Bioinformatic approaches to regulatory genomics and epigenomics
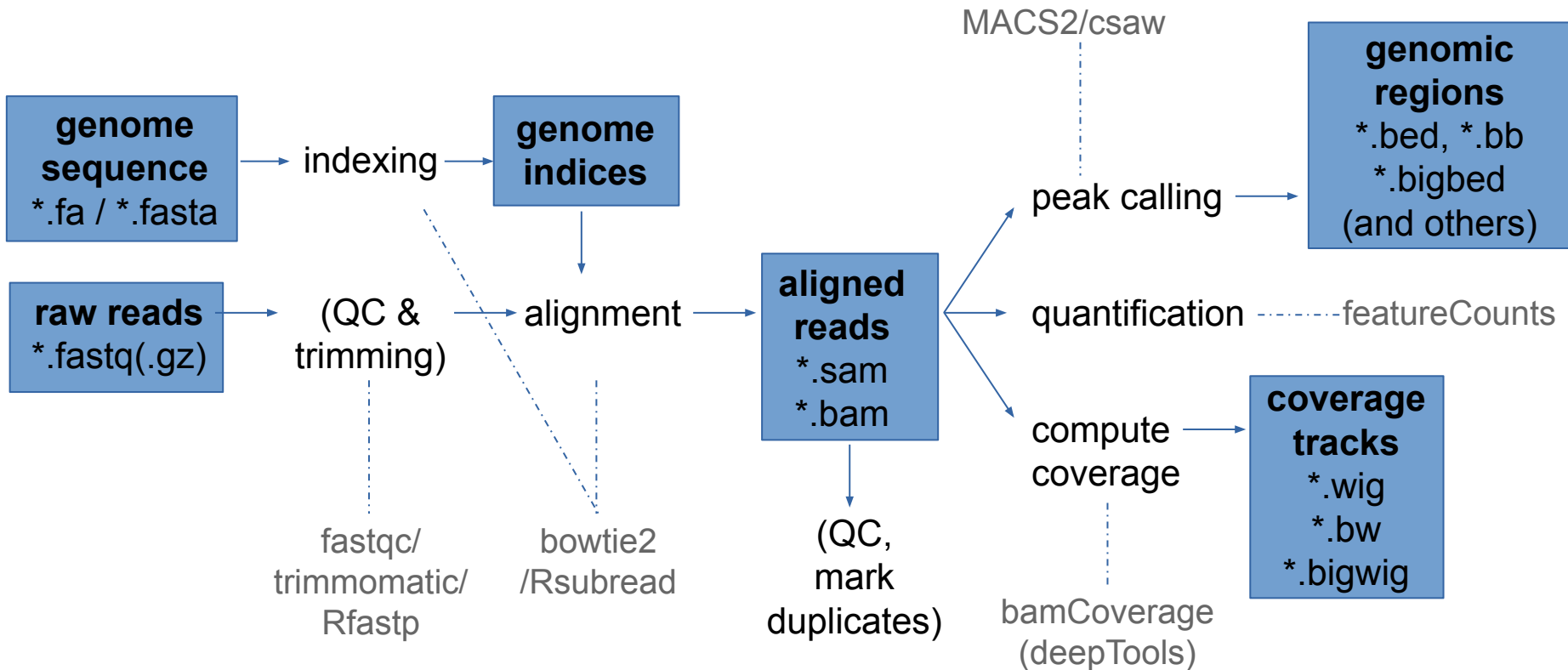
376-1347-00L - 2022 | week 04

Pierre-Luc Germain
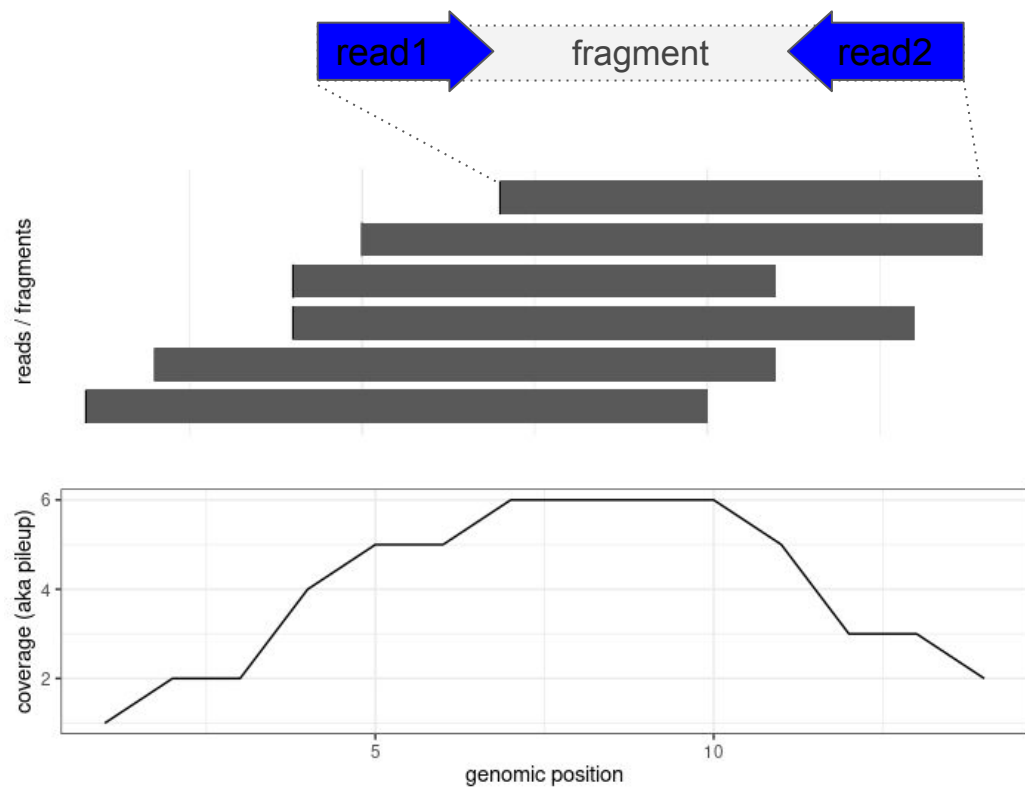
ETH Zürich

Plan for today

- Recap of last week

- Visualizing epigenomic signals

- Functional elements

- Finding data from the literature

Overview of a primary analysis pipeline (ChIP-seq and the likes)

genome sequence
*.fa / *.fasta → indexing → **genome indices**

raw reads
*.fastq(.gz) → (QC & trimming) → alignment → **aligned reads** *.sam *.bam

fastqc/ trimmomatic/ Rfastp

bowtie2 /Rsubread

(QC, mark duplicates)

MACS2/csaw

peak calling → **genomic regions** *.bed, *.bb *.bigbed (and others)

quantification ----- featureCounts

compute coverage → **coverage tracks** *.wig *.bw *.bigwig
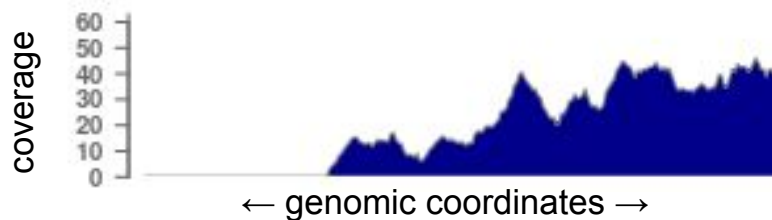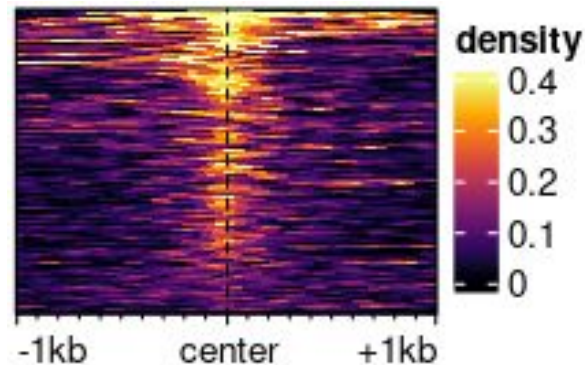
bamCoverage (deepTools)

Recap

Visualizations available in *epiwraps*

- Signal across one genomic region:
  `plotSignalTracks`

(Based on the *Gviz* R package)

- Signal across several genomic regions:
  `signal2Matrix` →
  `plotEnrichedHeatmaps`

(Mainly based on the EnrichedHeatmap R package, itself based on ComplexHeatmap)
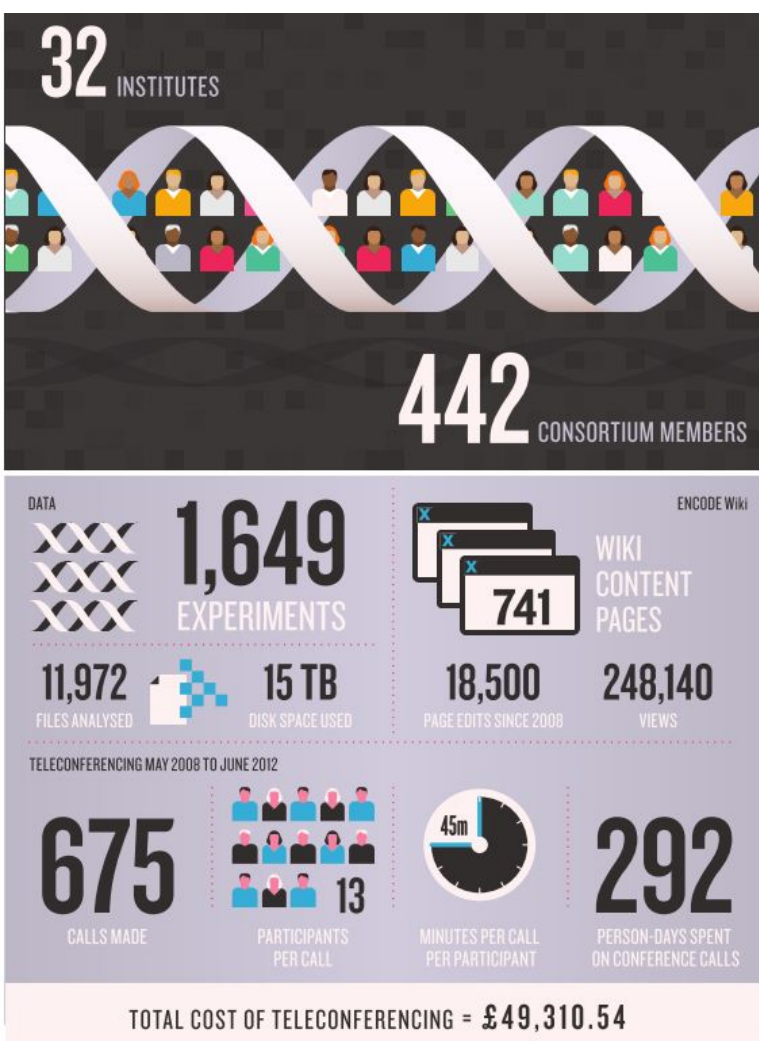
Read extension in coverage track generation

read1   fragment

read1   fragment

fragment   read1

Coverage without
read extension

Coverage with
read extension

The ENCyclopedia Of DNA Elements

~30 publications in September 2012

$288 million USD

… then an ENCODE2, 3, now working towards the 5…

An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium

Nature **489**, 57–74 (2012) | Cite this article

Bits of Mystery DNA, Far From 'Junk,' Play Crucial Role

The New York Times

by Gina Kolata

"At least 80 percent of this DNA is *active* and *needed*."

The evolutionary arguments for junk:
- 1% protein-coding
- ~4 to 10% evolutionarily conserved
- >50% transposable elements
- Onions have a 5 times bigger genome

The very angry response:
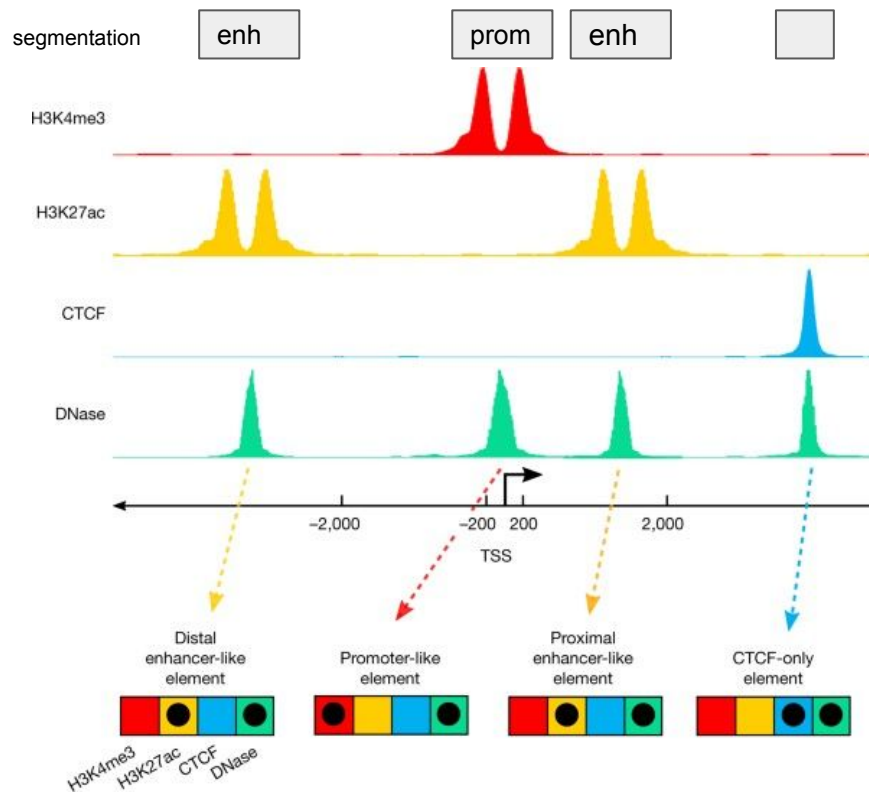- Graur et al., GBE 2013

NEWS&ANALYSIS

GENOMICS

ENCODE Project Writes Eulogy For Junk DNA

–ELIZABETH PENNISI

SCIENCE   VOL 337   7 SEPTEMBER 2012

(For more on the topic, see Germain, Ratti and Boem 2014; Ratti and Germain *in print*.)

A signature-based encyclopedia of DNA elements

ENCODE's "signature strategy":

- Different types of functional genetic elements are associated with different chemical signatures

- We can identify functional elements by identifying these signatures genome-wide
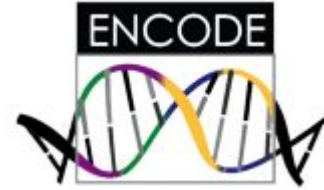
Generic repositories for NGS data

https://www.ncbi.nlm.nih.gov/geo/

https://www.ebi.ac.uk/ena/

https://www.ncbi.nlm.nih.gov/sra

Quality-controlled and uniformly processed human and mouse NGS datasets

www.roadmapepigenomics.org

www.encodeproject.org

(hematopoietic system)

Assignment

- Complete last week's assignment

- Find and download the peak-level data (i.e. bed-like format) for the following histone modifications in mouse embryonic stem cells (mESC) from ENCODE:

  - H3K4me3, H3K4me1, H3K27ac, and H3K27me3

- Of the p300 peaks you identified last week, what proportion overlap each of the marks?