

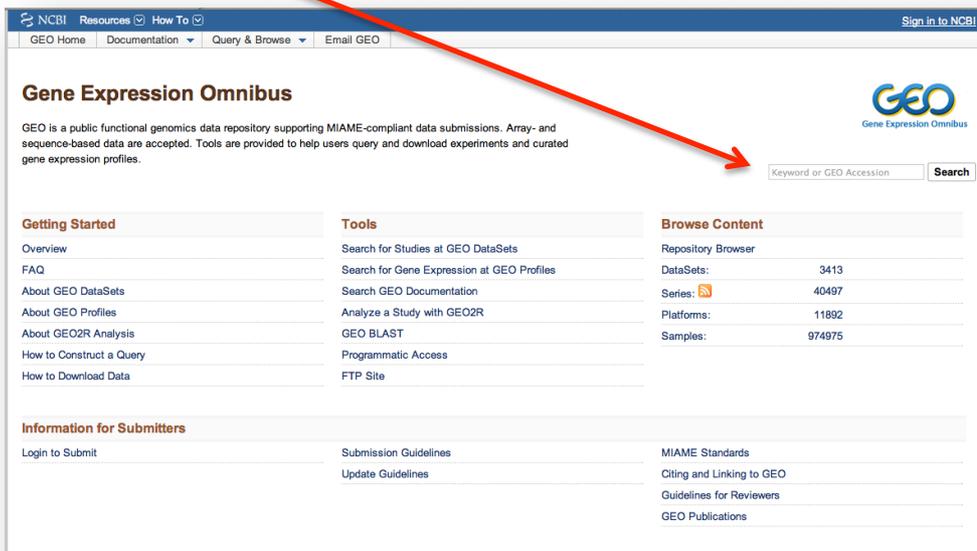
Downloading published fastq data from GEO

This guide will show you how to download fastq format data from published papers.

Look in the paper for the GEO accession number and then go to the GEO website:

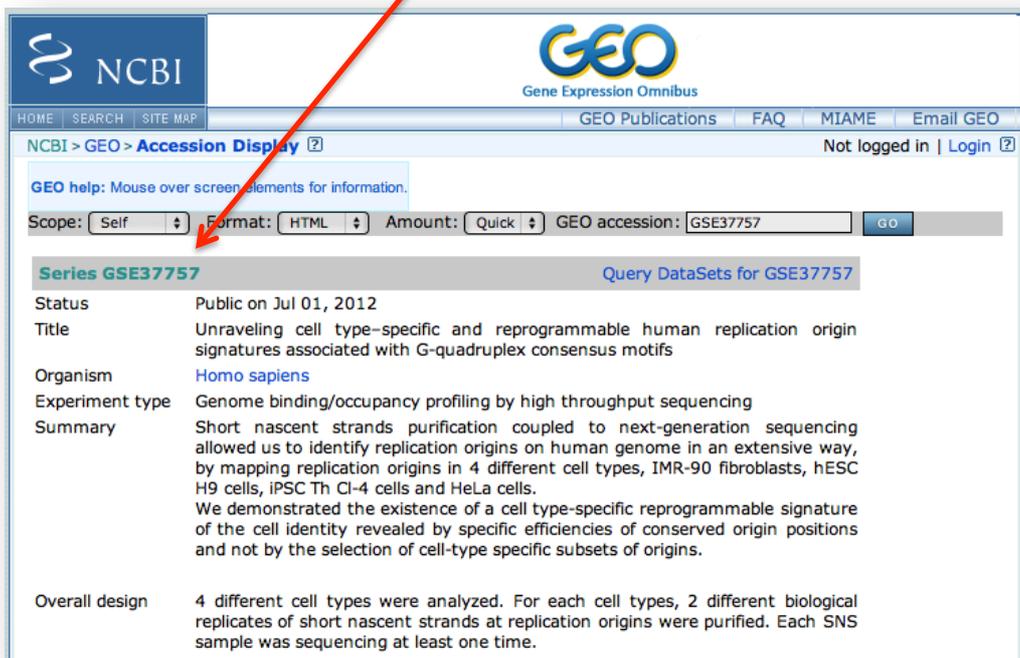
<http://www.ncbi.nlm.nih.gov/geo/>

Enter the GEO accession code for that data you want to download (e.g. **GSE37757**) and click **Search**



The screenshot shows the GEO website homepage. At the top, there is a navigation bar with "GEO Home", "Documentation", "Query & Browse", and "Email GEO". The main heading is "Gene Expression Omnibus". Below this, there is a search bar with the placeholder text "Keyword or GEO Accession" and a "Search" button. A red arrow points from the text above to this search bar. The page is divided into several sections: "Getting Started" (Overview, FAQ, About GEO DataSets, About GEO Profiles, About GEO2R Analysis, How to Construct a Query, How to Download Data), "Tools" (Search for Studies at GEO DataSets, Search for Gene Expression at GEO Profiles, Search GEO Documentation, Analyze a Study with GEO2R, GEO BLAST, Programmatic Access, FTP Site), "Browse Content" (Repository Browser, DataSets: 3413, Series: 40497, Platforms: 11892, Samples: 974975), and "Information for Submitters" (Login to Submit, Submission Guidelines, Update Guidelines, MIAME Standards, Citing and Linking to GEO, Guidelines for Reviewers, GEO Publications).

You will then see a page for that **Series**:



The screenshot shows the "Accession Display" page for Series GSE37757. At the top, there is a navigation bar with "HOME", "SEARCH", "SITE MAP", "GEO Publications", "FAQ", "MIAME", and "Email GEO". The main heading is "NCBI > GEO > Accession Display". Below this, there is a search bar with the placeholder text "GEO help: Mouse over screen elements for information." and a "GO" button. A red arrow points from the text above to the search bar. The search results show "Scope: Self", "Format: HTML", "Amount: Quick", and "GEO accession: GSE37757". The main content area displays the details for Series GSE37757, including the title "Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs", the organism "Homo sapiens", the experiment type "Genome binding/occupancy profiling by high throughput sequencing", and the summary "Short nascent strands purification coupled to next-generation sequencing allowed us to identify replication origins on human genome in an extensive way, by mapping replication origins in 4 different cell types, IMR-90 fibroblasts, hESC H9 cells, iPSC Th CI-4 cells and HeLa cells. We demonstrated the existence of a cell type-specific reprogrammable signature of the cell identity revealed by specific efficiencies of conserved origin positions and not by the selection of cell-type specific subsets of origins." The overall design section states "4 different cell types were analyzed. For each cell types, 2 different biological replicates of short nascent strands at replication origins were purified. Each SNS sample was sequencing at least one time."

Scroll down that page to find the 'Samples' section and click 'More' link if necessary to see all the samples in the entry.

NCBI > GEO > **Accession Display** [?](#) Not logged in | [Login](#) [?](#)

GEO help: Mouse over screen elements for information.

Scope: Format: Amount: GEO accession:

Series GSE37757 [Query DataSets for GSE37757](#)

Status Public on Jul 01, 2012

Title Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs

Organism [Homo sapiens](#)

Experiment type Genome binding/occupancy profiling by high throughput sequencing

Summary Short nascent strands purification coupled to next-generation sequencing allowed us to identify replication origins on human genome in an extensive way, by mapping replication origins in 4 different cell types, IMR-90 fibroblasts, hESC H9 cells, iPSC Th CI-4 cells and HeLa cells. We demonstrated the existence of a cell type-specific reprogrammable signature of the cell identity revealed by specific efficiencies of conserved origin positions and not by the selection of cell-type specific subsets of origins.

Overall design 4 different cell types were analyzed. For each cell types, 2 different biological replicates of short nascent strands at replication origins were purified. Each SNS sample was sequencing at least one time.

Contributor(s) [Besnard E](#), [Babled A](#), [Lapasset L](#), [Milhavet O](#), [Parrinello H](#), [Dantec C](#), [Marin J](#), [Lemaitre J](#)

Citation(s) Besnard E, Babled A, Lapasset L, Milhavet O et al. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol* 2012 Aug;19(8):837-44. PMID: [22751019](#)

Submission date May 04, 2012

Last update date Jul 01, 2013

Contact name Emilie Besnard

Organization name Institute of Functional Genomics

Lab Genome Plasticity and Aging

Street address 141 rue de la Cardonille

City MONTPELLIER

ZIP/Postal code 34000

Country France

Platforms (1) [GPL9115](#) Illumina Genome Analyzer II (Homo sapiens)

Samples (4) [GSM927235](#) IMR-90
[More...](#) [GSM927236](#) hESC H9
[GSM927237](#) iPSC Thomson clone 4

Relations

BioProject [PRJNA163241](#)

SRA [SRP012667](#)

Download family	Format
SOFT formatted family file(s)	SOFT ?
MINIML formatted family file(s)	MINIML ?
Series Matrix File(s)	TXT ?

Click on one the sample links eg 'GSM927238'

You will now see an entry for that sample

Scope: Format: Amount: GEO accession:

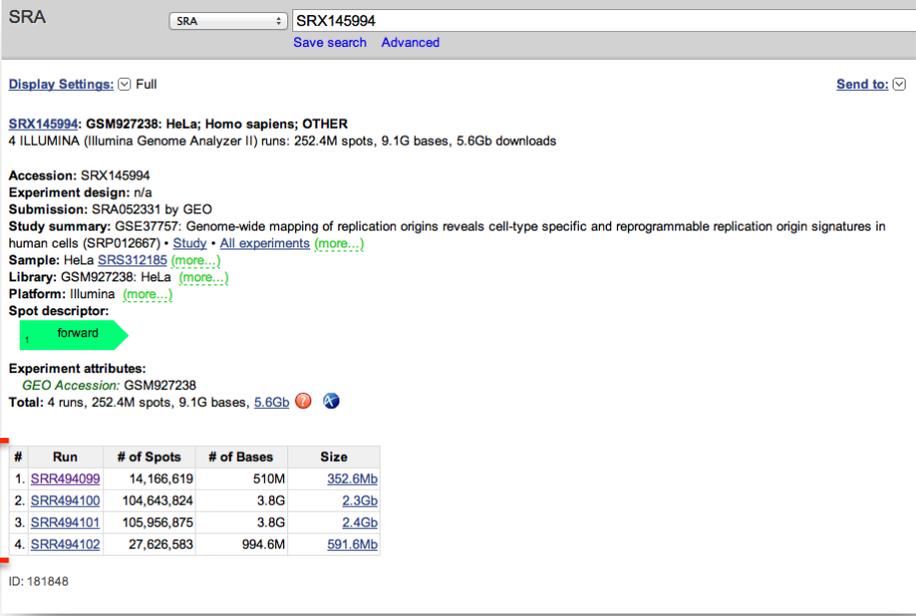
Sample GSM927238 [Query DataSets for GSM927238](#)

Status	Public on Jul 01, 2012
Title	HeLa
Sample type	SRA
Source name	cervical cancer cells
Organism	Homo sapiens
Characteristics	cell line: HeLa
Treatment protocol	No specific treatment
Growth protocol	HeLa cells were maintained in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen) containing 10 % heat-inactivated Fetal Bovine Serum (FBS, PAA), 1 % L-Glutamine (200 mM), 1 % penicillin and streptomycin (all from Invitrogen).
Extracted molecule	genomic DNA
Extraction protocol	We purified short nascent strands at replication origins in different cell types. We realized isolation of Short Nascent Strands at replication origins (SNS) of asynchronous human cells by adapting the protocol previously described (Cadoret et al. PNAS 2008). For each cell type, we purified the SNS of two biological replicates. For each sample, DNazol was used to extract DNA of 100 millions asynchronous cells. Then DNA fragments of 1-2 kb were isolated after sucrose gradient separation. After two runs of lambda exonuclease, the quality for each purified SNS preparation was tested, as previously described (Cadoret et al. PNAS 2008). Synthesis of second strands was realized with the kit bioprime DNA labeling system. In order to be reproducible, we realized, for each biological sample, this synthesis with the same amount of background DNA calculated on the basis of background around the replication origin at the Myc locus. In order to be able to compare the replication origins between different cells, we gave the same amount of background DNA to the sequencing platform.
Library strategy	OTHER
Library source	genomic
Library selection	other
Instrument model	Illumina Genome Analyzer II
Description	enrichment of short nascent strands (SNS) at replication origins with lambda exonuclease
Data processing	The next generation sequencing of purified and sonicated SNS was realized in the Montpellier GenomiX (MGX) facility in Montpellier, France using the Illumina's sequencing by synthesis technology. The sequencing library was clustered and then hybridized with sequencing primers. The sequencing of 36 bp reads by single end pairing was performed with HiSeq 2000 of Illumina.

Scroll down until you see the 'Relations' section and click on the SRA link:

Platform ID	GPL9115
Series (1)	GSE37757 Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs
Relations	
SRA	SRX145994
BioSample	SAMN00990948

Next you will see a 'SRA' page. Note the number of runs that make up this entry (in this case, 4) and note the SRR numbers for each:



SRA SRX145994
[Save search](#) [Advanced](#)

Display Settings: Full [Send to:](#)

SRX145994: GSM927238: HeLa; Homo sapiens; OTHER
4 ILLUMINA (Illumina Genome Analyzer II) runs: 252.4M spots, 9.1G bases, 5.6Gb downloads

Accession: SRX145994
Experiment design: n/a
Submission: SRA052331 by GEO
Study summary: GSE37757: Genome-wide mapping of replication origins reveals cell-type specific and reprogrammable replication origin signatures in human cells (SRP012667) • [Study](#) • [All experiments \(more...\)](#)
Sample: HeLa [SRS312185 \(more...\)](#)
Library: GSM927238: HeLa [\(more...\)](#)
Platform: Illumina [\(more...\)](#)
Spot descriptor:
forward

Experiment attributes:
GEO Accession: GSM927238
Total: 4 runs, 252.4M spots, 9.1G bases, [5.6Gb](#) 

#	Run	# of Spots	# of Bases	Size
1.	SRR494099	14,166,619	510M	352.6Mb
2.	SRR494100	104,643,824	3.8G	2.3Gb
3.	SRR494101	105,956,875	3.8G	2.4Gb
4.	SRR494102	27,626,583	994.6M	591.6Mb

ID: 181848

Now go to the EBI SRA page using the following URL substituting "SRR_number" for yours:

http://www.ebi.ac.uk/ena/data/view/SRR_number

eg: <http://www.ebi.ac.uk/ena/data/view/SRR494099>

You will be taken to the ENA – European Nucleotide Archive.

EMBL-EBI Services

ENA European Nucleotide Archive

[ENA Home](#) | [Search & Browse](#) | [Submit & Update](#) | [About ENA](#) | [Contact](#)

Please subscribe to ena-announce@ebi.ac.uk to receive alerts about ENA services.

[Text search](#) | [Advanced search](#) | [Sequence search](#)

Enter or paste text or ENA accession number: Upload file of accessions: no file selected

Read: SRR494099 : Illumina Genome Analyzer II sequencing; GSM927238: HeLa; Homo sapiens; OTHER

View: [XML](#) Download: [XML](#)
[Send Feedback](#)

Submitting Centre	Run Date	Platform	Model	Read Count	Base Count
GEO		ILLUMINA	Illumina Genome Analyzer II		
Library Layout	Library Strategy	Library Source	Library Selection	Library Name	
SINGLE	OTHER	GENOMIC	other	GSM927238: HeLa	
Broker Name	NCBI				

Navigation **Read Files**

This table contains the files for run SRR494099

[Download files](#) Download: [TEXT](#)

View: [TEXT](#)
[Select columns](#)

Showing results 1 - 1 of 1 results

Study accession	Secondary study accession	Sample accessions	Experiment accession	Run accession	Scientific name	Instrument model	Library layout	Fastq files (ftp)	Fastq files (galaxy)	Submitted files (ftp)	Submitted files (galaxy)
SRP012667	SRP012667	SRS312185	SRX145994	SRR494099	Homo sapiens	Illumina Genome Analyzer II	SINGLE	File 1	File 1	Not available	Not available

For Aspera download, please [download and install Aspera Connect](#)

Look for the fastq files (ftp) link and right-click on the link. A pop-up menu will appear – select **Copy Link**:

Navigation **Read Files**

This table contains the files for run SRR494099

[Download files](#) Download: [TEXT](#)

View: [TEXT](#)
[Select columns](#)

Showing results 1 - 1 of 1 results

Study accession	Secondary study accession	Sample accessions	Experiment accession	Run accession	Scientific name	Instrument model	Library layout	Fastq files (ftp)	Fastq files (galaxy)	Submitted files (ftp)	Submitted files (galaxy)
SRP012667	SRP012667	SRS312185	SRX145994	SRR494099	Homo sapiens	Illumina Genome Analyzer II	SINGLE	File 1	File 1	Not available	Not available

For Aspera download, please [download and install Aspera Connect](#)

- Open Link in New Window
- Open Link in New Tab
- Download Linked File
- Download Linked File As...
- Add Link to Bookmarks...
- Add Link to Reading List
- Copy Link
- Add to iTunes as a Spoken Track

EMBL-EBI Services

News By topic
Brochures By name (A-Z)

Research Overview
Publications Train at EBI

Paste this link into a text file, eg:

<ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR494/SRR494099/SRR494099.fastq.gz>

You can use this link with the unix command 'wget' to download the fastq file; connect to your CBRG account and move to your HTS space – *do not download HTS data under your home directory!* (please contact CBRG if you do not know where your HTS space is)

Then type

wget <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR494/SRR494099/SRR494099.fastq.gz> &