Twitter Thread by Ersa Flavinkins

Ersa Flavinkins
@flavinkins

Here is why you shouldn't trust anything that was sampled after the outbreak: https://t.co/VoC1cwmfdz all samples were collected at June 2020. they mentioned an "oligo array" which is how they get to all virus, no bacteria. they claim "pangolins" but there were only bat.

What they actually did: the first part of SRR13380247 were virus at the 1st read and nothing at the 2nd read, despite the quality score say otherwise. what this tells you is that they have used a blank bat sample and artificially added in reads of fake viruses.

all samples lacked non-coronavirus reads. one have traces of human Flu virus and other have trace (6Kbp) "picornavirales". both were likely probes that have fell off. the samples were very late, so the most likely route: High throughput gene synthesis->mix into blank sample

(they forgot to add the bacteria, again!) and then mock "sequencing". also, since this is already when SARS-CoV-2 have been widely circulating in THailand, there exist a change that this backbone may have spilled into bats and have recombined with the circulating strains.

If you have ever studied bat viromes, you known that a real bat sample would have more than just one viral families. ajority of bats have picoRNAviruses and other respiratory viruses, yet none of these samples gave any other bat viruses other than their single "Coronavirus"

sequences despite the use of an "respiratory oligonucleotide panel" that target all viral families. Coupled with a late collection date, any such "new data" should be completely ignored. they can just take some PCR amplicons (majority of reads were RdRp and N) and dip a blank

swab in it. samples were collected by Chinese authors. coupled with the lack of other circulating bat viruses, these samples can not be said to be legitimate in any way.

The direction of reads were scrambled, indicating that the source molecule is not RNA. the first part of the fastq have virus in the first read and nothing in the second read, despite being the same spot. this indicate that the DNA are likely synthetic.

Notice that this "respiratory virus oligo panel" is designed to perform enrichment AFTER tagmentation, using the same "enhanced PCR mix" and protocol as the TruSeq stranded mRNA preparation protocol for cDNA preparation and
https://t.co/1zB9CZILDV
https://t.co/VmT2T1J4jY

strand separation protocol, wich includes Uracil-containing DNA and the same UDP0XXX/Enhanced PCR Mix system as the TruSeq stranded mRNA kit. this system preserves strand information and any real Coronavirus reads should give a significant bias on the strand information itself.

However, these libraries were in stead to be found to have the correct strand information for anything non-Coronavirus and scrambled strand information for all Coronavirus reads, which implies that these "coronavirus reads" were added into an otherwise blank sample as dsDNA.

Consider that Twist Biosciences is already capable of performing High-throughput gene synthesis, the fact that the rRNA and bat mRNA had the correct strand information yet all Coronavirus strand information were scrambled, imply that these "sequences" were synthetic.

What they did is that they used CoV2, MP789, GX-P2V and ZC45 to creat a few imaginary "Coronavirus" sequences, ordered from Twist Biosciences for HTGS, then picked up a few blank swabs without any other virus, and poured these synthetic DNA sequences into them.

As the result, Only the Coronavirus sequences display a scrambled strand direction information, while everything else have the correct strand information expected from rRNA/mRNA mixture of a blank tissue sample.

and a few dsDNA molecules. the only sequences that don't show a strand direction bias in these libraries were those "coronavirus" sequences. ILLUMINA cDNA synthesis creates a negative first strand (dTTP) and a positive second strand (dUTP), which is quenched during later PCR

amplification and result in a conserved strand direction of F2R1. this F2R1 direction is found to be preserved in the mRNA and rRNA component within these libraries. Coronavirus RNA are strand biased, especially in rectal swabs since they have an origin in intact virions.

However, the Coronavirus reads within these libraries have a scrambled strand information of a nearly perfect 50:50 mixture of F2R1 and F1R2 read direction, indicating that the "Coronavirus" within these samples are NOT RNA molecules.

These samples are therefore easy fakes. it also show how dangerous modern gene synthesis have went.

These samples were collected AFTER the outbreak. this would allow two thing: 1: SARS-CoV-2 may have entered the local bat population and created recombinants with the otherwise normal bat CoVs there. 2: The CCP already had all the sequences they needed to concoct a fake sequence,

Synthesized them as double stranded DNA (which is how all the Coronavirus reads within the SRA and only the Coronavirus reads within the SRA becomes read-direction scrambled, but not any other things like bat reads which came from an original

blank RNA sample)

and then perform mock sequencing of these "samples". another interesting fact is that these MiSeq reads were all 100bp reads, not the 150bp ones used in SARS-CoV-2 sequencing of either Miseq or Hiseq runs. shorter runs are easier to conduct and don't reveal as many bad actors

as longer runs, and of course realistic clones of mRNA can also be synthesized usign the same system and spiked in, for added realism that forgot to preserve the strand direction information.

@threadreaderapp Unroll