

## Week 3 – Dataset identification

The following dataset was identified from NCBI for the purpose of transcriptomic data analysis and for studying differential expression of genes over the course of fibroblast conversion to neuron:

### Timeline RNAseq of fibroblast to neuron direction conversion

Accession: PRJEB32551 ID: 544668

Timeline RNAseq to identify gene expression dynamics over the course of conversion of fibroblasts into induced neurons.

Accession	PRJEB32551
Scope	Monoisolate
Submission	Registration date: 24-May-2019 European Bioinformatics Institute

#### Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
SRA Experiments	7
OTHER DATASETS	
BioSample	7

▼ SRA Data Details

Parameter	Value
Data volume, Gbases	11
Data volume, Mbytes	4905

This dataset was chosen due to the following reasons:



- It provides 7 SRA sequencing experiments which are comparatively easier to analyse
- The datasets were derived over a period of a month. A fibroblast control was first sequenced followed by transcriptome sequencing of the maturing samples at 1 day, 3 days, 6 days, 12 days, 18 days and 24 days of maturation. This can help us observe gene expression patterns over a period of time.
- The accession list was downloaded in .txt format using the 'Run selector' option on NCBI and the .fastq versions of the files were derived by using the "Fast download FASTQ files" option on galaxy, the files were then split from the collection using the "extract dataset" option and renamed with suitable metadata i.e. days of incubation passed prior to sequencing and restructured into a collection using the "Build dataset list" option.

Select




	Runs	Bytes	Bases	Download	Cloud Data Delivery	Computing
Total	7	4.57 Gb	10.99 G	Metadata or Accession List		
Selected	0	0	0	Metadata or Accession List Full List of Accessions in Recordset	Deliver Data	Galaxy


Found 7 Items

<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Run	BioSample	Bases	Bytes	cell_type	ENA-LAST-UPDATE (Sample)	Experiment	Experimental_Factor_time	external_id	INSDC_last_update	Libra
<input type="checkbox"/>	1	ERR3322430	SAMEA5605439	2.00 G	853.49 Mb	induced neuron	2019-05-13T11:58:15Z	ERX3347212	12	SAMEA5605439	2019-05-13T11:58:15Z	06_12d
<input type="checkbox"/>	2	ERR3322431	SAMEA5605440	1.60 G	683.14 Mb	induced neuron	2019-05-13T11:58:15Z	ERX3347213	18	SAMEA5605440	2019-05-13T11:58:15Z	06_18d
<input type="checkbox"/>	3	ERR3322432	SAMEA5605441	1.76 G	749.49 Mb	induced neuron	2019-05-13T11:58:15Z	ERX3347214	1	SAMEA5605441	2019-05-13T11:58:15Z	06_1d_s
<input type="checkbox"/>	4	ERR3322433	SAMEA5605442	1.55 G	658.41 Mb	induced neuron	2019-05-13T11:58:15Z	ERX3347215	24	SAMEA5605442	2019-05-13T11:58:15Z	06_24d
<input type="checkbox"/>	5	ERR3322434	SAMEA5605443	1.22 G	517.22 Mb	induced neuron	2019-05-13T11:58:15Z	ERX3347216	3	SAMEA5605443	2019-05-13T11:58:15Z	06_3d_s
<input type="checkbox"/>	6	ERR3322435	SAMEA5605444	1.74 G	738.17 Mb	induced neuron	2019-05-13T11:58:15Z	ERX3347217	6	SAMEA5605444	2019-05-13T11:58:15Z	06_6d_s
<input type="checkbox"/>	7	ERR3322436	SAMEA5605445	1.12 G	478.63 Mb	fibroblast	2019-05-13T11:58:16Z	ERX3347218	0	SAMEA5605445	2019-05-13T11:58:16Z	06_Fib_s

44: Single end files (RNA seq)  

a list with 7 **fastqsanger.gz** datasets






1: GP.txt   

Add Tags 

7 lines


format **txt**, database ?

uploaded txt file


    


ERR3322430  
ERR3322431  
ERR3322432  
ERR3322433  
ERR3322434


<< History: GP miniproject


Single end files (RNA seq) 


a list with 7 **fastqsanger.gz** datasets


Download 


1: ERR3322433\_24d 

2: ERR3322431\_18d 

3: ERR3322430\_12d 

4: ERR3322435\_6d 

5: ERR3322434\_3d 

6: ERR3322432\_1d 

7: ERR3322436\_Fibroblast 