steps happen a few times and are much quicker to do with a robot. It’ll take us around 20 minutes to complete. As an aside, 60mL of the magnetic beads costs the same as lots of people earn in a month ■ 15/n

We now have another tube of colourless liquid that contains DNA that we’re going to attach more things too. It’s incubating for another 20 mins. I did warn you all that live doesn’t always mean live action ■

More washing steps 16/n

So, after 4 hours we finally have what we call a “sequencing library” which is the DNA prepared with all the right bits and bobs attached to make the sequencing work. I remember now why we don’t normally do this manually ■ 17/n

Well done if you’ve stuck with me until this point! Meet the MinION, a nanopore sequencer. It’s definitely not a flashing stapler, even though that’s often how we refer to it. The snakey looking black plastic thing is called a Flowcell and that’s what does the hard work. 18/n

The sequencing library has been loaded onto the MinION and it’ll take a few hours to read all the ACTG’s on each DNA strand in the library. Nanopore sequencing is (relatively) new and it reads the DNA letters in a really interesting way 19/n

So tomorrow I'll go through nanopore sequencing, the data and analysis. If they were real samples they'd be prioritised and analysed in the next couple of hours. But they're not and I haven't had lunch yet ■■■■■

OK! So the raw data from the sequencer went to our server and now we need to do something called basecalling (i.e. call the letters ACTG from the data) and QC by trimming bad data and the sample ID barcodes. The biggest rule of bioinformatics applies: rubbish in, rubbish out 20/n

While that's processing, let me explain nanopore sequencing. Basically, a protein nanopore sits on an electrically resistant membrane with an ionic current passing through it. As the DNA feeds through the pore the current changes in an identifiable way 21/n

That allows us to assess both which base (DNA letter) is passing through at that point (well, technically, they're in groups of 5 bases) but also the quality of that base call. This is important as we want to remove bad quality sequences from our dataset. That's QC. 22/n

OK, nearly finished (mostly because we've had an influx of samples at Taxa and I need to deal with the chaos). Once we have a nicely QC'd dataset of ACTGs we need to assemble them into what we call a consensus sequence (i.e. the sequence of our sample). 23/n

Because it's COVID and because no-one has time to reinvent the wheel we use the @NetworkArtic protocol using their field bioinformatics environment. That simplifies the job significantly. Takes a couple of minutes to do multiple samples. 24/n

And to finish, we assign lineage to the consensus sequence with pangolin (very aptly named...). This takes about 30 seconds and we have a .csv lineage file that tells us our control belongs to lineage B.1 ■ 25/26

I hope the constant stream of tweets hasn't bored you all silly. There's so much more we can do with viral genomes (cluster patients, show transmission chains, describe new viruses etc) but 160 characters and multiple tweets isn't quite the medium for that I think!