

Download GEO Data (SRA Toolkit)

The Sequence Read Archive (SRA) is a publicly accessible archive for high throughput sequencing data. The [SRA Toolkit](#) from NCBI is a collection of tools for using data in the INSDC SRA. It takes the following steps to download data from SRA:

Install and Config SRA Toolkit

```
# Download and extract the latest version
$ wget --output-document sratoolkit.tar.gz https://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/2.10.9/sratoolkit.2.10.9-ubuntu64.tar.gz
$ tar -vxzf sratoolkit.2.10.9-ubuntu64.tar.gz

# Append the path to your PATH environment variable:
$ export PATH=$PATH:/path/to/sratoolkit.2.10.9-ubuntu64/bin

# Verify the installation
$ which fastq-dump
```

Download Data from SRA

1. Access the GEO summary page by searching "GSE93676" on [GEO website](#).



The screenshot shows the NCBI GEO website interface. At the top, there are logos for NCBI and GEO (Gene Expression Omnibus). Below the logos is a navigation bar with links: HOME, SEARCH, SITE MAP, GEO Publications, FAQ, MIAME, and Email GEO. The main content area is titled "NCBI > GEO > Accession Display" and includes a "Not logged in | Login" link. A "GEO help" message states: "Mouse over screen elements for information." Below this is a search bar with the following fields: Scope (Self), Format (HTML), Amount (Quick), GEO accession (GSE93676), and a GO button. The search results are displayed under the heading "Series GSE93676" with a link "Query DataSets for GSE93676". The results are organized into a table with the following rows:

Status	Public on Oct 31, 2017
Title	Viral and Cellular N6-methyladenosine (m6A) Epitranscriptomes in KSHV Life Cycle
Organisms	Homo sapiens ; Rattus norvegicus
Experiment type	Expression profiling by high throughput sequencing
Summary	Tan et al. discovered abundant conserved N6-methyladenosine (m6A) modifications on KSHV transcripts during latent and productive infection in different cell types. They also show that m6A readers YTHDF2 and YTHDF3 mediate KSHV replication, and KSHV optimizes both phases of viral replication by reprogramming cellular epitranscriptome to regulate distinct signaling pathways.
Overall design	m6A epitranscriptome profiles of four uninfected cell types, five latently infected cell types, and three lytic cell types. Each cell type has an immunoprecipitated (IP) and input sample.
Contributor(s)	Tan B , Liu H , Silva SR , Zhang L , Meng J , Cui X , Yuan H , Sorel O , Huang Y , Gao SJ

2. Find a link for "SRA" under the heading "Relations".

Submission date Jan 16, 2017
 Last update date May 15, 2019
 Contact name Hui Liu
 E-mail(s) lhcumt@hotmail.com
 Organization name China University of Mining and Technology
 Street address #1 Daxue Road
 City Xuzhou
 State/province Jiangsu
 ZIP/Postal code 221116
 Country China

Platforms (2) [GPL11154](#) Illumina HiSeq 2000 (Homo sapiens)
[GPL14844](#) Illumina HiSeq 2000 (Rattus norvegicus)

Samples (72) [GSM2460344](#) iSLK-uninf-input
[GSM2460345](#) iSLK-uninf-ip
[GSM2460346](#) iSLK-KSHV_BAC16-latent-input
[More...](#)

Relations
 BioProject [PRJNA361517](#)
 SRA [SRP096845](#)

3. Click on the link (SRP096845) which sends you to a page of all the biological samples with specific runs and files in this study.

NCBI Resources How To Sign in to NCBI

SRA SRA SRP096845 Search

Create alert Advanced Help

Access Public (72) Summary 20 per page Send to: Filters: Manage Filters

Source RNA (72) View results as an expanded interactive table using the RunSelector. [Send results to Run selector](#)

Library Layout single (72)

Platform Illumina (72)

Strategy Exome (72)

Data in Cloud GS (72) S3 (72)

File Type fastq (72)

Clear all Show additional filters

Search results
 Items: 1 to 20 of 72

1. [GSM2754257: KMM-KSHV_BAC36-latent-input-R3: Rattus norvegicus: OTHER](#)
 7 ILLUMINA (Illumina HiSeq 2000) runs: 25.5M spots, 1.3G bases, 462.5Mb downloads
 Accession: SRX3135237

2. [GSM2754256: KMM-KSHV_BAC36-latent-ip-R3: Rattus norvegicus: OTHER](#)
 7 ILLUMINA (Illumina HiSeq 2000) runs: 26.3M spots, 1.3G bases, 486.5Mb downloads
 Accession: SRX3135236

3. [GSM2754255: MM-uninf-input-R3: Rattus norvegicus: OTHER](#)
 7 ILLUMINA (Illumina HiSeq 2000) runs: 38.4M spots, 2G bases, 668.8Mb downloads
 Accession: SRX3135235

4. [GSM2754254: MM-uninf-ip-R3: Rattus norvegicus: OTHER](#)
 7 ILLUMINA (Illumina HiSeq 2000) runs: 39.5M spots, 2G bases, 694Mb downloads
 Accession: SRX3135234

Results by taxon
 Top Organisms [\[Tree\]](#)
 Homo sapiens (60)
 Rattus norvegicus (12)

Search in related databases

Database	Access		all
	public	controlled	
BioSample			
BioProject			
dbGaP			
GEO Datasets	1		1

Find related data
 Database: [Select](#)
[Find items](#)

4. To find files of interest in one comprehensive list, navigate to the bottom of the page then click: "send to" > "Run Selector" > "go". Use "Filter List" to narrow down the choices.

NCBI SRA Run Selector Log in to NIH

Filters List

- ☐ Antibody
- ☐ Bases
- ☐ Bytes
- ☐ Cell_Line
- ☐ Infection_status
- ☒ Organism
- ☐ ReleaseDate
- ☐ Treatment

Organism

- ☒ homo sapiens 364
- ☐ rattus norvegicus 76

Common Fields

BioProject	PRJNA361517
Consent	PUBLIC
Assay Type	OTHER
AvgSpotLen	51
Center Name	GEO
DATASTORE filetype	FASTQ, SRA
DATASTORE provider	GS, NCBI, S3
DATASTORE region	gs.US, ncbi.public, s3.us-east-1
Instrument	Illumina HiSeq 2000

Select

	Runs	Bytes	Bases	Download	Cloud Data Delivery	Computing
Total	440	81.77 Gb	181.31 G	Metadata or Accession List		
Selected	0	0	0	Metadata or Accession List or JWT Cart	Deliver Data	Galaxy

Found 364 items Search within results

5. Extract FastQ files from SRA-accession using SRA-Toolkit

```
#!/bin/bash
cd /path/to/raw_data/homo/
fetch_dump(){
  prefetch $1
  fastq-dump $1
}
export -f fetch_dump
for s in SRR5978827 SRR5978828 SRR5978829 SRR5978834 SRR5978835 SRR5978836
SRR5978869 SRR5978870 SRR5978871 SRR5179446 SRR5179447 SRR5179448
do
  fetch_dump ${s}
done
```

```
#!/bin/bash
cd /path/to/raw_data/mm10/
fetch_dump(){
  prefetch $1
  fastq-dump $1
}
export -f fetch_dump
for s in SRR866997 SRR866998 SRR866999 SRR867000 SRR867001 SRR867002 SRR866991
SRR866992 SRR866993 SRR866994 SRR866995 SRR866996
do
  fetch_dump ${s}
done
```

For paired-end data, you need to add `--split-files` option in the `fastq-dump` command.

```
fastq-dump --split-files SRR866997
```