

In [3]:

```
import pandas as pd
import numpy as np

df=pd.read_csv("metadata_processed_files_hg38.tsv",sep='\t')
df.head()
```

Out[3]:

	File accession	File format	File type	File format type	Output type	File assembly	Experiment accession	Assay	
0	ENCFF404WOP	bam	bam	NaN	alignments	GRCh38	ENCSR977OXG	eCLIP	donors
1	ENCFF080HMO	bam	bam	NaN	alignments	GRCh38	ENCSR977OXG	eCLIP	donors
2	ENCFF224ZJZ	bigWig	bigWig	NaN	minus strand signal of unique reads	GRCh38	ENCSR977OXG	eCLIP	donors
3	ENCFF394GOQ	bigWig	bigWig	NaN	minus strand signal of unique reads	GRCh38	ENCSR977OXG	eCLIP	donors
4	ENCFF292ZPN	bigWig	bigWig	NaN	plus strand signal of unique reads	GRCh38	ENCSR977OXG	eCLIP	donors

5 rows × 59 columns

In [4]:

```
df=df.loc[:,["File accession","File format","File assembly","Biosample term name","Experiment accession"]]
```

In [5]:

```
df=df[(df["File format"]=="bed narrowPeak")&(df["Biological replicate(s)"]=="1, 2")&(df["File assembly"]=="GRCh38")]
```

In [6]:

```
df.to_csv("metadata_processed_files_hg38_version_1.tsv")
```

In [7]:

```
df.shape
```

Out[7]:

```
(238, 8)
```