

From: Tabak, Lawrence (NIH/OD) [E]
Sent: Wed, 8 Sep 2021 14:44:08 -0500
To: Hallett, Adrienne (NIH/OD) [E]; Allen-Gifford, Patrice (NIH/OD) [E]
Cc: Cramer, Lindsay (NIH/OD) [E]
Subject: Re: Signature Approval for SRA letters

please proceed.
thanks

From: "Hallett, Adrienne (NIH/OD) [E]" (b) (6)
Date: Wednesday, September 8, 2021 at 3:40 PM
To: "Tabak, Lawrence (NIH/OD) [E]" (b) (6), "Allen-Gifford, Patrice (NIH/OD) [E]" (b) (6)
Cc: "Cramer, Lindsay (NIH/OD) [E]" (b) (6)
Subject: Signature Approval for SRA letters

Hey LAT,

Exec Sec needs your formal approval to sign the SRA response letters. I've attached the back and forth with FC last week to jog your memory if need be.

Thanks!
Adrienne

From: Downs, Austin (NIH/OD) [C]
Sent: Mon, 28 Jun 2021 14:24:32 -0500
To: McMahon, Christine (NIH/OD) [E]
Subject: RE: Incoming Congressional from Blackburn/Grassley/Marshall (WF 400559)

Per guidance from Higgins, please assign the letter to OLPA with necessary action/input from NLM.

Please let me know if there is anything else I can help you with on this matter.

Thank you kindly,
Austin

From: McMahon, Christine (NIH/OD) [E] (b) (6)
Sent: Monday, June 28, 2021 2:08 PM
To: Downs, Austin (NIH/OD) [C] (b) (6) >
Subject: RE: Incoming Congressional from Blackburn/Grassley/Marshall (WF 400559)

Great, thanks very much for checking!

Best,
Christine

From: Downs, Austin (NIH/OD) [C] (b) (6)
Sent: Monday, June 28, 2021 2:08 PM
To: McMahon, Christine (NIH/OD) [E] (b) (6)
Subject: RE: Incoming Congressional from Blackburn/Grassley/Marshall (WF 400559)

Good afternoon Christine, let me check with Lauren Higgins and get back to you with how we would like to proceed.

Thank you,
Austin

From: McMahon, Christine (NIH/OD) [E] (b) (6)
Sent: Monday, June 28, 2021 1:44 PM
To: Downs, Austin (NIH/OD) [C] (b) (6)
Subject: Incoming Congressional from Blackburn/Grassley/Marshall (WF 400559)

Hi Austin,

I wanted to check with you on how ES should assign the attached incoming Congressional letter. Is OLPA taking the lead, or should NLM be the lead?

Thank you!
Christine

Christine McMahon
Executive Secretariat
National Institutes of Health

(b) (6)

United States Senate
WASHINGTON, DC 20510

September 16, 2021

VIA ELECTRONIC TRANSMISSION

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

Dear Dr. Collins:

On June 28, 2021, we wrote to you requesting answers to seven questions pertaining to the NIH's role and responsibility with respect to the Sequence Read Archive (SRA) relating to COVID-19 data. On September 8, 2021, your office provided a response that failed to fully and completely answer all seven questions and failed to provide the requested records.

As we have made clear to you, Congress has a constitutional responsibility to engage in oversight of the executive branch and the executive branch has an obligation to Congress and the American people to substantively respond. In light of our responsibility and your obligation in that regard, we are reposing the unanswered questions from our June 28, 2021, letter. If you are unable to respond to each question and provide the requested records, please explain why that is the case.

In addition, we'd like to note that in your most recent letter, you stated that the National Center for Biotechnology Information has initiated "an independent review of the [Sequence Read Archive] processes and standard operating procedures" with respect to the withdrawal request relating to the COVID-19 data.¹ You also noted that NIH "conducted an analysis of withdrawal requests from January 2020 through June 2021."² With that in mind, we request the final report of investigation, or an equivalent document, for the latter review and an explanation for why that review was done. Further, we request that you answer several additional questions, provide the requested records and schedule a briefing to address your answers no later than September 30, 2021:

¹ Letter from Lawrence A. Tabak, Principal Deputy Director, NIH, to Sens. Grassley, Blackburn, Marshall (Sept. 10, 2021).

² *Id.*

1. With respect to deleting data from the NIH Sequence Archive, please name all personnel that have the authority to do so. In your answer, please provide the names and titles of the personnel that were involved in the deletion of SARS-CoV-2 data.
2. With respect to the *Wall Street Journal* report, which Chinese researcher(s) requested that the data be deleted from the NIH Sequence Read Archive?³ When did the deletion occur?
3. After deletion, does the NIH Sequence Read Archive maintain any accessible back-up of the deleted data? If so, please provide all records to us.
4. Please list all collaborating partners to the NIH Sequence Read Archive.
5. In the past five years, how many researchers and other personnel associated with the communist Chinese government, as well as those not associated with the Chinese regime, have requested that data be deleted from the NIH Sequence Read Archive? Please list by requestor, date, reason, and the information to be deleted. Please also note whether and when that material was in fact deleted.
6. More specifically, in the past five years, how many researchers and other personnel associated with the communist Chinese government, as well as those not associated with the Chinese regime, have requested that data be deleted from the NIH Sequence Read Archive relating to coronaviruses? Please list by requestor, date, reason, and the information to be deleted. Please also note whether and when that material was in fact deleted.
7. In your letter, you stated NCBI has initiated “an independent review of the SRA processes and standard operating procedures” with respect to the withdrawal request relating to the COVID-19 data.
 - a. When was that review initiated?
 - b. Which unit and personnel will be involved in the review?
 - c. What are the other countries that have requested withdrawal of data from the database?
 - d. Do you plan to analyze withdrawal requests that occurred prior to the COVID-19 outbreak? If not, why not?

³ Amy Dockser Marcus, Betsy McKay, Drew Hinshaw, *Chinese Covid-19 Gene Data That Could Have Aided Pandemic Research Removed from NIH Database*, WALL ST. J. (June 23, 2021), <https://www.wsj.com/articles/covid-19-gene-data-that-could-have-aided-research-on-early-epidemic-removed-from-database-11624472105>; Amy Dockser Marcus and Drew Hinshaw, *After Covid-19 Data Is Deleted, NIH Reviews How its Gene Archive Is Handled*, WALL ST. J. (Sept. 13, 2021), <https://www.wsj.com/articles/after-covid-19-data-is-deleted-nih-reviews-how-its-gene-archive-is-handled-11631545490>

Thank you for your attention to this important matter.

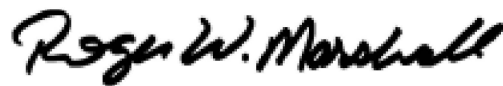
Sincerely,



Charles E. Grassley
Ranking Member
Committee on the Judiciary



Marsha Blackburn
U.S. Senator



Roger Marshall, M.D.
U.S. Senator

United States Senate

WASHINGTON, DC 20510

June 28, 2021

VIA ELECTRONIC TRANSMISSION

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

Dear Dr. Collins:

On June 23, 2021, the *Wall Street Journal* reported that Chinese researchers “directed” the National Institutes of Health (NIH) “to delete gene sequences of early COVID-19 cases from a key scientific database,” called the NIH Sequence Read Archive.¹ The article states that NIH confirmed that it deleted the sequences.² The article further reports that the deleted data includes genomic sequences from SARS-CoV-2 and that these sequences were from viral samples collected in Wuhan “in January and February 2020” from patients in the hospital.³

This type of data may contain important and relevant information that could help to better determine the virus’s origins. The efforts by Chinese researchers to delete the data demands additional explanation. As you are aware, the Chinese government has failed, from the beginning, to be open and transparent with the world with respect to its role in the pandemic.

The COVID-19 pandemic has resulted in more than 600,000 deaths, and Congress has spent trillions of dollars to support the American people, businesses, and the economy during these difficult times. Simply put, the American people deserve to know what their government knows about the origins of this global illness. As part of our continuing oversight with respect to NIH’s role during the COVID-19 pandemic, we request additional information about the NIH Sequence Read Archive and the actions taken by Chinese researchers to have NIH delete SARS-CoV-2 related data. Accordingly, please answer the following no later than July 12, 2021:

1. Please describe, in detail, how and under what circumstances data can be provided to the NIH Sequence Read Archive and how and under what circumstances data can be deleted from the same.

¹ https://www.wsj.com/articles/covid-19-gene-data-that-could-have-aided-research-on-early-epidemic-removed-from-database-11624472105?mod=searchresults_pos1&page=1

² *Id.*

³ *Id.*

2. With respect to deleting data from the NIH Sequence Archive, please name all personnel that have the authority to do so. In your answer, please provide the names and titles of the personnel that were involved in the deletion of SARS-CoV-2 data.
3. With respect to the *Wall Street Journal* report, which Chinese researcher(s) requested that the data be deleted from the NIH Sequence Read Archive? When was the request made and when did the deletion occur?
4. After deletion, does the NIH Sequence Read Archive maintain any accessible back-up of the deleted data? If so, please provide all records to us.
5. Please list all collaborating partners to the NIH Sequence Read Archive.
6. In the past five years, how many researchers and other personnel associated with the communist Chinese government have requested that data be deleted from the NIH Sequence Read Archive? Please list by requestor, date, reason, and the information to be deleted. Please also note whether and when that material was in fact deleted.
7. More specifically, in the past five years, how many researchers and other personnel associated with the communist Chinese government have requested that data be deleted from the NIH Sequence Read Archive relating to coronaviruses? Please list by requestor, date, reason, and the information to be deleted. Please also note whether and when that material was in fact deleted.

Thank you for your attention to this important matter.

Sincerely,



Marsha Blackburn
U.S. Senator



Charles E. Grassley
U.S. Senator



Roger Marshall
U.S. Senator



September 8, 2021

The Honorable Marsha Blackburn
United States Senate
Washington, DC 20510

Dear Senator Blackburn:

Thank you for your June 28, 2021, letter regarding the National Library of Medicine's (NLM) Sequence Read Archive (SRA).

Early in the pandemic, the National Institutes of Health (NIH) and other federal agencies moved quickly to make COVID-19 open-access data and computational resources freely available to researchers. NIH's National Library of Medicine (NLM) has a broad portfolio of open-access databases, including the Sequence Read Archive (SRA), the world's largest publicly available repository of high-throughput sequencing data. In the past year, SRA received approximately 2.4 million submissions of sequence data.

SRA is managed by NLM's National Center for Biotechnology Information (NCBI), which is the U.S. participating member of the International Nucleotide Sequence Database Collaboration (INSDC) since 1987. NCBI follows the INSDC [policies](#)¹ and [guidelines for data submission and change requests](#),² and collaborates with participating organizations in updating policies and guidelines as described in this 2018 [article](#).³

The guidelines describe the criteria for which submitting researchers can request a change in data status (for example, if the data have been corrupted) and actions taken if the criteria are met. Submitting institutions must contact NLM/NCBI to request that data be withdrawn. NLM/NCBI staff review the request against the INSDC guidelines.

In March 2020, the SARS-CoV-2 sequences you mention in your letter were submitted by a researcher at Wuhan University for public release status via SRA. The researcher from Wuhan University published relevant information about these sequences [by preprint in March 2020](#)⁴ and in an international scientific [journal in June 2020](#).⁵ In June 2020, NCBI received a request to withdraw the sequences from the same researcher. The reason given by the researcher was they were depositing updated data in a different database and they wanted to prevent version confusion. In response, NCBI withdrew the sequences from the SRA but they remained

¹ <https://www.insdc.org/policy.html>

² <https://www.insdc.org/documents/insdc-status-document>

³ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753279/>

⁴ <https://www.medrxiv.org/content/10.1101/2020.03.04.20029538v1>

⁵ <https://onlinelibrary.wiley.com/doi/full/10.1002/sml.202002169>

available to the world's scientists and researchers through the publication and on other platforms. No conditions were made on that request and no conditions were granted.

NIH subsequently conducted an analysis of withdrawal requests from January 2020 through June 2021. Withdrawal requests were generally submitted for scientific reasons. Examples include problems with the quality of the data, data were not correctly attributed to their owner, or data were not correctly identified. In that time, six institutions requested withdrawal of SARS-CoV-2 submission packages through NLM/NCBI services. This included the one requested by the researcher at Wuhan University and the rest from researchers at institutions from other countries, predominantly the U.S. In addition, NCBI received withdrawal requests from five institutions through INSDC collaborators for sequence data that were in the collaborating repositories and were replicated within the SRA.

While NIH considers the policies and guidelines of the INSDC sound, NCBI has initiated an independent review of SRA processes and standard operating procedures to determine whether the appropriate steps were taken to assess this withdrawal request. Withdrawal makes the data undiscoverable but does not erase it. Per the INSDC guidelines, NCBI retains withdrawn data for the scientific record and for disaster recovery. Pending outcome of the review, NCBI will work with INSDC to assign the data to the appropriate status.

Thank you for your interest in NIH's National Library of Medicine's broad portfolio of open-access databases. I hope this information is helpful to you. I have also sent an identical letter to the co-signers of your letter.

Sincerely,

(b) (6)

Lawrence A. Tabak, D.D.S., Ph.D.
Principal Deputy Director



September 8, 2021

The Honorable Charles E. Grassley
United States Senate
Washington, DC 20510

Dear Senator Grassley:

Thank you for your June 28, 2021, letter regarding the National Library of Medicine's (NLM) Sequence Read Archive (SRA).

Early in the pandemic, the National Institutes of Health (NIH) and other federal agencies moved quickly to make COVID-19 open-access data and computational resources freely available to researchers. NIH's National Library of Medicine (NLM) has a broad portfolio of open-access databases, including the Sequence Read Archive (SRA), the world's largest publicly available repository of high-throughput sequencing data. In the past year, SRA received approximately 2.4 million submissions of sequence data.

SRA is managed by NLM's National Center for Biotechnology Information (NCBI), which is the U.S. participating member of the International Nucleotide Sequence Database Collaboration (INSDC) since 1987. NCBI follows the INSDC [policies](#)¹ and [guidelines for data submission and change requests](#),² and collaborates with participating organizations in updating policies and guidelines as described in this 2018 [article](#).³

The guidelines describe the criteria for which submitting researchers can request a change in data status (for example, if the data have been corrupted) and actions taken if the criteria are met. Submitting institutions must contact NLM/NCBI to request that data be withdrawn. NLM/NCBI staff review the request against the INSDC guidelines.

In March 2020, the SARS-CoV-2 sequences you mention in your letter were submitted by a researcher at Wuhan University for public release status via SRA. The researcher from Wuhan University published relevant information about these sequences [by preprint in March 2020](#)⁴ and in an international scientific [journal in June 2020](#).⁵ In June 2020, NCBI received a request to withdraw the sequences from the same researcher. The reason given by the researcher was they were depositing updated data in a different database and they wanted to prevent version confusion. In response, NCBI withdrew the sequences from the SRA but they remained

¹ <https://www.insdc.org/policy.html>

² <https://www.insdc.org/documents/insdc-status-document>

³ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753279/>

⁴ <https://www.medrxiv.org/content/10.1101/2020.03.04.20029538v1>

⁵ <https://onlinelibrary.wiley.com/doi/full/10.1002/sml.202002169>

available to the world's scientists and researchers through the publication and on other platforms. No conditions were made on that request and no conditions were granted.

NIH subsequently conducted an analysis of withdrawal requests from January 2020 through June 2021. Withdrawal requests were generally submitted for scientific reasons. Examples include problems with the quality of the data, data were not correctly attributed to their owner, or data were not correctly identified. In that time, six institutions requested withdrawal of SARS-CoV-2 submission packages through NLM/NCBI services. This included the one requested by the researcher at Wuhan University and the rest from researchers at institutions from other countries, predominantly the U.S. In addition, NCBI received withdrawal requests from five institutions through INSDC collaborators for sequence data that were in the collaborating repositories and were replicated within the SRA.

While NIH considers the policies and guidelines of the INSDC sound, NCBI has initiated an independent review of SRA processes and standard operating procedures to determine whether the appropriate steps were taken to assess this withdrawal request. Withdrawal makes the data undiscoverable but does not erase it. Per the INSDC guidelines, NCBI retains withdrawn data for the scientific record and for disaster recovery. Pending outcome of the review, NCBI will work with INSDC to assign the data to the appropriate status.

Thank you for your interest in NIH's National Library of Medicine's broad portfolio of open-access databases. I hope this information is helpful to you. I have also sent an identical letter to the co-signers of your letter.

Sincerely,

(b) (6)

Lawrence A. Tabak, D.D.S., Ph.D.
Principal Deputy Director



September 8, 2021

The Honorable Roger Marshall
United States Senate
Washington, DC 20510

Dear Senator Marshall:

Thank you for your June 28, 2021, letter regarding the National Library of Medicine's (NLM) Sequence Read Archive (SRA).

Early in the pandemic, the National Institutes of Health (NIH) and other federal agencies moved quickly to make COVID-19 open-access data and computational resources freely available to researchers. NIH's National Library of Medicine (NLM) has a broad portfolio of open-access databases, including the Sequence Read Archive (SRA), the world's largest publicly available repository of high-throughput sequencing data. In the past year, SRA received approximately 2.4 million submissions of sequence data.

SRA is managed by NLM's National Center for Biotechnology Information (NCBI), which is the U.S. participating member of the International Nucleotide Sequence Database Collaboration (INSDC) since 1987. NCBI follows the INSDC [policies](#)¹ and [guidelines for data submission and change requests](#),² and collaborates with participating organizations in updating policies and guidelines as described in this 2018 [article](#).³

The guidelines describe the criteria for which submitting researchers can request a change in data status (for example, if the data have been corrupted) and actions taken if the criteria are met. Submitting institutions must contact NLM/NCBI to request that data be withdrawn. NLM/NCBI staff review the request against the INSDC guidelines.

In March 2020, the SARS-CoV-2 sequences you mention in your letter were submitted by a researcher at Wuhan University for public release status via SRA. The researcher from Wuhan University published relevant information about these sequences [by preprint in March 2020](#)⁴ and in an international scientific [journal in June 2020](#).⁵ In June 2020, NCBI received a request to withdraw the sequences from the same researcher. The reason given by the researcher was they were depositing updated data in a different database and they wanted to prevent version confusion. In response, NCBI withdrew the sequences from the SRA but they remained

¹ <https://www.insdc.org/policy.html>

² <https://www.insdc.org/documents/insdc-status-document>

³ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753279/>

⁴ <https://www.medrxiv.org/content/10.1101/2020.03.04.20029538v1>

⁵ <https://onlinelibrary.wiley.com/doi/full/10.1002/sml.202002169>

available to the world's scientists and researchers through the publication and on other platforms. No conditions were made on that request and no conditions were granted.

NIH subsequently conducted an analysis of withdrawal requests from January 2020 through June 2021. Withdrawal requests were generally submitted for scientific reasons. Examples include problems with the quality of the data, data were not correctly attributed to their owner, or data were not correctly identified. In that time, six institutions requested withdrawal of SARS-CoV-2 submission packages through NLM/NCBI services. This included the one requested by the researcher at Wuhan University and the rest from researchers at institutions from other countries, predominantly the U.S. In addition, NCBI received withdrawal requests from five institutions through INSDC collaborators for sequence data that were in the collaborating repositories and were replicated within the SRA.

While NIH considers the policies and guidelines of the INSDC sound, NCBI has initiated an independent review of SRA processes and standard operating procedures to determine whether the appropriate steps were taken to assess this withdrawal request. Withdrawal makes the data undiscoverable but does not erase it. Per the INSDC guidelines, NCBI retains withdrawn data for the scientific record and for disaster recovery. Pending outcome of the review, NCBI will work with INSDC to assign the data to the appropriate status.

Thank you for your interest in NIH's National Library of Medicine's broad portfolio of open-access databases. I hope this information is helpful to you. I have also sent an identical letter to the co-signers of your letter.

Sincerely,

(b) (6)

Lawrence A. Tabak, D.D.S., Ph.D.
Principal Deputy Director

June 28, 2021

VIA ELECTRONIC TRANSMISSION

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

Dear Dr. Collins:

On June 23, 2021, the *Wall Street Journal* reported that Chinese researchers “directed” the National Institutes of Health (NIH) “to delete gene sequences of early COVID-19 cases from a key scientific database,” called the NIH Sequence Read Archive.¹ The article states that NIH confirmed that it deleted the sequences.² The article further reports that the deleted data includes genomic sequences from SARS-CoV-2 and that these sequences were from viral samples collected in Wuhan “in January and February 2020” from patients in the hospital.³

This type of data may contain important and relevant information that could help to better determine the virus’s origins. The efforts by Chinese researchers to delete the data demands additional explanation. As you are aware, the Chinese government has failed, from the beginning, to be open and transparent with the world with respect to its role in the pandemic.

The COVID-19 pandemic has resulted in more than 600,000 deaths, and Congress has spent trillions of dollars to support the American people, businesses, and the economy during these difficult times. Simply put, the American people deserve to know what their government knows about the origins of this global illness. As part of our continuing oversight with respect to NIH’s role during the COVID-19 pandemic, we request additional information about the NIH Sequence Read Archive and the actions taken by Chinese researchers to have NIH delete SARS-CoV-2 related data. Accordingly, please answer the following no later than July 12, 2021:

1. Please describe, in detail, how and under what circumstances data can be provided to the NIH Sequence Read Archive and how and under what circumstances data can be deleted from the same.

Response:

[REDACTED] (b) (5)

¹ https://www.wsj.com/articles/covid-19-gene-data-that-could-have-aided-research-on-early-epidemic-removed-from-database-11624472105?mod=searchresults_pos1&page=1

² Id.

³ Id.

(b) (5)

2. With respect to deleting data from the NIH Sequence Archive, please name all personnel that have the authority to do so. In your answer, please provide the names and titles of the personnel that were involved in the deletion of SARS-CoV-2 data.

Response:

(b) (5)

3. With respect to the *Wall Street Journal* report, which Chinese researcher(s) requested that the data be deleted from the NIH Sequence Read Archive? When was the request made and when did the deletion occur?

Response:

(b) (5)

4. After deletion, does the NIH Sequence Read Archive maintain any accessible back-up of the deleted data? If so, please provide all records to us.

Response:

(b) (5)

5. Please list all collaborating partners to the NIH Sequence Read Archive.

Response:

(b) (5)

6. In the past five years, how many researchers and other personnel associated with the communist Chinese government have requested that data be deleted from the NIH Sequence Read Archive? Please list by requestor, date, reason, and the information to be deleted. Please also note whether and when that material was in fact deleted.

Response:

(b) (5)

7. More specifically, in the past five years, how many researchers and other personnel associated with the communist Chinese government have requested that data be deleted from the NIH Sequence Read Archive relating to coronaviruses? Please list by requestor, date, reason, and the information to be deleted. Please also note whether and when that material was in fact deleted.

Response:

(b) (5)

Thank you for your attention to this important matter.

Sincerely,

Marsha Blackburn
U.S. Senator

Charles E. Grassley
U.S. Senator

Roger Marshall
U.S. Senator