**Onco-Wiki Biological Data Search**

1. **Biological Data Search**

Biological data for RNA-Sequencing can be found in public repositories on the internet and are published in the form of ‘reads.’ Reads are short nucleotide fragments complementary to the original mRNA strand which are mapped to the annotated reference genome as a measure of gene expression in RNA-Sequencing. Reads from original publications can be search for at these public repositories:

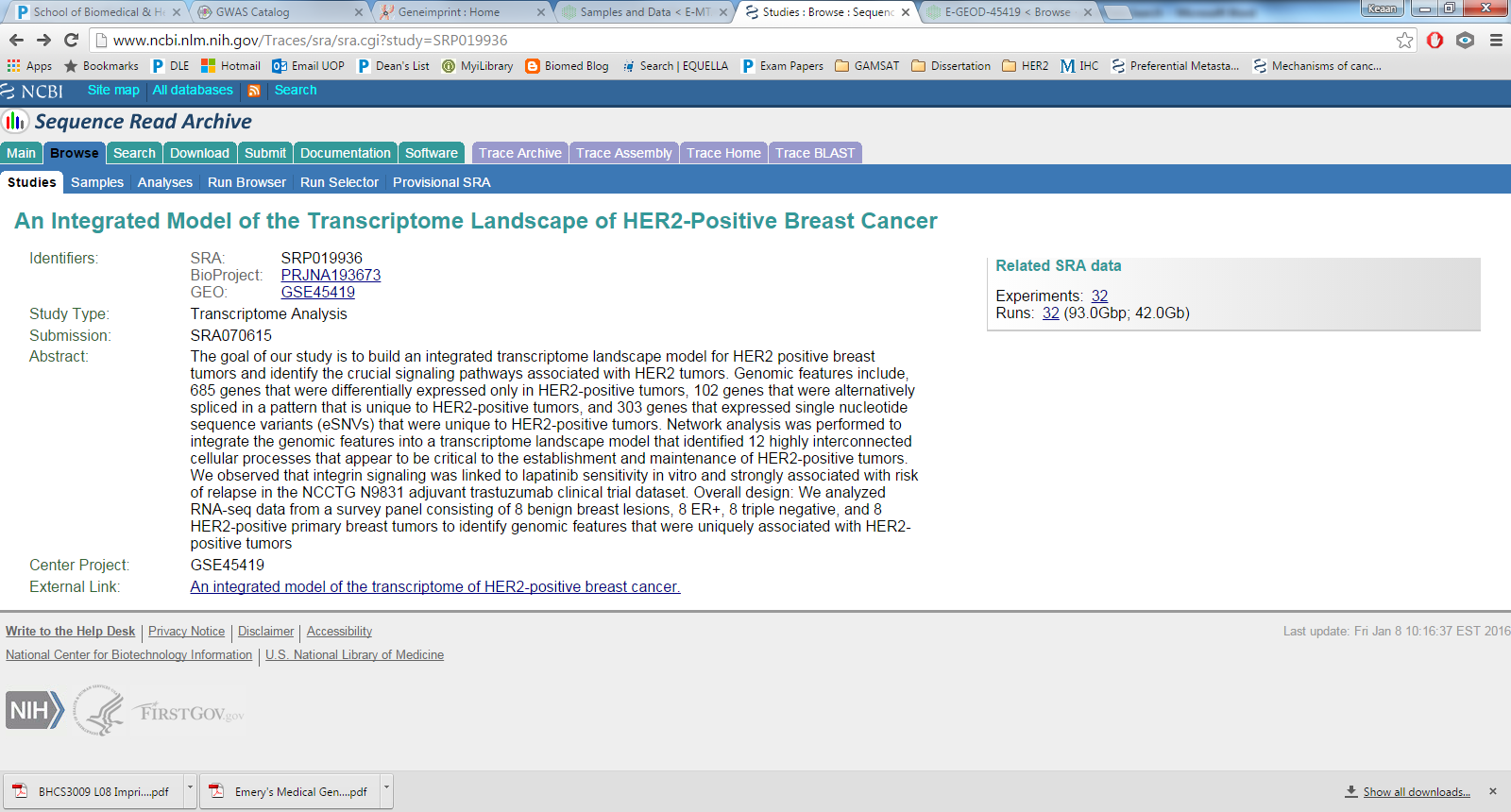
|  |  |
| --- | --- |
| Sequence Read Archive (NCBI) | ArrayExpress (EBI) |
| <http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=studies> | [**https://www.ebi.ac.uk/arrayexpress/browse.html**](https://www.ebi.ac.uk/arrayexpress/browse.html) |

1. **Considerations**

Ensure that the biological data you find arises from RNA-Sequencing experiments and this published data was used by the original authors towards their own scientific paper publication. Ideally, try to find studies that have used human patient tissue biopsies and have been explicit in doing so. If this is not possible, look to use human cell line ‘in vitro’ data over organoid, stem cell or xenograft data.

1. **Biological Data Download**

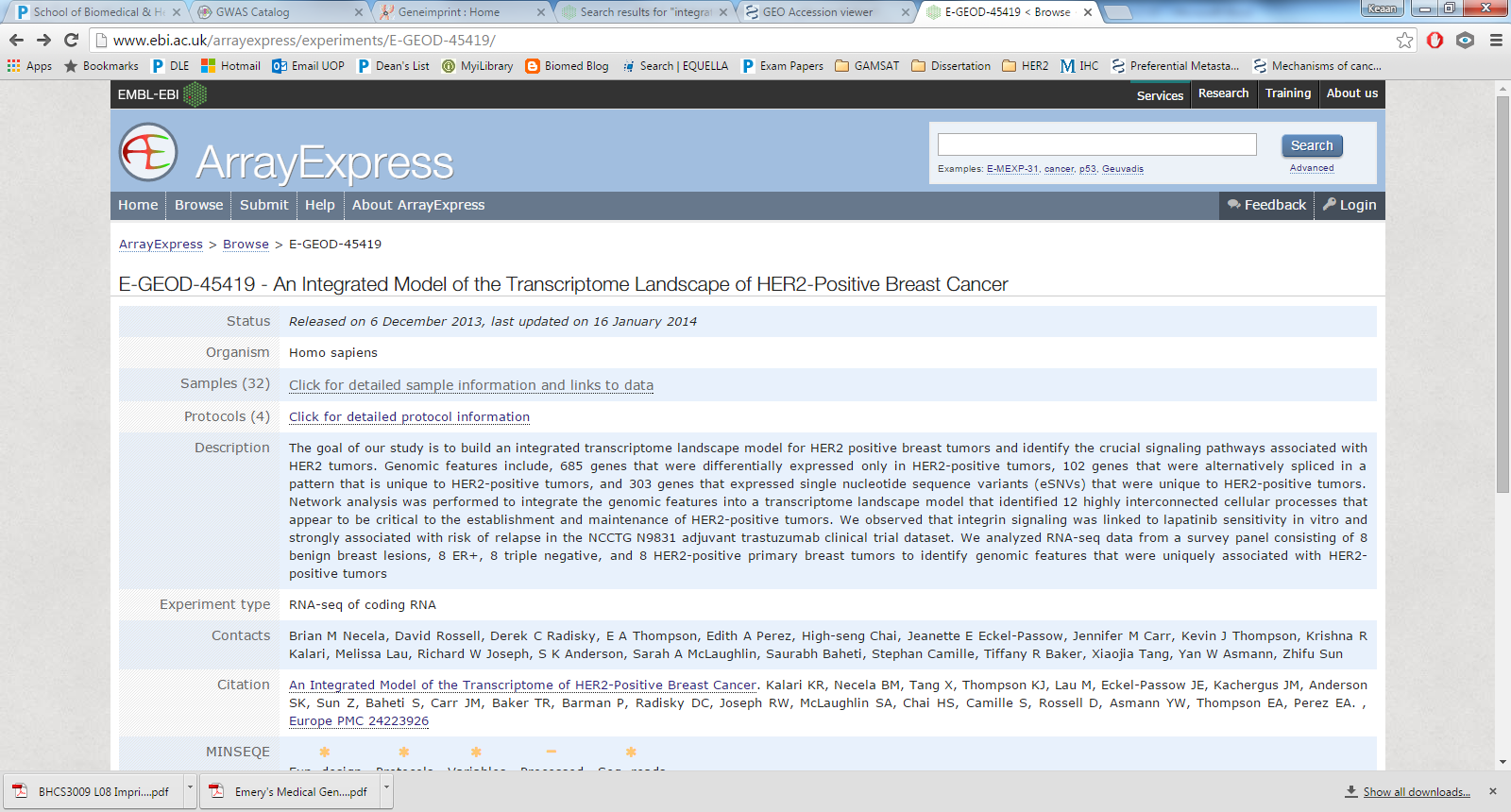
Prior to data download, ensure that you save your data to an external hard drive with enough memory capacity. Biological Data for each sample can be up 14-16 GB per sample).

**Sequence Read Archive** – After searching for a representative study, click on the SRP Accession number and the following screen should appear. Click on the GEO number (e.g. GSE45419)

On the GEO Accession Viewer screen on the right, scroll down to sample information and click on an individual sample. Scroll down to the bottom of this screen and find the SRX number under relations. Save this SRX number and enter the SRX number via this link: <http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=search_seq_name>

Choose FASTA format and download for each sample. The file format extension should be: .fasta.gz

**ArrayExpress** – After searching for a representative study, click on the Accession number and the following screen should appear. Select ‘Samples’ for detailed sample information and links to data.



Under the Links to Data subheading, a download via FASTA or FASTQ option may be shown. Select FASTA (smaller file size) if offered or FASTQ. If not, click on ENA, scroll down and download under the FASTQ files (ftp) subheading. Ensure that you download all files corresponding to the individual sample and repeat for each different sample.