(b) (5)

Please let me know right away what can be learned about this.

Francis

From: Collins, Francis (NIH/OD) [E]
Sent: Friday, June 18, 2021 10:01 PM
To: Bloom PhD, Jesse D <jbloom@fredhutch.org>; Sherry, Steve (NIH/NLM/NCBI) [E] (b) (6); Fauci, Anthony (NIH/NIAID) [E] (b) (6)
Subject: RE: SARS-CoV-2 data deleted from the NIH/NCBI SRA

Dear Jesse,

This is truly intriguing. I'll be interested in Steve's thoughts about the deleted SRA entries and whether there is any way to recover information about how that happened.

Francis

From: Bloom PhD, Jesse D <jbloom@fredhutch.org>
Sent: Friday, June 18, 2021 7:00 PM
To: Collins, Francis (NIH/OD) [E] (b) (6); Sherry, Steve (NIH/NLM/NCBI) [E] (b) (6); Fauci, Anthony (NIH/NIAID) [E] (b) (6)
Subject: SARS-CoV-2 data deleted from the NIH/NCBI SRA

Hi Francis, Stephen, and Toni,

I'm just writing to give you a heads up that I identified a data set of early Wuhan SARS-CoV-2 sequences that has been deleted from the NIH's Sequence Read Archive (SRA). I was able to recover the deleted files from the Google Cloud and analyze the sequences, and have attached a pre-print on the analysis that I just submitted for posting by *bioRxiv*.

Since SARS-CoV-2 origins and early spread has become a hot-button topic, I wanted to give you a heads up. I made sure to emphasize in the discussion that the SRA has many sequences and so isn't in a position to vet all deletions. Nonetheless, I think it would be highly worthwhile to do a comprehensive analysis of SRA (and other NIH) data that might be relevant to this topic that could have been deleted or otherwise overlooked. If I can be of any assistance, let me know.

I have been running a pipeline to identify additional deleted SRA data using various heuristics including those in described in the attached pre-print, but have not yet completed the analysis enough to know the extent that the data I have recovered is relevant to SARS-CoV-2's origins or early spread. But as I mention in the pre-print, there are two known SRR deletions that are worth looking at. I definitely think it would be good to perform a SRA side search as well, since that will obviously be easier and more efficient, and could identify deleted data not on the cloud.