Congress of the United States
Washington, DC 20515

July 9, 2021

Francis S. Collins, M.D., Ph.D.
Director, National Institutes of Health
9000 Rockville Pike
Bethesda, Maryland 20892

Dear Dr. Collins,

I am writing to request information on what the National Institutes of Health (NIH) is doing in response to reports that some sequences of SARS-CoV-2 in Wuhan were deleted from the NIH's Sequence Read Archive.

In March 2020, scientists at Wuhan University uploaded 241 genetic sequences to the Sequence Read Archive, which is managed by the National Library of Medicine within the NIH. However, those scientists requested for some of these sequences to be withdrawn from the database in June 2020, and that request was approved. The reason for removal given to the archive managers was that the data was being updated on another database. Dr. Jesse Bloom, a virologist at the Fred Hutchinson Cancer Research Center, says he recovered 13 missing sequences from Google Cloud files, which may provide insight into the origin of COVID-19 and the possibility that the virus existed prior to December 2019 in Wuhan, China.[1] Given that the true origin of SARS-CoV-2 remains a mystery, the NIH must fully account for the decision to delete information that could bring the world closer to a conclusion.

In light of these issues, please provide responses to the following:

1. Is the NIH conducting a review of the report that COVID-19 sequences were requested to be deleted from the Sequence Read Archive and why that request may have been granted? In addition, who at the NIH granted the request, and onto what other database was the deleted data uploaded?
2. Can you rule out the possibility that the deletion of these COVID-19 sequences was done with corrupt intent by the Chinese government or by scientists facing government pressure in China?
3. What safeguards does the NIH have in place to protect this and other scientific data it hosts and controls from hostile foreign actors or other nefarious activities?
4. Are you evaluating archiving measures to ensure that such data removed from similar databases, justifiably or otherwise, is preserved in the future?
5. What plans does the NIH have to investigate the entirety of the SARS-CoV-2 data sequence to ensure that we have the complete dataset?

The comprehensive, scientific understanding of SARS-CoV-2 will allow us to properly respond to this current pandemic and prevent future pandemics. I request a response to my above questions by July 14, 2021.

Sincerely,

Raja Krishnamoorthi
Member of Congress

Mark Green
Member of Congress

[1] Scientist Finds Early Virus Sequences That Had Been Mysteriously Deleted. New York Times. June 23, 2021