Instructions for controlled access to TOPMed sequence data on the cloud

Tom Blackwell - University of Michigan - October 21, 2019

There are six separate components which need to be lined up in advance.

Permissions:

(1) An NIH eRA Commons login.

(2) An approved dbGaP Data Access Request ("DAR") for the TOPMed study or studies you wish to use.

(3) If someone else made the dbGaP Data Access Request, ask them to assign you as a "Designated Downloader" for the Data Access Request.

(4) An ".ngc" key file from the dbGaP "My Research Projects" page. To obtain this, follow the separate instructions for sample identifiers in item (6) below.

(5) Access to either a Google or Amazon billing account. The dbGaP access mechanism is designed to run identically on either platform, and TOPMed sequence data are stored in duplicate on both platforms.

Sample Identifiers:

(6) dbGaP access to whole genome sequence data identifies sequenced DNA samples by their NCBI Sequence Read Archive "SRR" run accession numbers. The correspondence to familiar TOPMed "NWD" sample identifiers is shown in the NCBI Sequence Read Archive "Run Selector". This is a java web interface to NCBI's internal database and it runs in a browser on the user's desktop. Separate, detailed instructions are provided in an appendix below, “Listing TOPMed sequence data available in dbGaP”. These instructions give the only way I know to obtain an ".ngc" key file. It would be difficult to script access to the Run Selector, or to access it directly from a cloud instance. The dbGaP access mechanism also allows combining TOPMed data with sequences from other studies. After locating the samples of interest, create a plain text file showing all of their SRR numbers, one SRR number per line.

Cloud Access:

(7) The fusera software is available from github.com/mitre/fusera. Detailed usage and troubleshooting instructions are in a wiki: github.com/mitre/fusera/wiki.

(8) Alternately, we provide a Docker machine image with samtools and the fusera software already installed: https://hub.docker.com/r/statgen/cram-access-tools, currently running fusera v-1.0.0. This also contains the build 38 human genome reference sequence, which is needed in order to read each .cram file.

(9) Start a computing instance on either Google or Amazon. From my desktop, this is:

```
gcloud init
gcloud compute instances create  fusera-trial-00  --zone  us-central1-c      \
    --image-project  ubuntu-os-cloud  --image-family  ubuntu-1604-lts        \
    --create-disk  size=50GB
```

If you are using Google's web interface instead of the command line, do not check "Container", do check "Allow default access" and do not check either box under "Firewall".

(10) ssh into the running compute instance, create a working directory and a subdirectory, say '/working' and '/working/genomes', and copy both the '*.ngc' key file and the list of SRR numbers from steps 4 and 6 into the working directory on the compute instance from your desktop.

(11) Install Docker-CE ("Community Edition") on the compute instance, following instructions from https://docs.docker.com/install/linux/docker-ce/ubuntu/#set-up-the-rep down through 'sudo docker run hello-world'.

(12) Copy the statgen docker image from docker hub to the compute instance and start it with special permissions:

```
sudo docker pull  statgen/cram-access-tools
sudo docker run --rm -it -v /working:/working --privileged --cap-add SYS_ADMIN   \
   --cap-add MKNOD --device /dev/fuse  statgen/cram-access-tools
```

(13) Run fusera inside the docker container as:

```
nohup fusera mount --verbose  --ngc  <ngc.key.file>  --accession  <srr.list>     \
   /working/genomes  >  /working/fusera.log  2>&1  &
```

<srr.list> may be either a comma-separated list on the command line with one or more literal SRR numbers or a path to a text file of SRR numbers. The log file collects any error messages.

(14) This creates a 'fuse' virtual file system in the Docker image, mounted at /working/genomes, which contains the sequence data in a directory structure like:

```
/working/genomes/
      SRR12345678/
          NWD113355.b38.irc.v1.cram
          NWD113355.b38.irc.v1.cram.crai
      SRR12345679/
          NWD224466.b38.irc.v1.cram
          NWD224466.b38.irc.v1.cram.crai
      SRR12345680/
          NWD335577.b38.irc.v1.cram
          NWD335577.b38.irc.v1.cram.crai
```

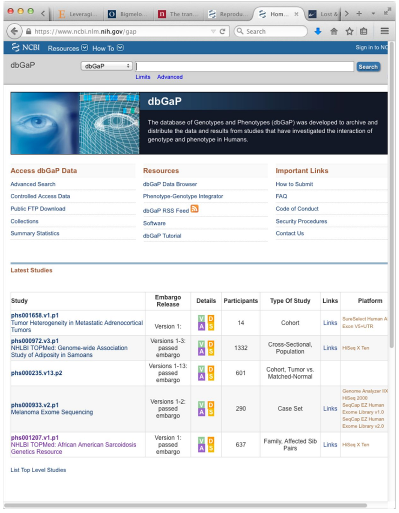
The fusera process must stay running in the background while other tools access the sequence data.

Appendix: Listing TOPMed sequence data available in dbGaP

This page provides expanded instructions to list the NCBI "SRR" accession numbers associated with TOPMed samples and to download a "*.ngc" key file which gives access to these data on the cloud. This is step (6) in the preceding "Instructions for controlled access to TOPMed sequence data on the cloud".

Start at the dbGaP main page: www.ncbi.nlm.nih.gov/gap   (panel A)
Select "Controlled Access Data" (second item in the left hand column).
This resolves to: dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login   (panel B)

The tab "Authorized Access" is already open. Select "Log In to dbGaP". (This is on the left, below the yellow banner and above "dbGaP Data Download".) It takes you to the login page with a very long address beginning:
auth.nih.gov/CertAuthV2/forms/NIHPivOrFormLogin.aspx?...   (panel C)



dbGaP

The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the data and results from studies that have investigated the interaction of genotype and phenotype in Humans.

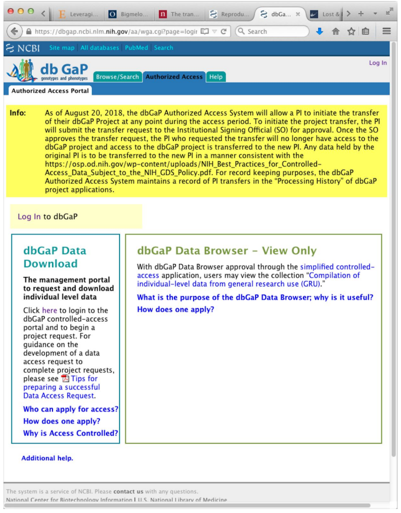
Access dbGaP Data	Resources	Important Links
Advanced Search	dbGaP Data Browser	How to Submit
Controlled Access Data	Phenotype-Genotype Integrator	FAQ
Public FTP Download	dbGaP RSS Feed	Code of Conduct
Collections	Software	Security Procedures
Summary Statistics	dbGaP Tutorial	Contact Us

Latest Studies

Study	Embargo Release	Details	Participants	Type Of Study	Links	Platform
phs001658.v1.p1 Tumor Heterogeneity in Metastatic Adrenocortical Tumors	Version 1:	V D A S	14	Cohort	Links	SureSelect Human A Exon V5+UTR
phs000972.v3.p1 NHLBI TOPMed: Genome-wide Association Study of Adiposity in Samoans	Versions 1-3: passed embargo	V D A S	1332	Cross-Sectional, Population	Links	HiSeq X Ten
phs000235.v13.p2	Versions 1-13: passed embargo	V D A S	601	Cohort, Tumor vs. Matched-Normal		
phs000933.v2.p1 Melanoma Exome Sequencing	Versions 1-2: passed embargo	V D A S	290	Case Set	Links	Genome Analyzer IIX HiSeq 2000 SeqCap EZ Human Exome Library v1.0 SeqCap EZ Human Exome Library v2.0
phs001207.v1.p1 NHLBI TOPMed: African American Sarcoidosis Genetics Resource	Version 1: passed embargo	V D A S	637	Family, Affected Sib Pairs	Links	HiSeq X Ten

List Top Level Studies

A



Info: As of August 20, 2018, the dbGaP Authorized Access System will allow a PI to initiate the transfer of their dbGaP Project at any point during the access period. To initiate the project transfer, the PI will submit the transfer request to the Institutional Signing Official (SO) for approval. Once the SO approves the transfer request, the PI who requested the transfer will no longer have access to the dbGaP project and access to the dbGaP project is transferred to the new PI. Any data held by the original PI is to be transferred to the new PI in a manner consistent with the https://osp.od.nih.gov/wp-content/uploads/NIH_Best_Practices_for_Controlled-Access_Data_Subject_to_the_NIH_GDS_Policy.pdf. For record keeping purposes, the dbGaP Authorized Access System maintains a record of PI transfers in the "Processing History" of dbGaP project applications.

Log In to dbGaP

dbGaP Data Download

The management portal to request and download individual level data

Click here to login to the dbGaP controlled-access portal and to begin a project request. For guidance on the development of a data access request to complete project requests, please see Tips for preparing a successful Data Access Request.

Who can apply for access?
How does one apply?
Why is Access Controlled?

dbGaP Data Browser – View Only

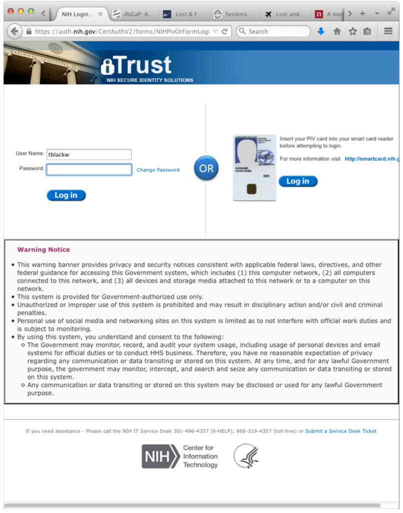
With dbGaP Data Browser approval through the simplified controlled-access application, users may view the collection "Compilation of individual-level data from general research use (GRU)."

What is the purpose of the dbGaP Data Browser; why is it useful?
How does one apply?

Additional help.

The system is a service of NCBI. Please **contact us** with any questions.
National Center for Biotechnology Information | U.S. National Library of Medicine

B


NIH Login...
dbGaP: A...
Lost & F...
Systems ...
Lost and...
A met
https://auth.nih.gov/CertAuthV2/forms/NIHPivOrFormLogi
Search
iTrust
NIH SECURE IDENTITY SOLUTIONS
Change Password
OR
Log in
Log in
NIH
Center for Information Technology

User Name: tblackw
Password:

Insert your PIV card into your smart card reader before attempting to login.
For more information visit http://smartcard.nih.g

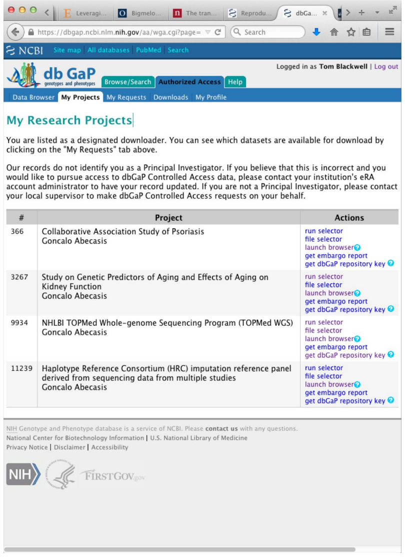
Warning Notice

- This warning banner provides privacy and security notices consistent with applicable federal laws, directives, and other federal guidance for accessing this Government system, which includes (1) this computer network, (2) all computers connected to this network, and (3) all devices and storage media attached to this network or to a computer on this network.
- This system is provided for Government-authorized use only.
- Unauthorized or improper use of this system is prohibited and may result in disciplinary action and/or civil and criminal penalties.
- Personal use of social media and networking sites on this system is limited as to not interfere with official work duties and is subject to monitoring.
- By using this system, you understand and consent to the following:
  - The Government may monitor, record, and audit your system usage, including usage of personal devices and email systems for official duties or to conduct HHS business. Therefore, you have no reasonable expectation of privacy regarding any communication or data transiting or stored on this system. At any time, and for any lawful Government purpose, the government may monitor, intercept, and search and seize any communication or data transiting or stored on this system.
  - Any communication or data transiting or stored on this system may be disclosed or used for any lawful Government purpose.

If you need assistance - Please call the NIH IT Service Desk 301-496-4357 (6-HELP); 866-319-4357 (toll-free) or Submit a Service Desk Ticket

C

Provide your eRA commons user name and password. This takes you to:
dbgap.ncbi.nlm.nih.gov/aa/wga.cgi/page=list_wishlists with the "Authorized Access" and "My Projects" tabs already open (panel D).



Logged in as **Tom Blackwell** | Log out

My Research Projects

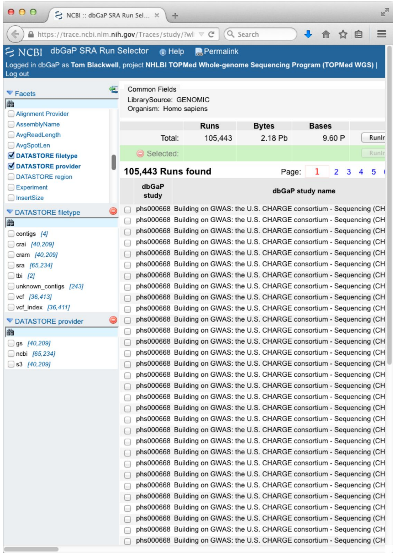
You are listed as a designated downloader. You can see which datasets are available for download by clicking on the "My Requests" tab above.

Our records do not identify you as a Principal Investigator. If you believe that this is incorrect and you would like to pursue access to dbGaP Controlled Access data, please contact your institution's eRA account administrator to have your record updated. If you are not a Principal Investigator, please contact your local supervisor to make dbGaP Controlled Access requests on your behalf.

#	Project	Actions
366	Collaborative Association Study of Psoriasis Goncalo Abecasis	run selector file selector launch browser get embargo report get dbGaP repository key
3267	Study on Genetic Predictors of Aging and Effects of Aging on Kidney Function Goncalo Abecasis	run selector file selector launch browser get embargo report get dbGaP repository key
9934	NHLBI TOPMed Whole-genome Sequencing Program (TOPMed WGS) Goncalo Abecasis	run selector file selector launch browser get embargo report get dbGaP repository key
11239	Haplotype Reference Consortium (HRC) imputation reference panel derived from sequencing data from multiple studies Goncalo Abecasis	run selector file selector launch browser get embargo report get dbGaP repository key

NIH Genotype and Phenotype database is a service of NCBI. Please **contact us** with any questions.
National Center for Biotechnology Information | U.S. National Library of Medicine
Privacy Notice | Disclaimer | Accessibility

D


NCBI :: dbGaP SRA Run Sel...
https://trace.ncbi.nlm.nih.gov/Traces/study/?wl
Search
NCBI
dbGaP SRA Run Selector
Help
Permalink
RunIn
RunIn

Logged in dbGaP as **Tom Blackwell**, project **NHLBI TOPMed Whole-genome Sequencing Program (TOPMed WGS)** | Log out

Facets
- ☐ Alignment Provider
- ☐ AssemblyName
- ☐ AvgReadLength
- ☐ AvgSpotLen
- ☑ DATASTORE filetype
- ☑ DATASTORE provider
- ☐ DATASTORE region
- ☐ Experiment
- ☐ InsertSize

DATASTORE filetype
- ☐ contigs [4]
- ☐ crai [40,209]
- ☐ cram [40,209]
- ☐ sra [65,234]
- ☐ tbi [2]
- ☐ unknown_contigs [243]
- ☐ vcf [36,413]
- ☐ vcf_index [36,411]

DATASTORE provider
- ☐ gs [40,209]
- ☐ ncbi [65,234]
- ☐ s3 [40,209]

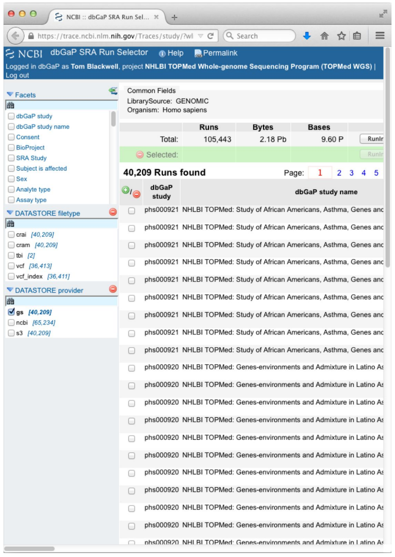
Common Fields
LibrarySource: GENOMIC
Organism: Homo sapiens

	Runs	Bytes	Bases
Total:	105,443	2.18 Pb	9.60 P
Selected:			

105,443 Runs found   Page: 1 2 3 4 5

	dbGaP study	dbGaP study name
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH
☐	phs000668	Building on GWAS: the U.S. CHARGE consortium - Sequencing (CH

E


NCBI :: dbGaP SRA Run Sel...
https://trace.ncbi.nlm.nih.gov/Traces/study/?wl
Search
NCBI
dbGaP SRA Run Selector
Help
Permalink
RunIn
RunIn

Logged in dbGaP as **Tom Blackwell**, project **NHLBI TOPMed Whole-genome Sequencing Program (TOPMed WGS)** | Log out

Facets
- ☐ dbGaP study
- ☐ dbGaP study name
- ☐ Consent
- ☐ BioProject
- ☐ SRA Study
- ☐ Subject is affected
- ☐ Sex
- ☐ Analyte type
- ☐ Assay type

DATASTORE filetype
- ☐ crai [40,209]
- ☐ cram [40,209]
- ☐ tbi [2]
- ☐ vcf [36,413]
- ☐ vcf_index [36,411]

DATASTORE provider
- ☑ gs [40,209]
- ☐ ncbi [65,234]
- ☐ s3 [40,209]

Common Fields
LibrarySource: GENOMIC
Organism: Homo sapiens

	Runs	Bytes	Bases
Total:	105,443	2.18 Pb	9.60 P
Selected:			

40,209 Runs found   Page: 1 2 3 4 5

/	dbGaP study	dbGaP study name
☐	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes anc
☐	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes anc
☐	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes anc
☐	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes anc
☐	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes anc
☐	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes anc
☐	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes anc
☐	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes anc
☐	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes anc
☐	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino As
☐	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino As
☐	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino As
☐	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino As
☐	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino As
☐	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino As
☐	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino As
☐	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino As
☐	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino As
☐	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino As
☐	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino As

F

From here on, the screenshots will inevitably show what I see and not exactly what you will see. TOPMed is the third among the four projects shown in panel D. Before going further, click "get dbGaP repository key" for the TOPMed project (last item in the right hand column for the TOPMed project). This generates and downloads to your desktop the "*.ngc" key file needed for cloud access to TOPMed sequence data. Set local permissions on this file so that only you can read it. Giving this file to anyone else will give them access to the sequence data.

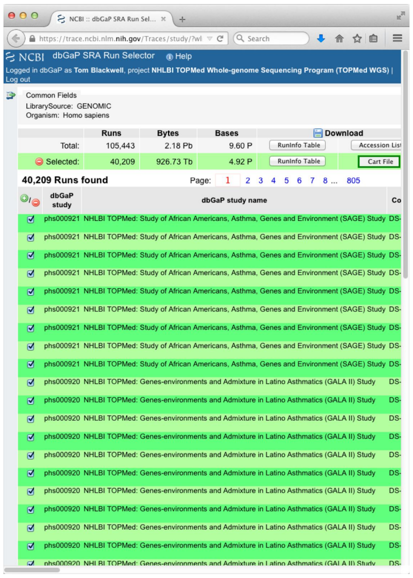
Next, select "run selector", first item in the right hand column for the TOPMed project. This takes you to the "dbGaP SRA Run Selector" trace.ncbi.nlm.nih.gov/Traces/study/?wlid=... .

This page (panel E) may take 60 seconds or more to load. Always switch to the “old” Run Selector if the new one fails to find any data. For me, as of December 2018, it returns a table of 105,440 rows and 44 columns. In order to negotiate this table, use faceted search from the left hand column of the web page. The top left panel, "Facets", with a small vertical scrollbar, shows all 44 column names. In panel E, both "DATASTORE filetype" and "DATASTORE provider" are checked. Each one opens a panel below showing all possible data values in that column and the number of records with each value.

TOPMed sequence data available on the cloud are exclusively .cram files with associated .crai index files. Files with .sra filetype are TOPMed phase 1 data mapped to build 37 plus older ESP exome sequence data from some TOPMed parent studies. These are hosted by NCBI in .sra format. .vcf and .vcf_index filetypes refer to TOPMed freeze 5b genotype data in single-sample .vcf format provided in March 2018 as an experiment for the NHLBI Data STAGE developers.

The "DATASTORE provider" tab shows that the same 40,209 records are available from both "gs" and "s3", while 65,234 (.sra) records are available from "ncbi".

Checking the box for "gs" (shown as panel F) restricts the display to just the 40,209 records with that provider. Then pressing "+" in the pair of green and red buttons left of the column headings in the main table, just below "40,209 Runs found", selects all 40,209 records, turns them green and enables you to select "RunInfo Table" above (panel G). This downloads a tab-delimited file of 40,209 rows x 44 columns to your desktop, from which you can select the studies and samples of interest. Their NWD identifiers are shown in column 27 "Sample name" and the corresponding "SRR" run accession numbers are shown in column 24 "Run" (panel H). These SRR identifiers are what you will need for cloud access to the sequence data.


NCBI :: dbGaP SRA Run Sel...
https://trace.ncbi.nlm.nih.gov/Traces/study/?wl
Search
NCBI dbGaP SRA Run Selector
Help

Logged in dbGaP as **Tom Blackwell**, project **NHLBI TOPMed Whole-genome Sequencing Program (TOPMed WGS)** | Log out

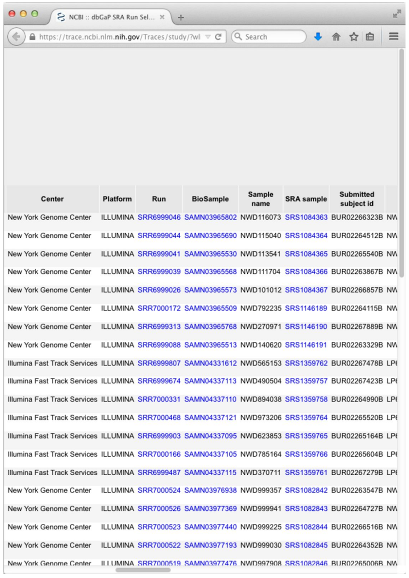
Common Fields
LibrarySource: GENOMIC
Organism: Homo sapiens

	Runs	Bytes	Bases	Download	
Total:	105,443	2.18 Pb	9.60 P	RunInfo Table	Accession List
Selected:	40,209	926.73 Tb	4.92 P	RunInfo Table	Cart File

40,209 Runs found Page: 1 2 3 4 5 6 7 8 ... 805

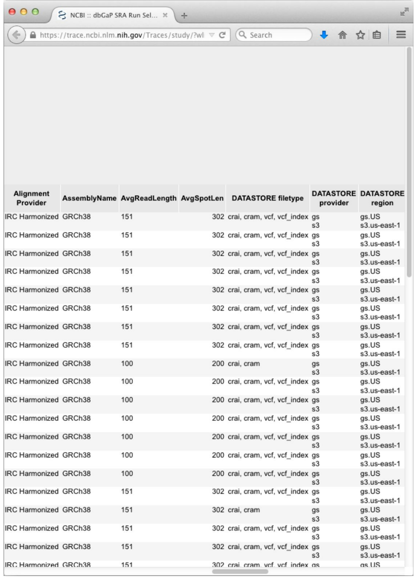
+/−	dbGaP study	dbGaP study name	Co
☑	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes and Environment (SAGE) Study	DS-
☑	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes and Environment (SAGE) Study	DS-
☑	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes and Environment (SAGE) Study	DS-
☑	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes and Environment (SAGE) Study	DS-
☑	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes and Environment (SAGE) Study	DS-
☑	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes and Environment (SAGE) Study	DS-
☑	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes and Environment (SAGE) Study	DS-
☑	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes and Environment (SAGE) Study	DS-
☑	phs000921	NHLBI TOPMed: Study of African Americans, Asthma, Genes and Environment (SAGE) Study	DS-
☑	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino Asthmatics (GALA II) Study	DS-
☑	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino Asthmatics (GALA II) Study	DS-
☑	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino Asthmatics (GALA II) Study	DS-
☑	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino Asthmatics (GALA II) Study	DS-
☑	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino Asthmatics (GALA II) Study	DS-
☑	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino Asthmatics (GALA II) Study	DS-
☑	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino Asthmatics (GALA II) Study	DS-
☑	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino Asthmatics (GALA II) Study	DS-
☑	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino Asthmatics (GALA II) Study	DS-
☑	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino Asthmatics (GALA II) Study	DS-
☑	phs000920	NHLBI TOPMed: Genes-environments and Admixture in Latino Asthmatics (GALA II) Study	DS-

G


NCBI :: dbGaP SRA Run Sel...
https://trace.ncbi.nlm.nih.gov/Traces/study/?wl
Search

Center	Platform	Run	BioSample	Sample name	SRA sample	Submitted subject id	
New York Genome Center	ILLUMINA	SRR6999046	SAMN03965802	NWD116073	SRS1084363	BUR02266323B	NW
New York Genome Center	ILLUMINA	SRR6999044	SAMN03965690	NWD115040	SRS1084364	BUR02264512B	NW
New York Genome Center	ILLUMINA	SRR6999041	SAMN03965530	NWD113541	SRS1084365	BUR02265540B	NW
New York Genome Center	ILLUMINA	SRR6999039	SAMN03965568	NWD111704	SRS1084366	BUR02263867B	NW
New York Genome Center	ILLUMINA	SRR6999026	SAMN03965573	NWD101012	SRS1084367	BUR02266857B	NW
New York Genome Center	ILLUMINA	SRR7000172	SAMN03965509	NWD792235	SRS1146189	BUR02264115B	NW
New York Genome Center	ILLUMINA	SRR6999313	SAMN03965768	NWD270971	SRS1146190	BUR02267889B	NW
New York Genome Center	ILLUMINA	SRR6999088	SAMN03965513	NWD140620	SRS1146191	BUR02263329B	NW
Illumina Fast Track Services	ILLUMINA	SRR6999807	SAMN04331612	NWD565153	SRS1359762	BUR02267478B	LP
Illumina Fast Track Services	ILLUMINA	SRR6999674	SAMN04337113	NWD490504	SRS1359757	BUR02267423B	LP
Illumina Fast Track Services	ILLUMINA	SRR7000331	SAMN04337110	NWD894038	SRS1359758	BUR02264990B	LP
Illumina Fast Track Services	ILLUMINA	SRR7000468	SAMN04337121	NWD973206	SRS1359764	BUR02265520B	LP
Illumina Fast Track Services	ILLUMINA	SRR6999903	SAMN04337095	NWD623853	SRS1359765	BUR02265164B	LP
Illumina Fast Track Services	ILLUMINA	SRR7000166	SAMN04337105	NWD785164	SRS1359766	BUR02265604B	LP
Illumina Fast Track Services	ILLUMINA	SRR6999487	SAMN04337115	NWD370711	SRS1359761	BUR02267279B	LP
New York Genome Center	ILLUMINA	SRR7000524	SAMN03976938	NWD999357	SRS1082842	BUR02263547B	NW
New York Genome Center	ILLUMINA	SRR7000526	SAMN03977369	NWD999941	SRS1082843	BUR02264727B	NW
New York Genome Center	ILLUMINA	SRR7000523	SAMN03977440	NWD999225	SRS1082844	BUR02266516B	NW
New York Genome Center	ILLUMINA	SRR7000522	SAMN03977193	NWD999030	SRS1082845	BUR02264352B	NW
New York Genome Center	ILLUMINA	SRR7000519	SAMN03977476	NWD997908	SRS1082846	BUR02265006B	NW

H


NCBI :: dbGaP SRA Run Sel...
https://trace.ncbi.nlm.nih.gov/Traces/study/?wl
Search

Alignment Provider	AssemblyName	AvgReadLength	AvgSpotLen	DATASTORE filetype	DATASTORE provider	DATASTORE region
IRC Harmonized	GRCh38	151	302	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	151	302	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	151	302	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	151	302	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	151	302	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	151	302	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	151	302	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	151	302	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	100	200	crai, cram	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	100	200	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	100	200	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	100	200	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	100	200	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	100	200	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	100	200	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	151	302	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	151	302	crai, cram	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	151	302	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	151	302	crai, cram, vcf, vcf_index	gs s3	gs.US s3.us-east-1
IRC Harmonized	GRCh38	151	302	crai, cram, vcf, vcf_index	gs	gs.US

I