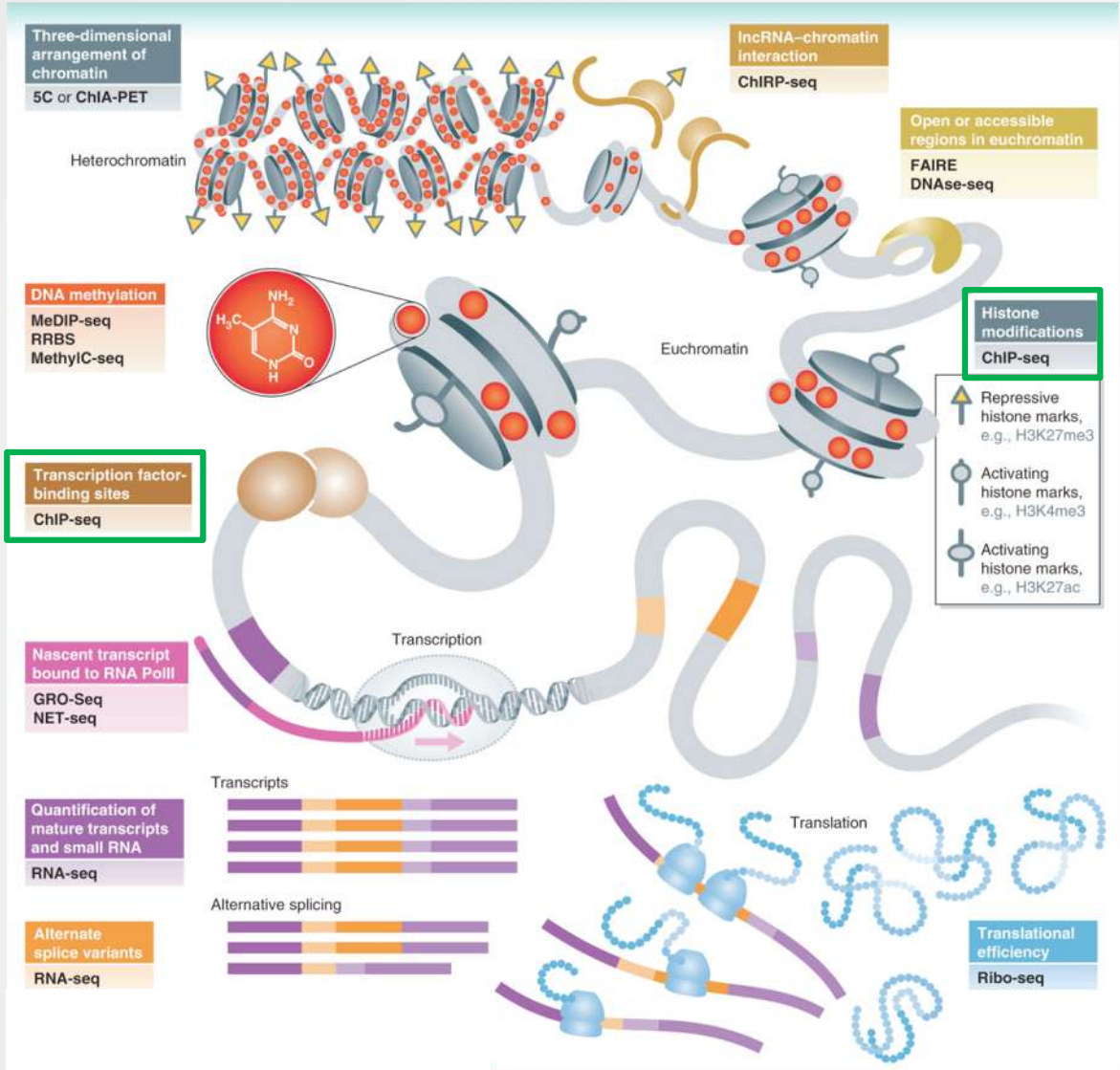
ChIP-sequencing: Library preparation and experimental design

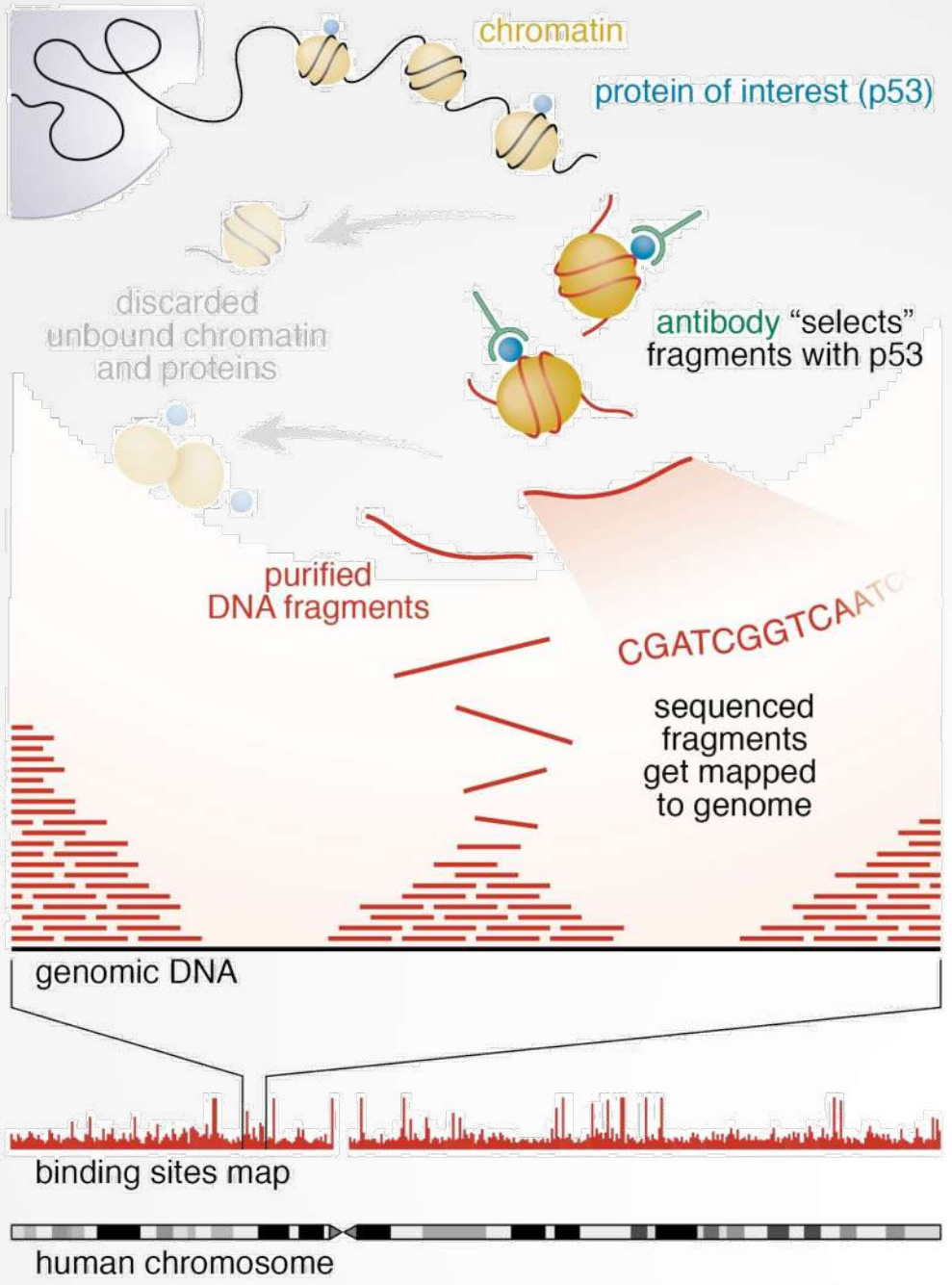
Stéphanie Le Gras
(slegras@igbmc.fr)

Sequencing technologies and their uses



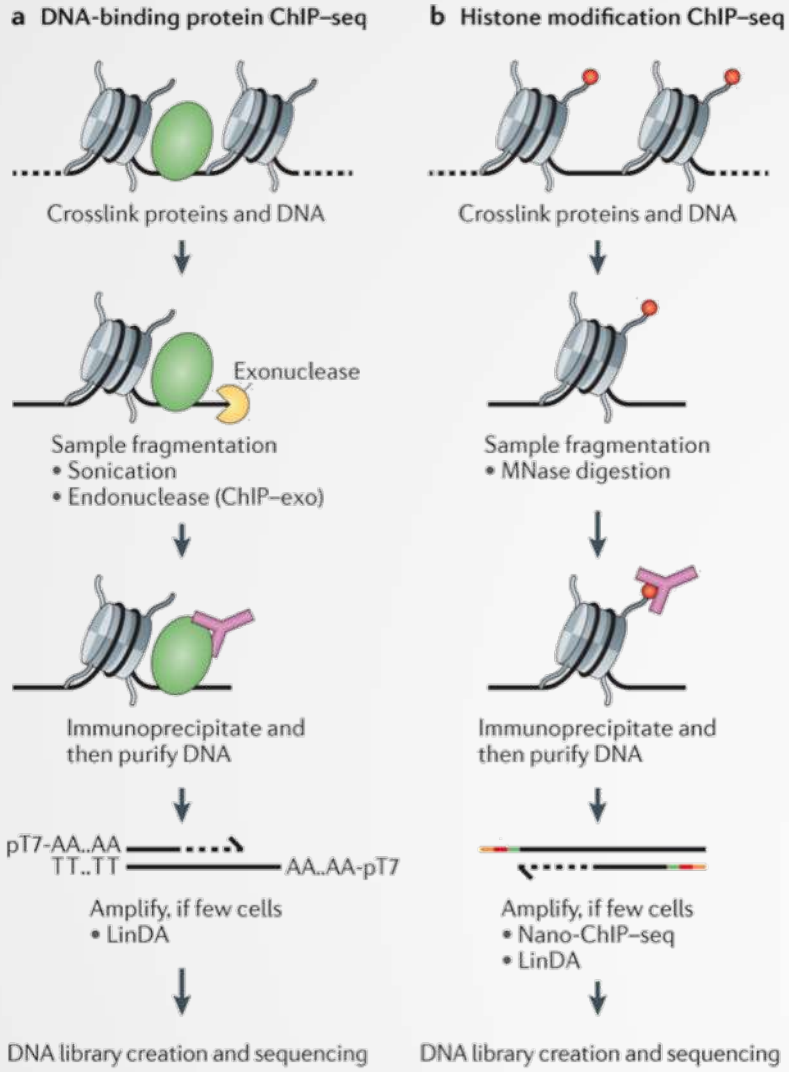
Soon et al, 2013

ChIP-seq


chromatin
protein of interest (p53)
discarded
unbound chromatin
and proteins
antibody "selects"
fragments with p53
purified
DNA fragments
CGATCGGTCAATC
sequenced
fragments
get mapped
to genome
genomic DNA
binding sites map
human chromosome

ChIP and library prep considerations

Chromatin ImmunoPrecipitation



Nature Reviews Genetics 13, 840-852 (December 2012) doi:10.1038/nrg3306

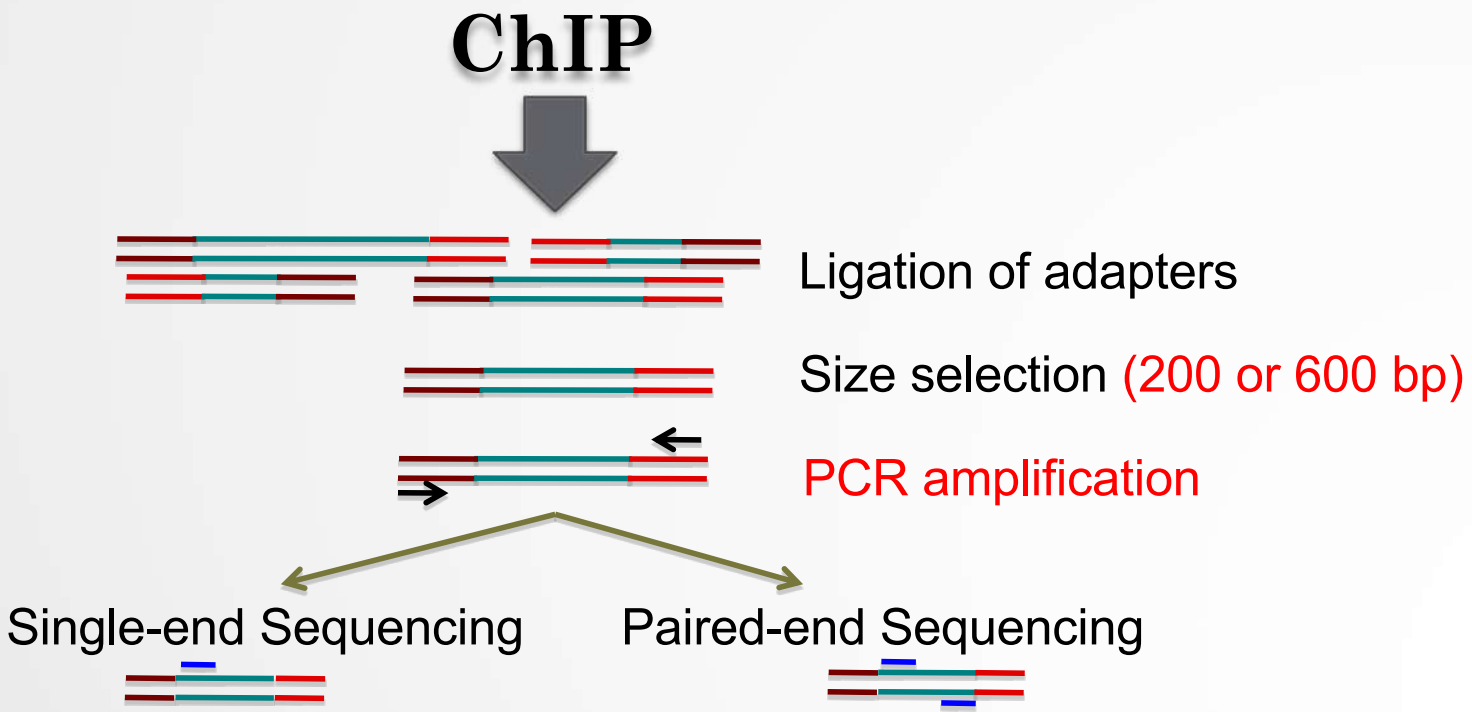
Considerations on chIP

- Antibody
  - Antibody quality varies, even between independently prepared lots of the same antibody (Egelhofer, T. A. *et al*. 2011)
- Number of cells
  - large number of cells are required for a ChIP experiment (limitation for small organisms)
    - Nano-ChIP–seq (Adli et al, 2011)
    - LinDA (Shankaranarayanan et al, 2011)
- Shearing of DNA (Mnase I, sonication, Covaris): trying to narrow down the size distribution of DNA fragments

→ **Complexity in DNA fragments**

Library prep

- Step between chIP and sequencing
- The goal is to prepare DNA for the sequencing
- Starting material: ChIP sample (1-10ng of sheared DNA)



Library prep

- PCR amplification : to increase amount of starting DNA
  - Number of PCR cycles as low as possible
  - Protocols with low quantity of starting material

Sequencing

- Sequencer : Illumina HiSeq 4000
- No. of reads per run, per sample :
  - 1st run on the GAIIx : 10-20 millions of reads per lane
  - (HiSeq 2500) 4 samples per lane :~41 millions per sample
  - (HiSeq 4000) 8 samples per lane :~43 millions per sample
- Length of DNA fragment : ~200-600bp
- No. of cycle per run : 50

Single end or paired end?

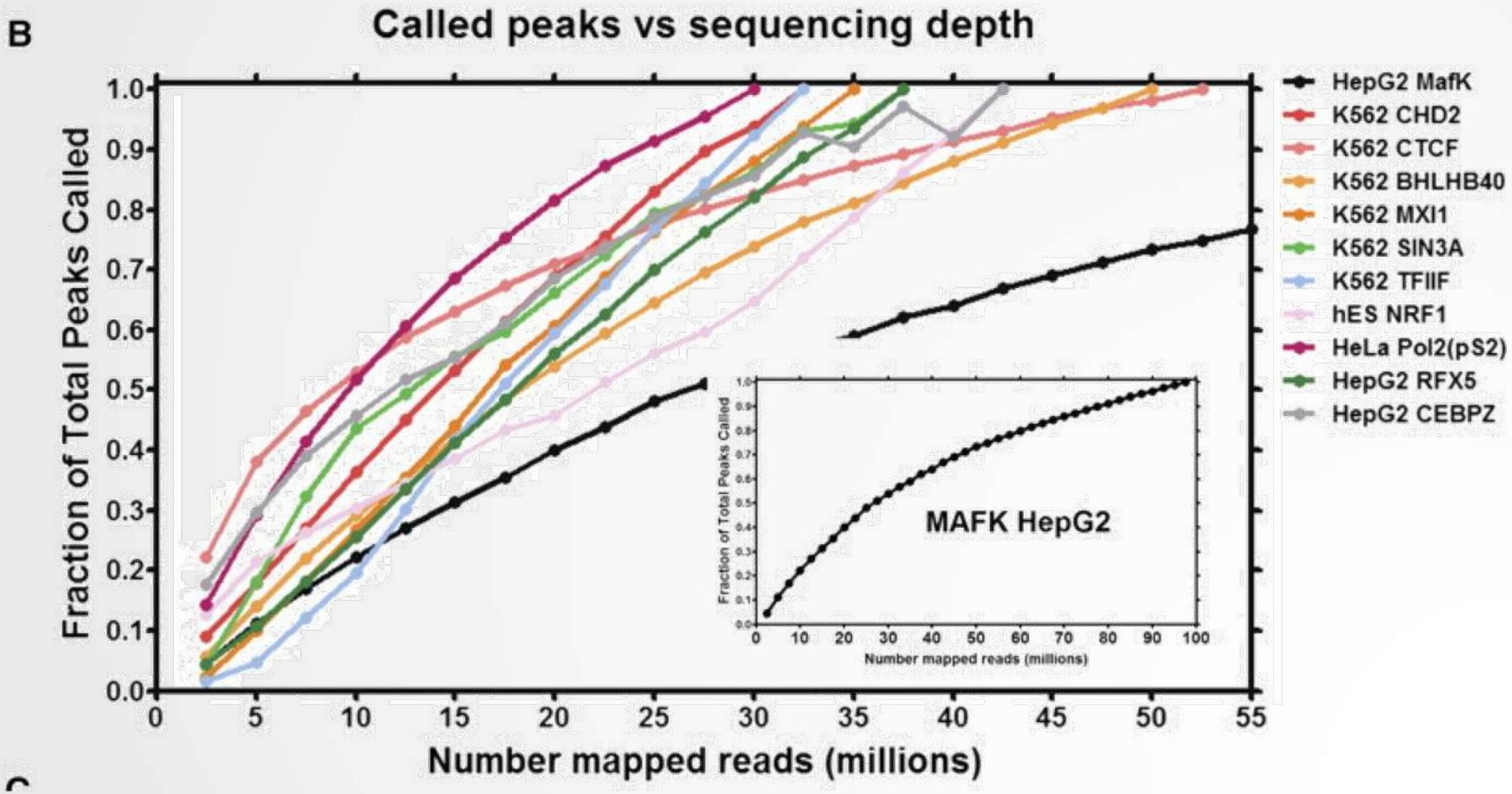
- Single end (most of the time)
- Paired-end sequencing
  - Improve identification of duplicated reads
  - Better estimation of the fragment size distribution
  - Increase the mapping efficiency to **repeat regions**
  - The price!

Sequencing depth

Consider the depth needed depending on:

- chipped protein,

Sequencing depth



Landt et al, 2012

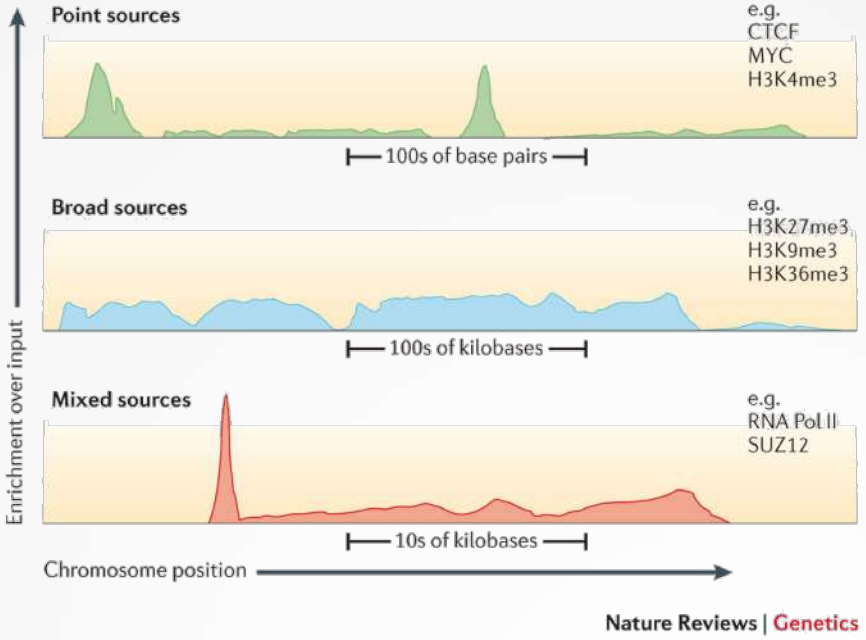
Sequencing depth

Consider the depth needed depending on:

- chipped protein,
- type of expected profile,
- number of expected binding sites,
- size of the genome of interest.

Ex:

- For human genomes, 20 million uniquely mapped read sequences are suggested for point-source peaks, or 40 million for broad-source peaks.
- For fly genome: 8 million reads
- For worm genome: 10 million reads



Controls

- Used mostly to filter out false positives (high level of noise)
  - Idea: potential false positive will be enriched in both treatment and control.
- A control will fail to filter out false positives if its enrichment profile is very different from the enrichment profile of false positive regions in the treatment sample
- 3 types of control are commonly used :
  - Input DNA : a portion of DNA sample removed prior to IP
  - DNA from non specific IP : DNA obtained from IP with an antibody not known to be involved in DNA binding or chromatin modification such as IgG
  - Mock IP DNA : DNA obtained from IP without antibodies
- Choice of control is extremely important
- It is recommended to cover the control in a higher extend than the IPs

Replicates

- A minimum of two replicates should be carried out per experiment.
- Each replicate should be a biological rather than a technical replicate; that is, it represents an independent cell culture, embryo pool or tissue sample.

ENCODE

- The Encyclopedia of DNA Elements (ENCODE) Consortium has carried out hundreds of ChIP–seq experiments and has used this experience to develop a set of working standards and guidelines

See: https://www.encodeproject.org/about/experiment-guidelines/

Data used in this course

Sample name	No. of raw reads
MITF	31,334,257
Ctrl	29,433,042
H3K4me3	11,192,622
polII	10,404,820