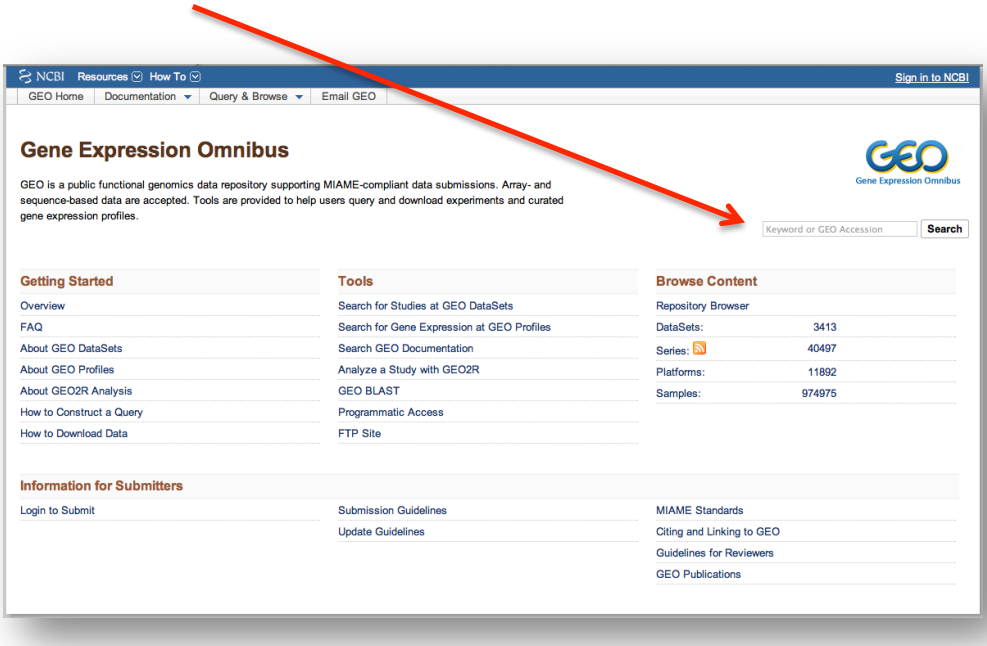
Downloading published fastq data from GEO

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

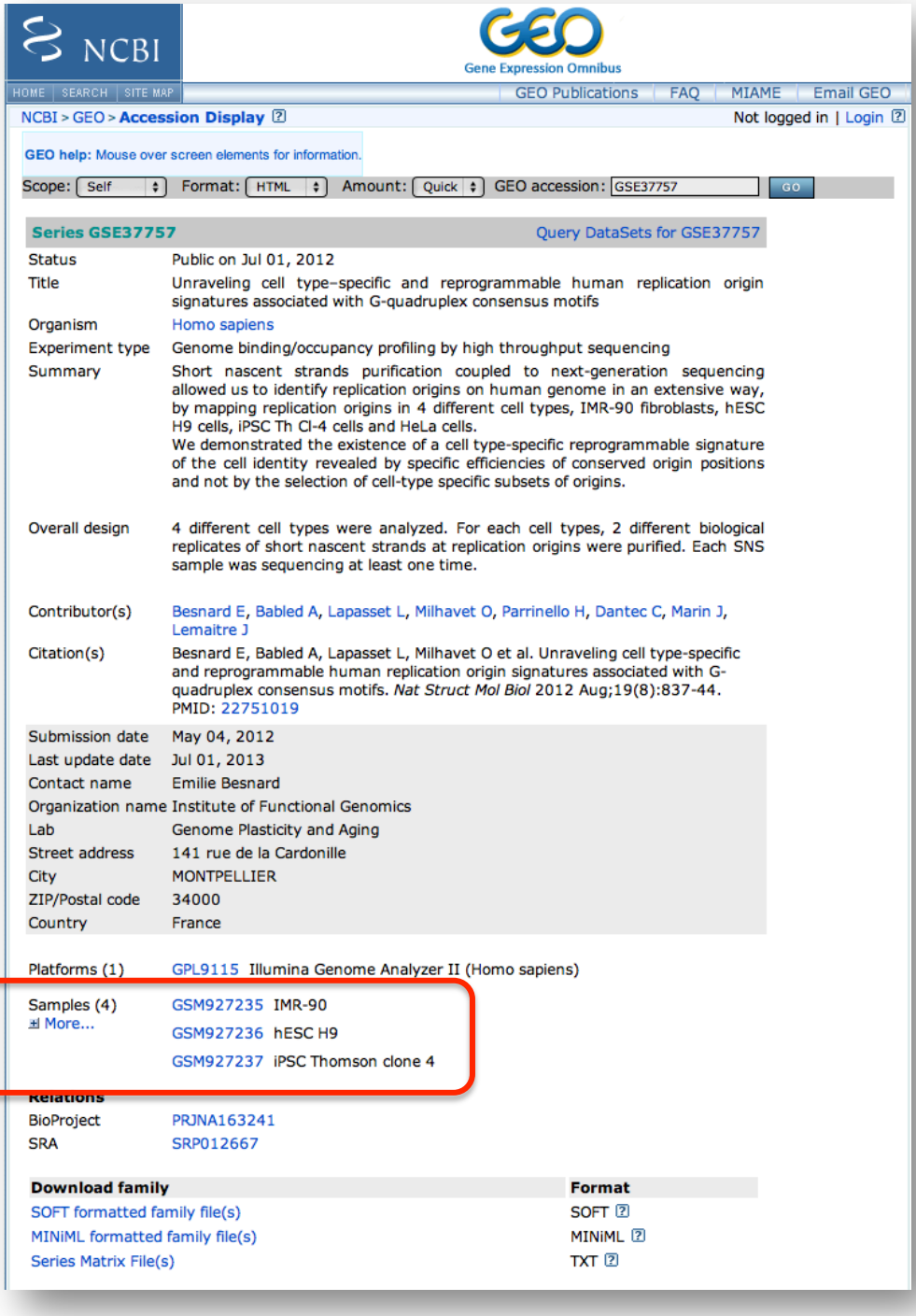
Look in the paper for the GEO accession number and then go to the GEO website: http://www.ncbi.nlm.nih.gov/geo/

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



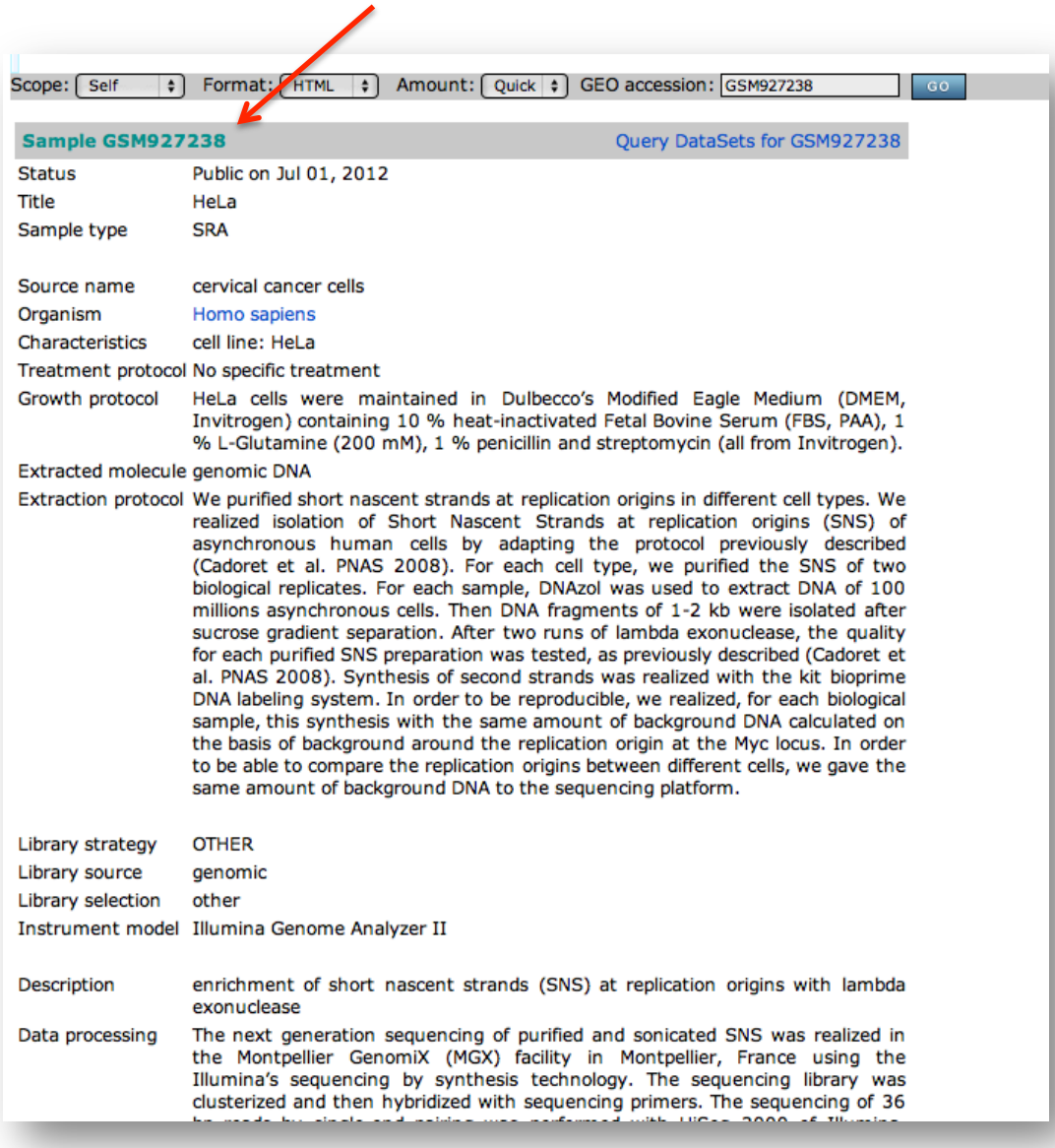
You will then see a page for that **Series**:

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

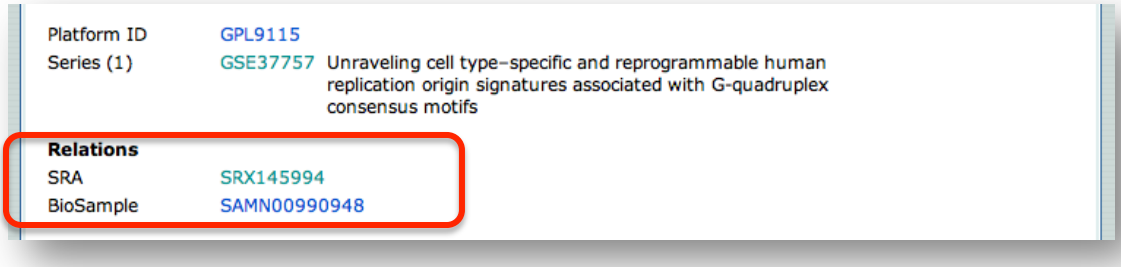


Click on one the sample links eg '**GSM927238'**

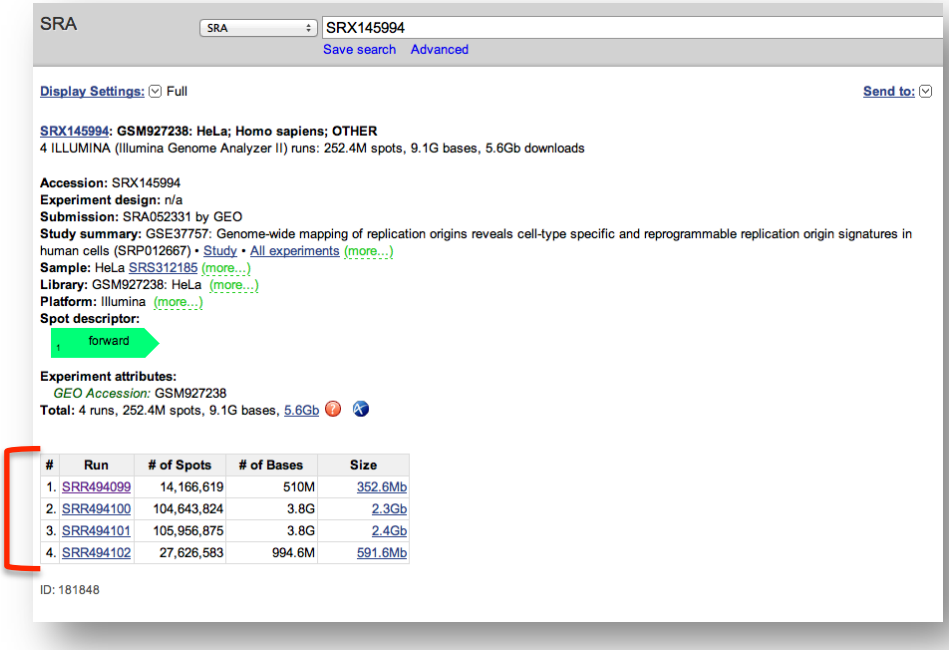
You will now see an entry for that sample



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



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:

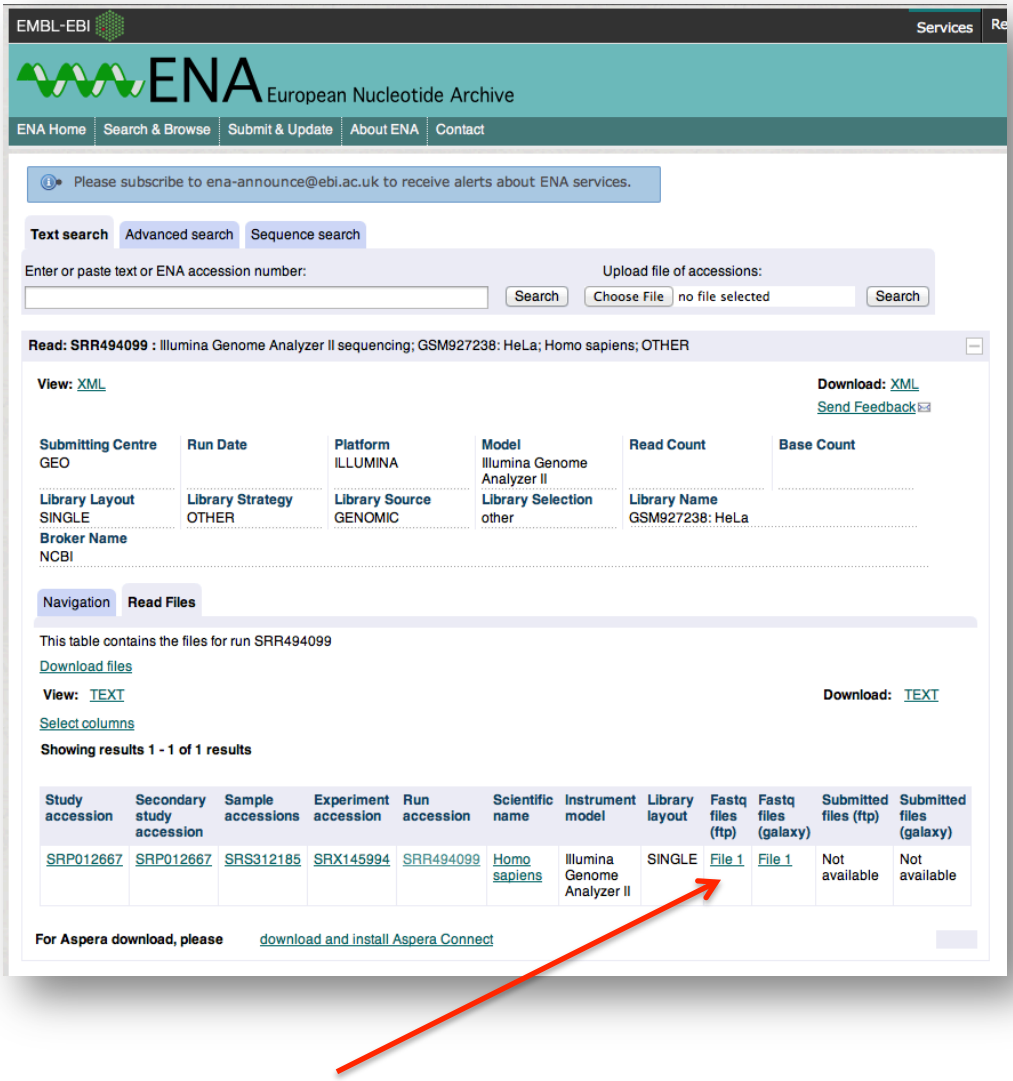


Now go to the EBI SRA page using the following URL substituting “SRR_number” for yours:

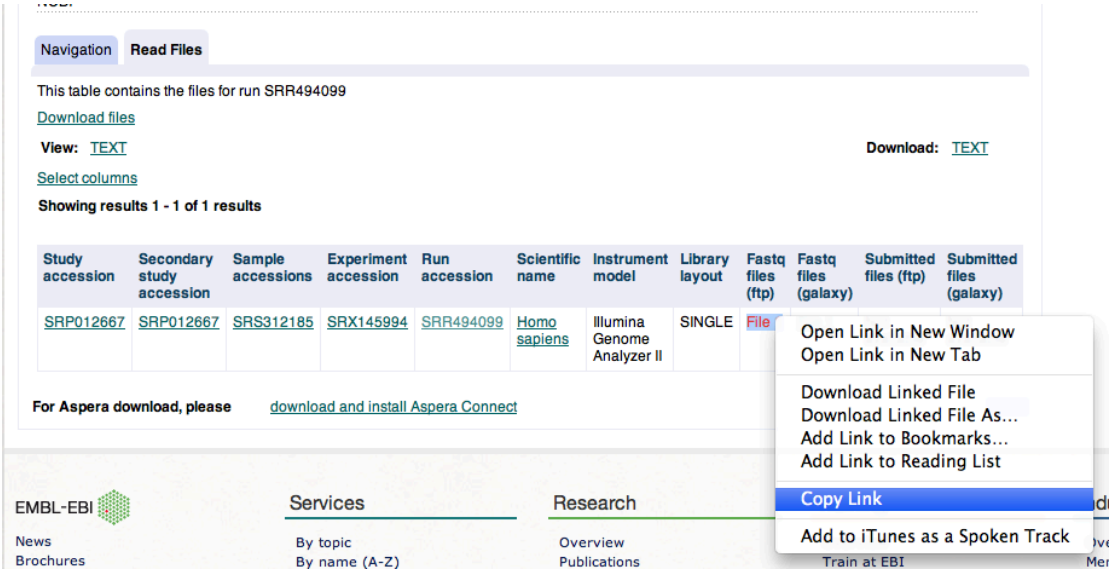
http://www.ebi.ac.uk/ena/data/view/SRR_number

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

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



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



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 &
```