Analysis of broad occupancy patterns

Workshop on **ChIP-seq data analysis**

Stockholm, 8 November 2018

Agata Smialowska
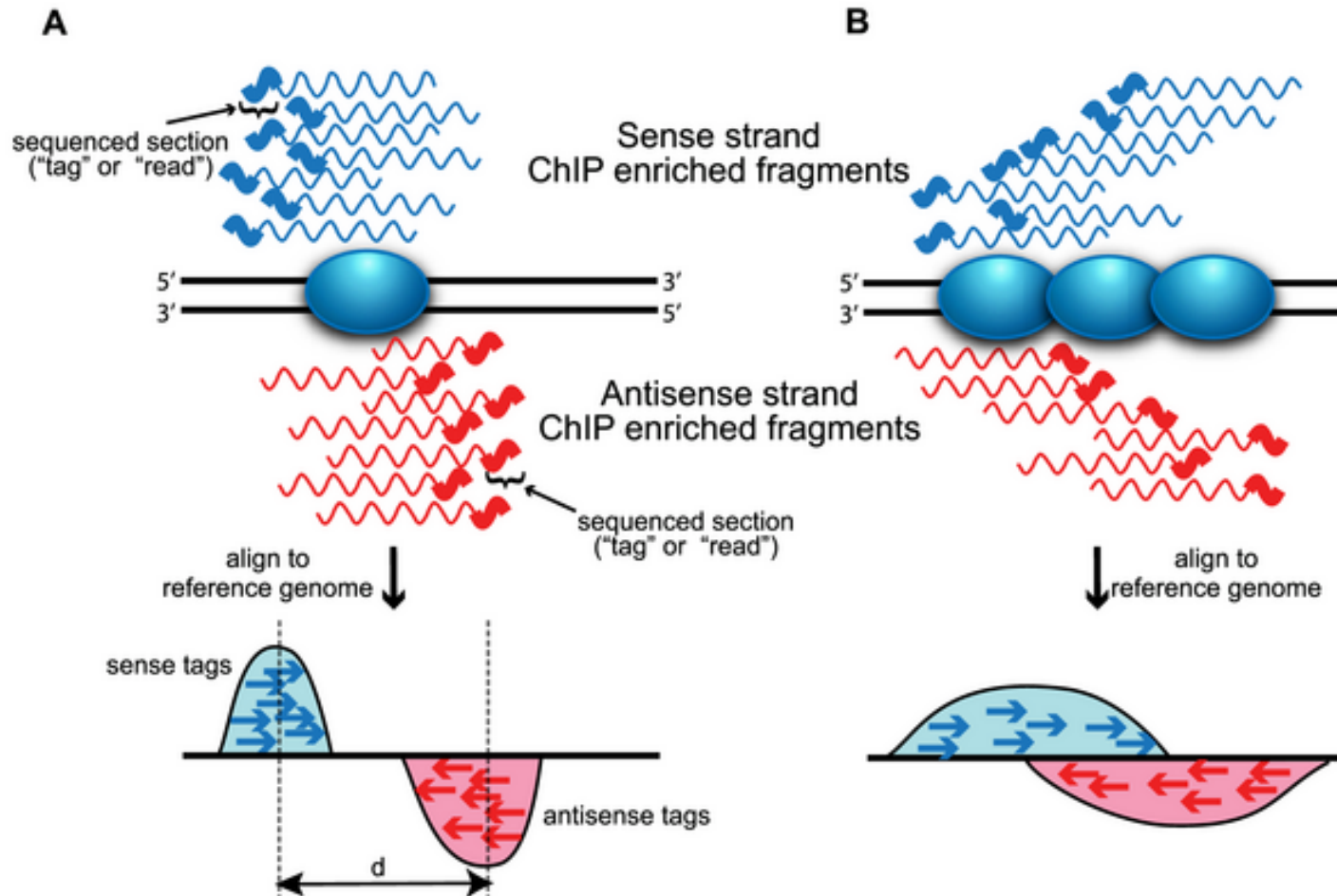
NBIS, SciLifeLab, Stockholm University
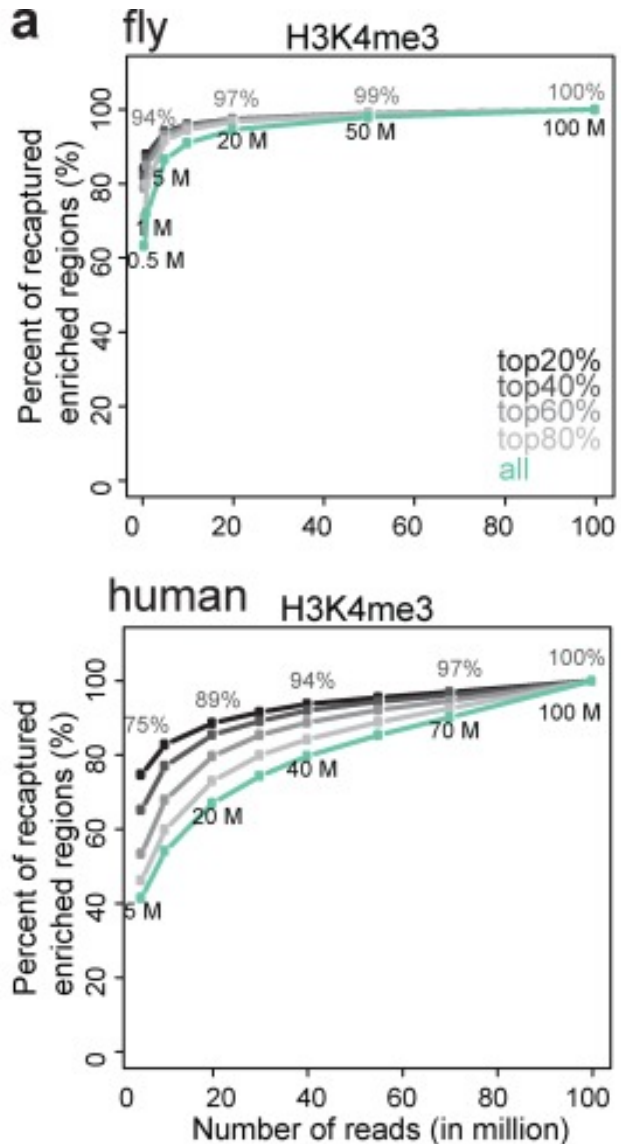
Point-source vs. broad peak detection

Sequence-specific binding (TFs)   Distributed binding (histones, RNApol2)

How to define sufficient read depth?

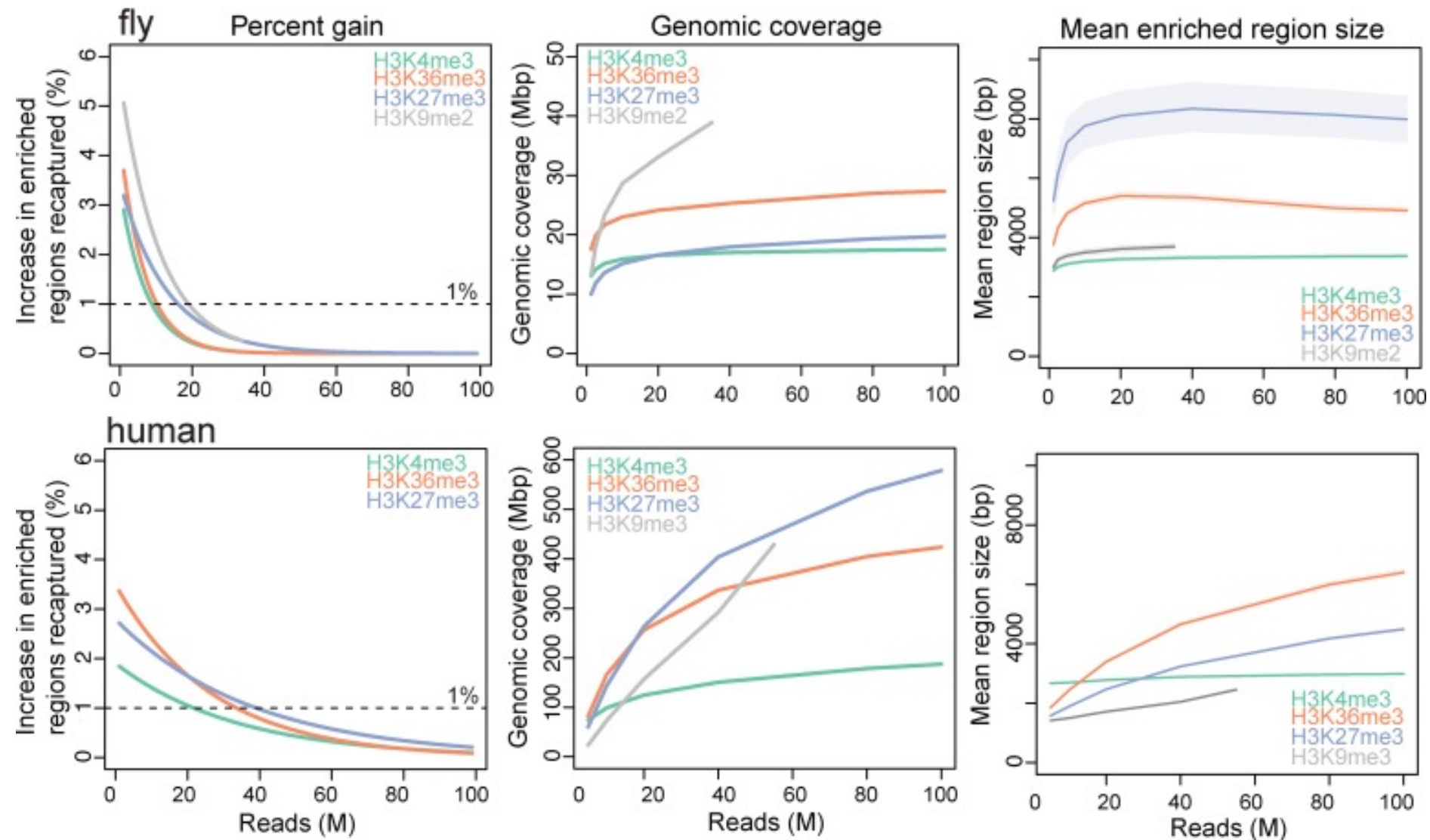detection sensitivity - recovery of true enrichment region

percent increase in enriched regions recaptured when an additional 1 million reads are sequenced
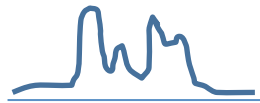
'sufficient depth' - the sequencing depth at which the percent gain per 1 million additional sequence reads falls below 1%

Jung et al, NAR 2014
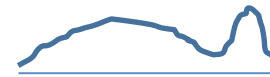
Effect of sequencing depth on site recovery

Sequencing depth

Chromatin
Remodellers

Histone marks

Chromatin
Remodellers

Histone marks

RNA polymerase II

mixed signal

broad signal

Human:
3.2Gb
Drosophila:
300Mb

H3K36me3: 25 M

H3K36me3: 11 M

H3K27me3: 35 M

H3K27me3: 20M

H3K9me3: >55 M

H3K9me3: 20 M

No clear guidelines for mixed and broad type of peaks

Source: The ENCODE consortium; Jung et al, NAR 2014

Peak calling

Chromatin
Remodellers

Histone marks

Chromatin
Remodellers

Histone marks
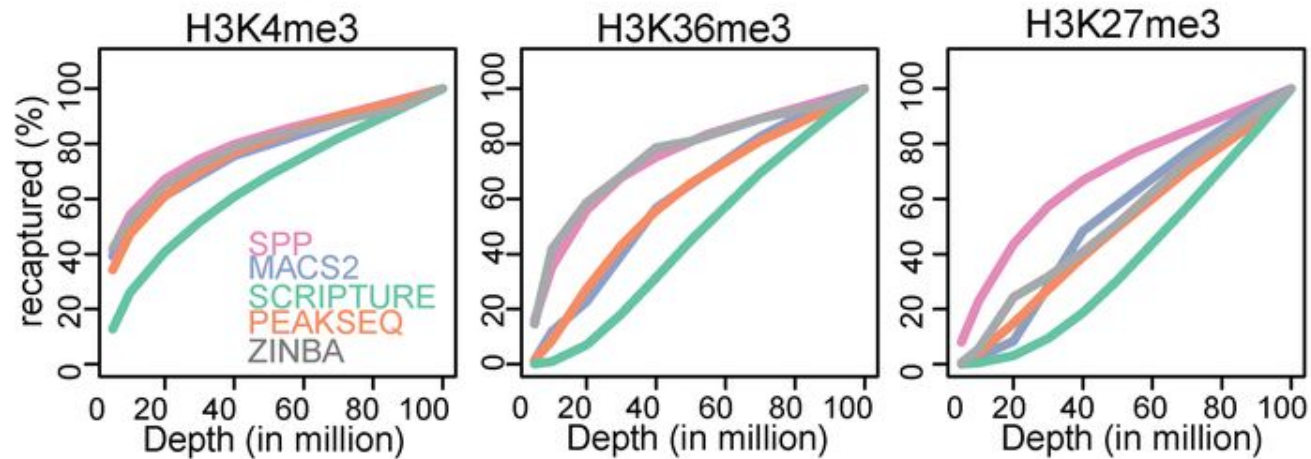
RNA polymerase II

mixed signal

broad signal

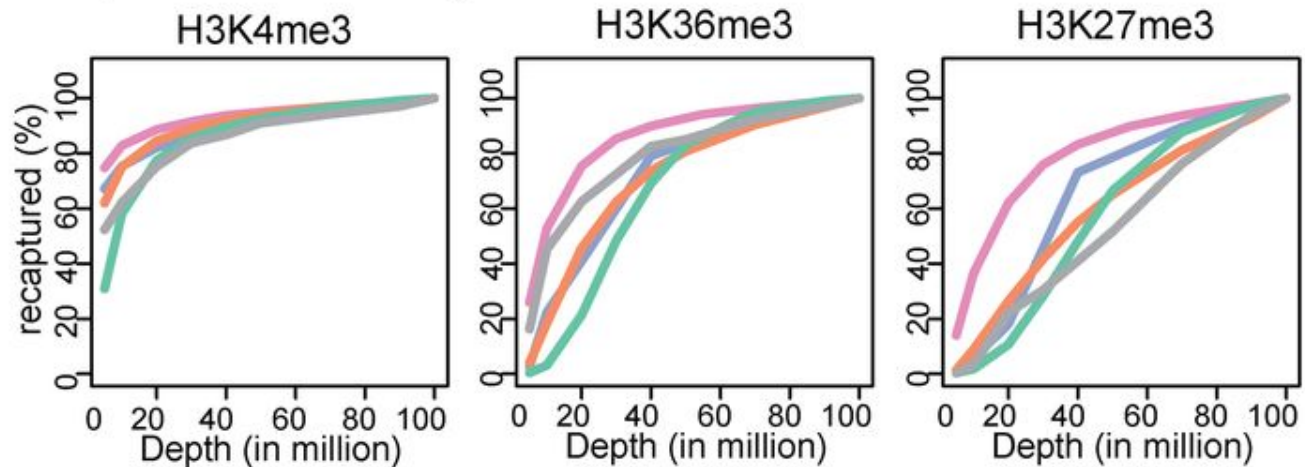- MACS2 in broad mode
- moving windows (csaw)

Effect of sequencing depth on regions detected by various algorithms
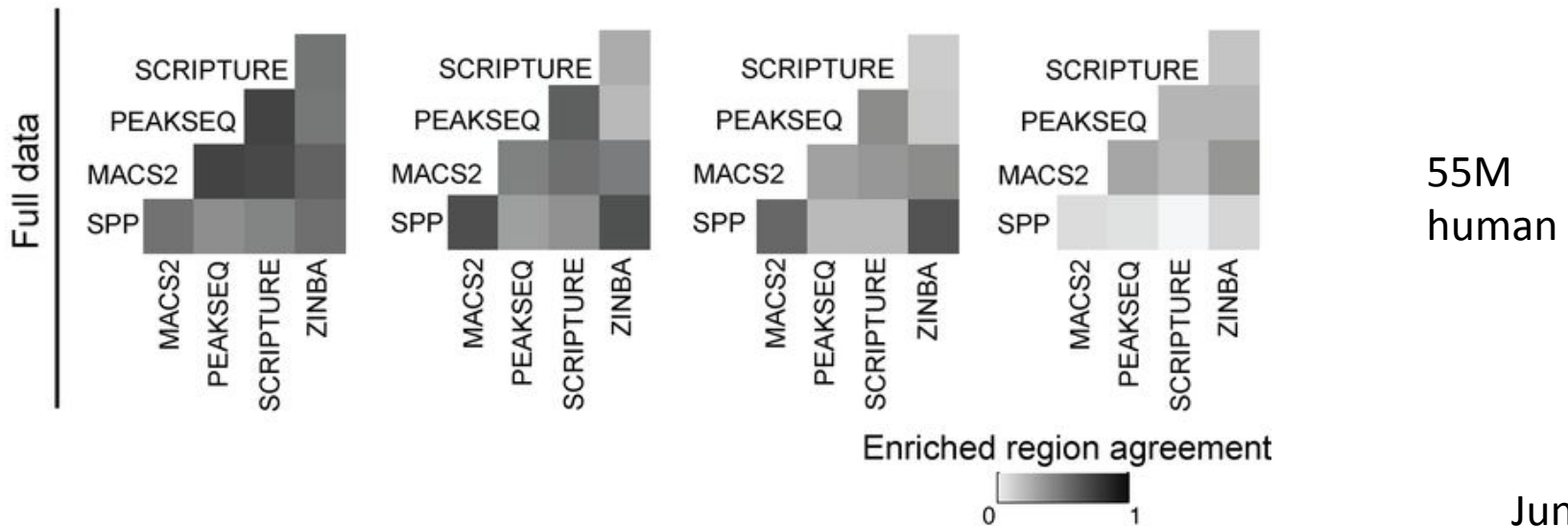

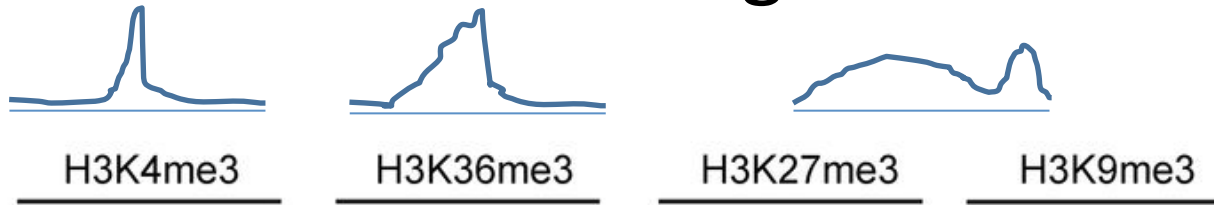**b** Percent of recaptured enriched regions
All enriched regions

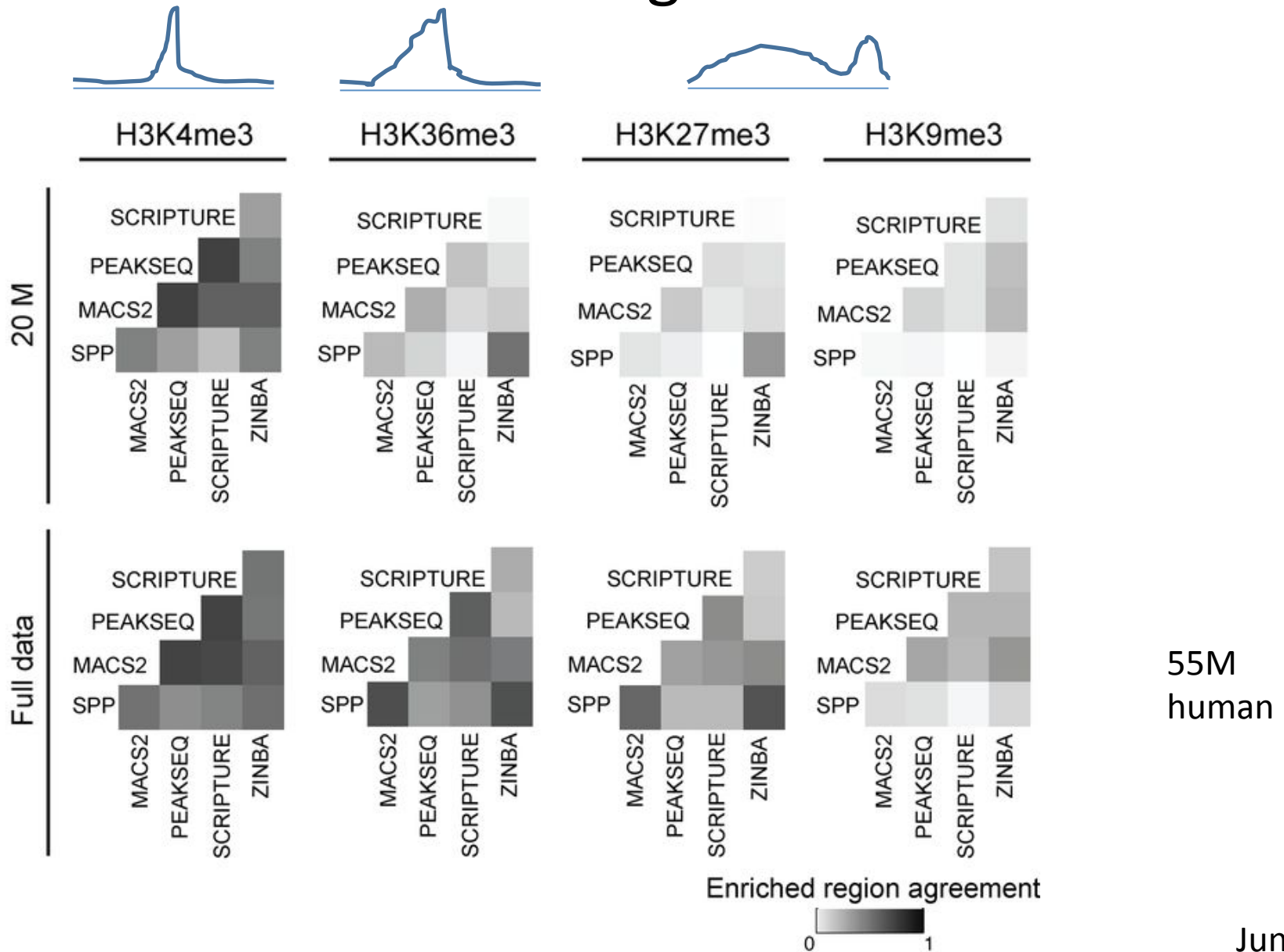Top 20% enriched regions

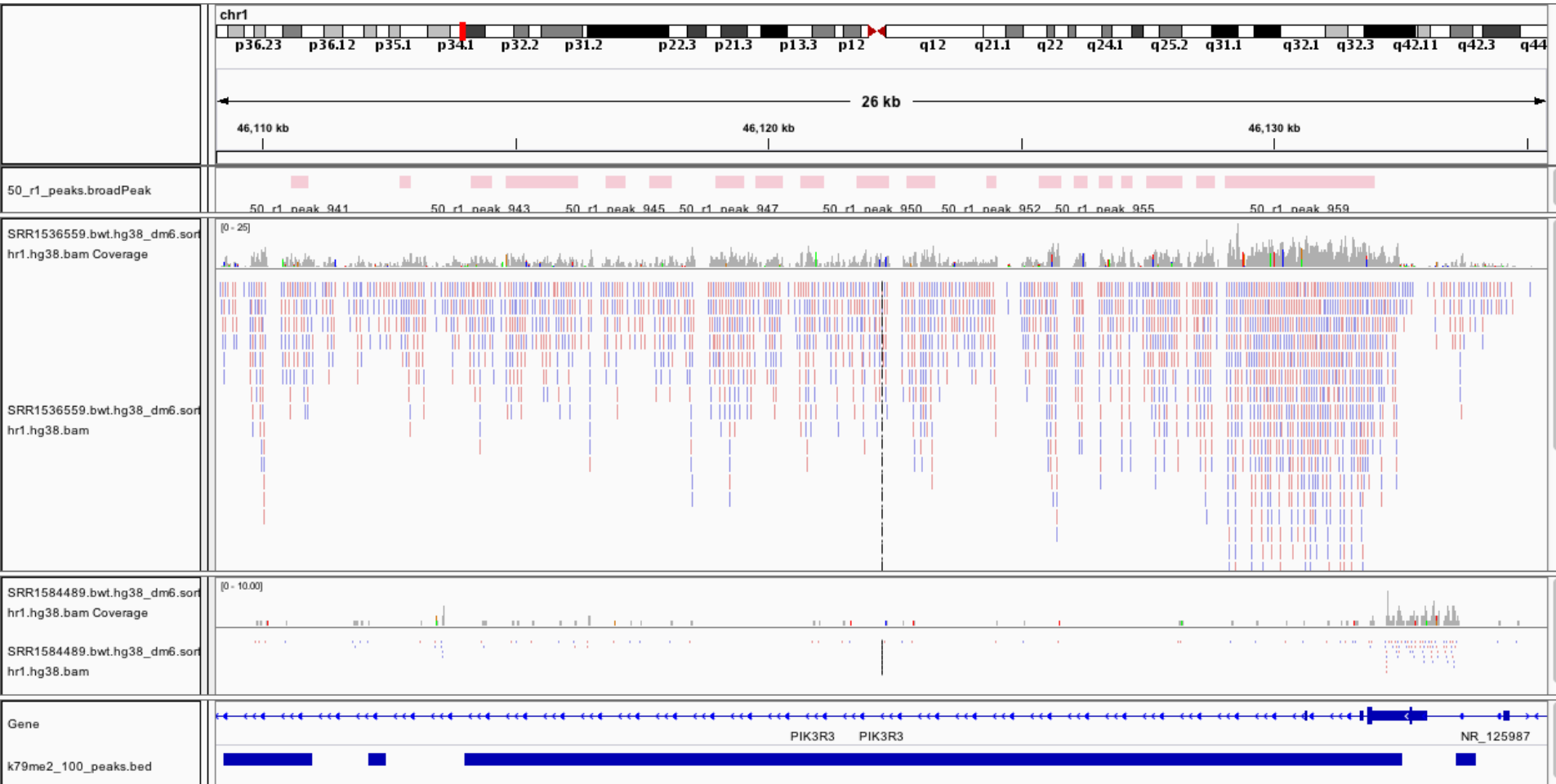# Comparison of enriched regions detected by various algorithms

55M human

Jung 2014

Comparison of enriched regions detected by various algorithms

Enriched region agreement

Jung 2014

Exercise

- MACS2 in `broad peak` mode;

- csaw: Detection of differentially bound regions in ChIP-seq data with sliding windows, with methods for normalization and proper FDR control;

- Low sequencing depth data; H3K79me2

MACS2 vs. csaw

Resources for broad region analysis

- https://omictools.com/peak-calling-category

- https://www.encodeproject.org/chip-seq/histone/