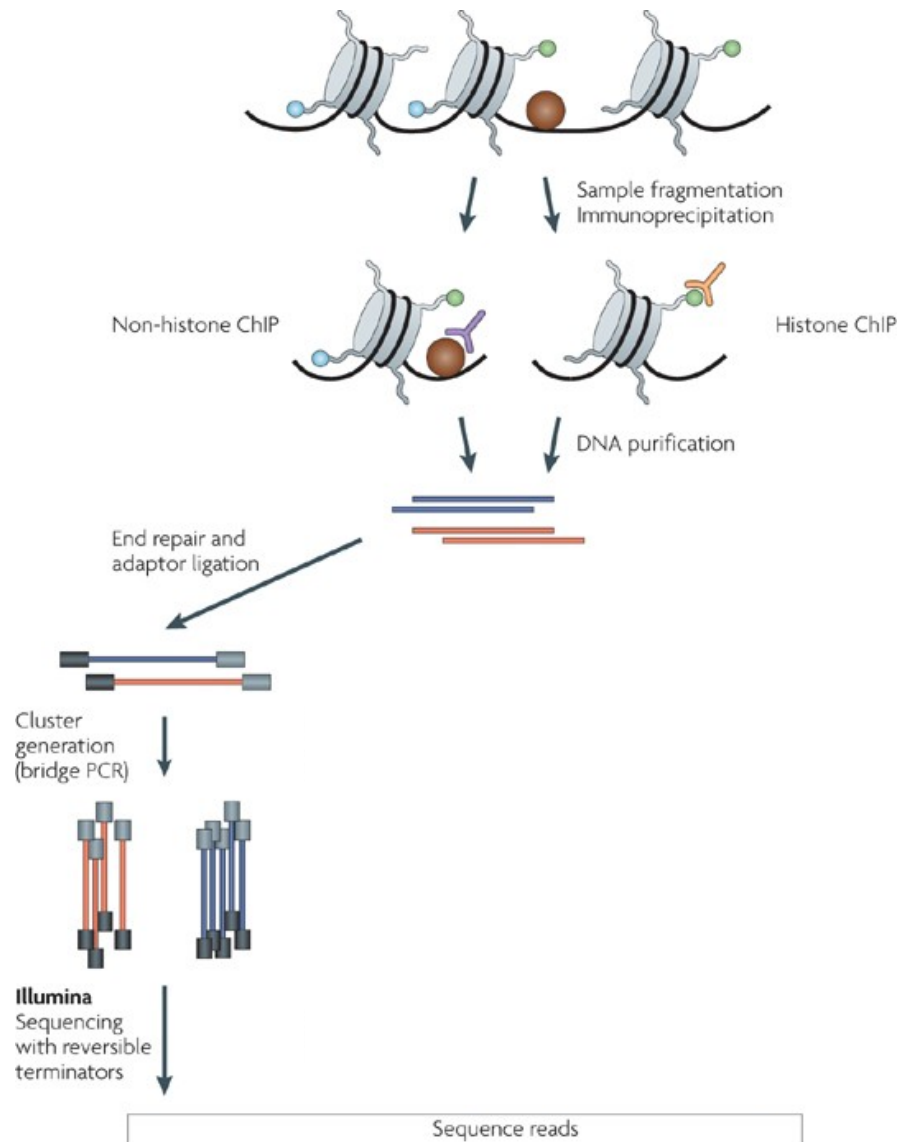


# Introduction to ChIP-sequencing

Simon van Heeringen  
November 4, 2014

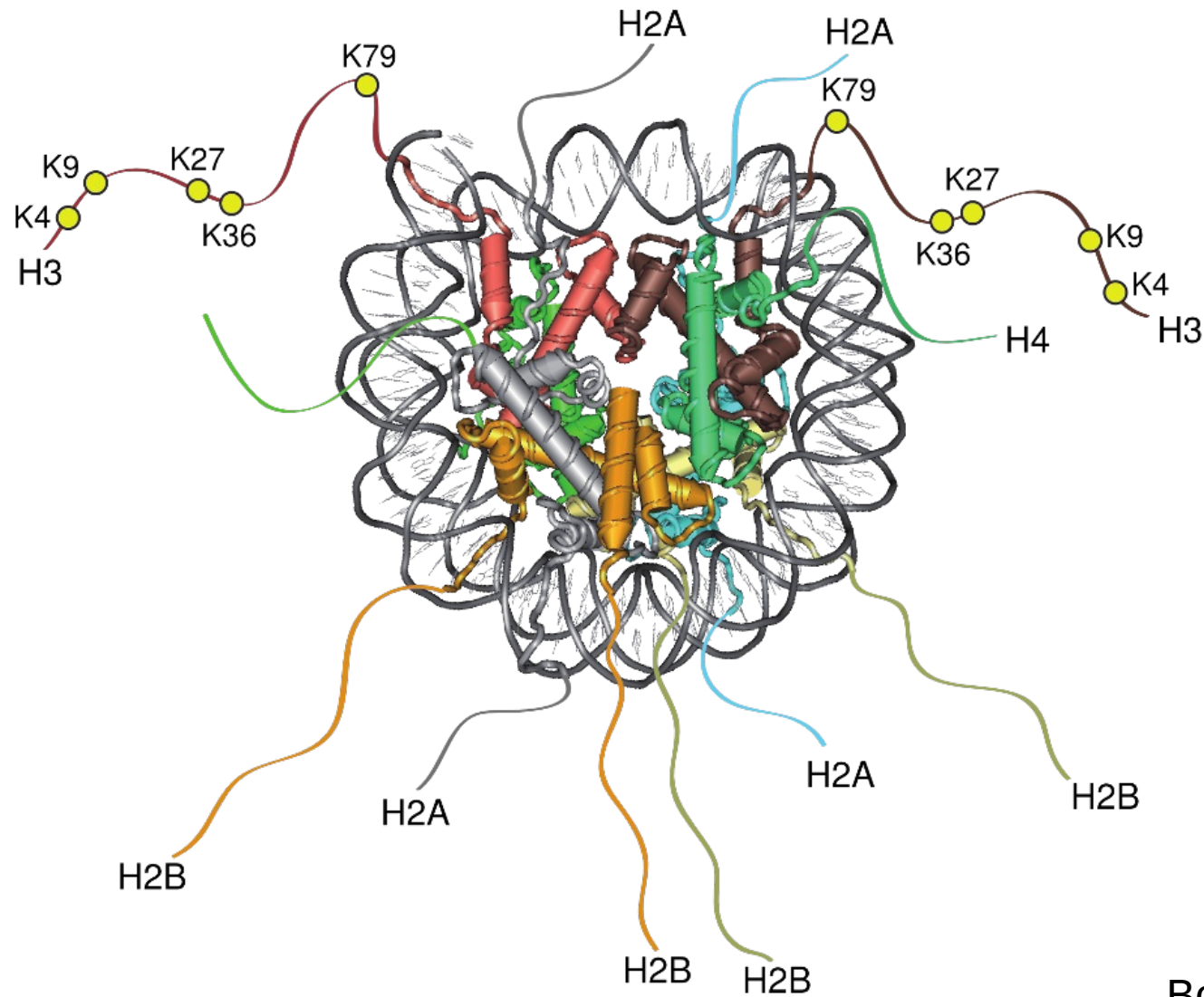
# A basic ChIP-seq experiment



- Cross-link protein and DNA (formaldehyde)
- Fragmentation (sonication)
- IP with antibody
- Reverse crosslinks
- Amplification, library preparation
- Sequencing

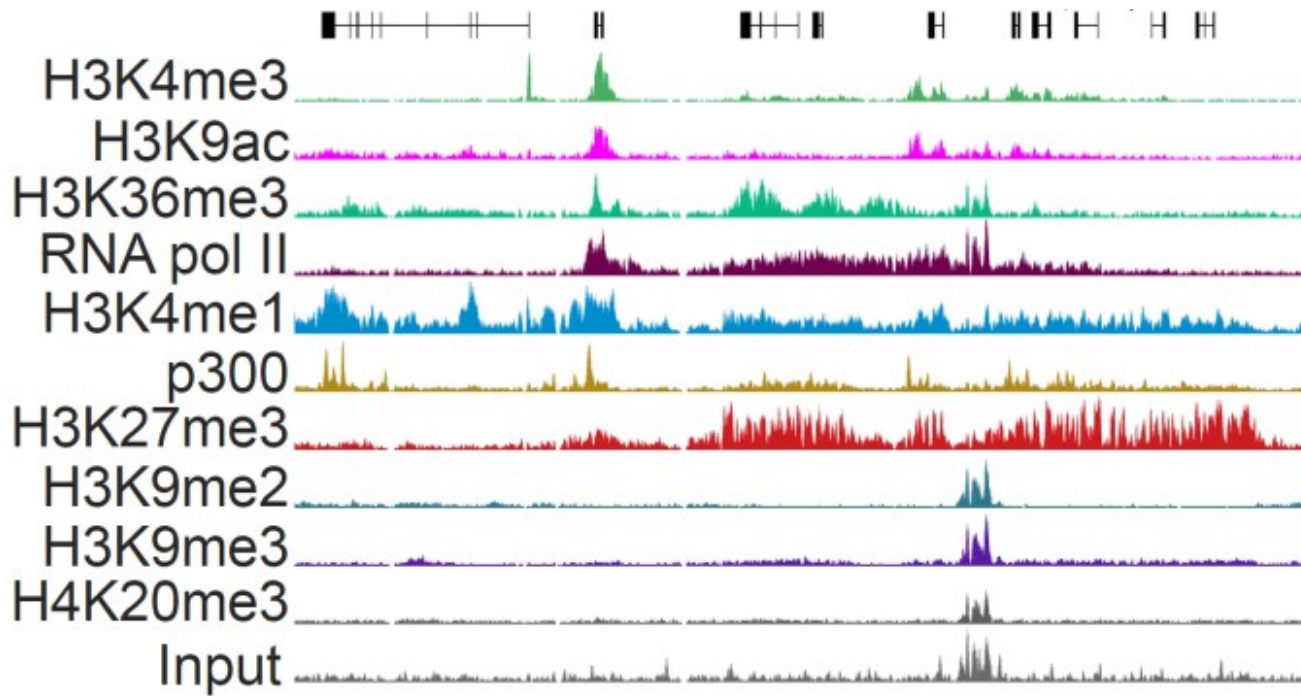
Adapted from Park, 2009

# Histone modifications



Bogdanovic, 2012

# Epigenome



CAAAATACTGATATATACAACATGAACGAATGTCAGACAGTACATTGAAGGACAGAAGCCCGACAAAAATGAGCACATAATGTATGATTCCCC  
GTGTAGTGGCACGATCTTGGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCACCTCCCGAATAGCTGGGATTACAG  
GGGGATTACACGTTGGCCACGCTGGTCTGGAACCTCCTATCCTCAAGTAATCCGCCCGCCTCGGCCTCCCAAAGTGCAGGCGTGAGCCAC  
AAAATGGTTATGGAGATCAAAATAAAGGTGGGGTCGGGAATCGACTGGGAAGAGACGTGATGAAACGTTTCTGGGACGATGAAAAGGGTCTC

# ChIP-seq workflow

Data type

File formats

```
TAACGTGAACCCCTCTATCTTCCTTCACAGATTG
TAGTTTCTCACTTCAAGTTATCCAGCAACCTTGGA
TTTGAACATCATGTTCTGTCATGTTTTGGTGCTTG
GACTGCTAATATCCTTATCATTTACAAAAGGGTAC
GCTTTTTACATTGCGACCACTTAATAAATGACTAG
TCCTTATCCTATGCTCTTATACCCCATATTACTGC
CAGAACAGGAATGAGGGGTCTCTAAATGGCTGATA
CTGCTAAATGTCAATAACTATAATAGCTATGATTT
TGTGGTATTTTATCAAATACATGTTTAAACAAATG
TCCCTATCTTTAAAATCCAGTGCACCTAAAGAATTG
```

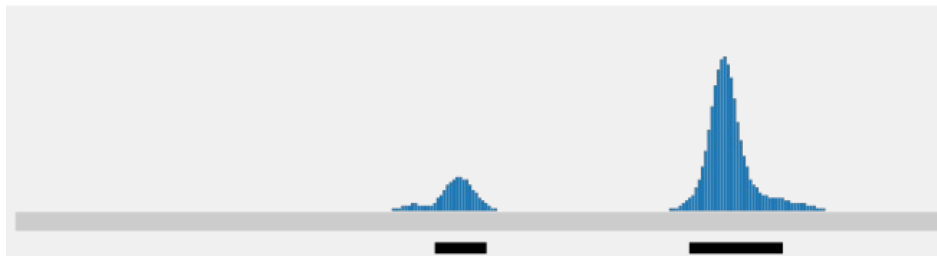
raw  
reads

FASTQ  
SAM / BAM



aligned  
data

BED  
SAM / BAM



visualization,  
peak  
calling

BED  
WIG



# Basic downstream analysis

- Visualization
  - Genome Browser
- Identifying enriched regions (peaks)
- Functional analysis
- Transcription factor motifs

# Peak calling

- Sounds easy, doesn't it?

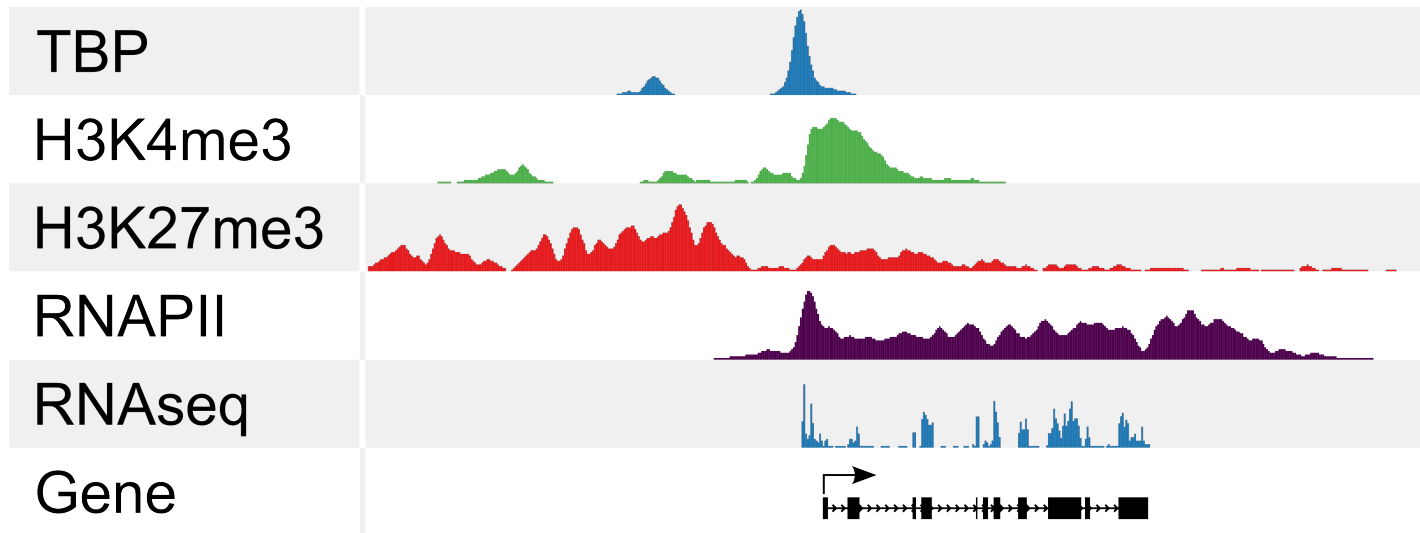
TBF gastrula



Peak!

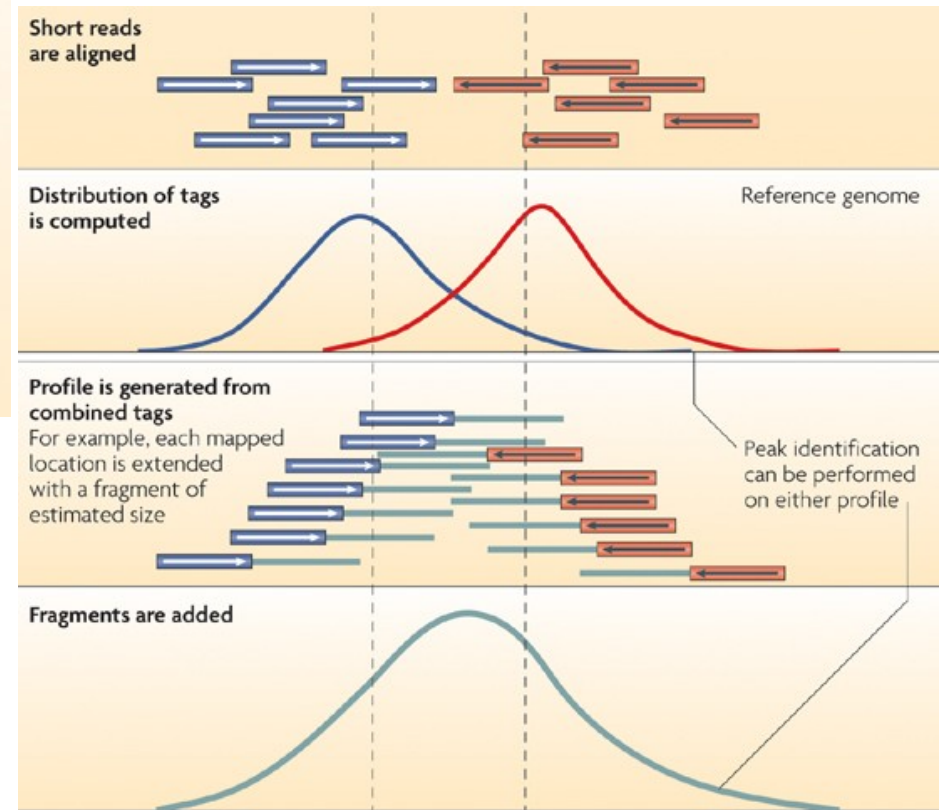
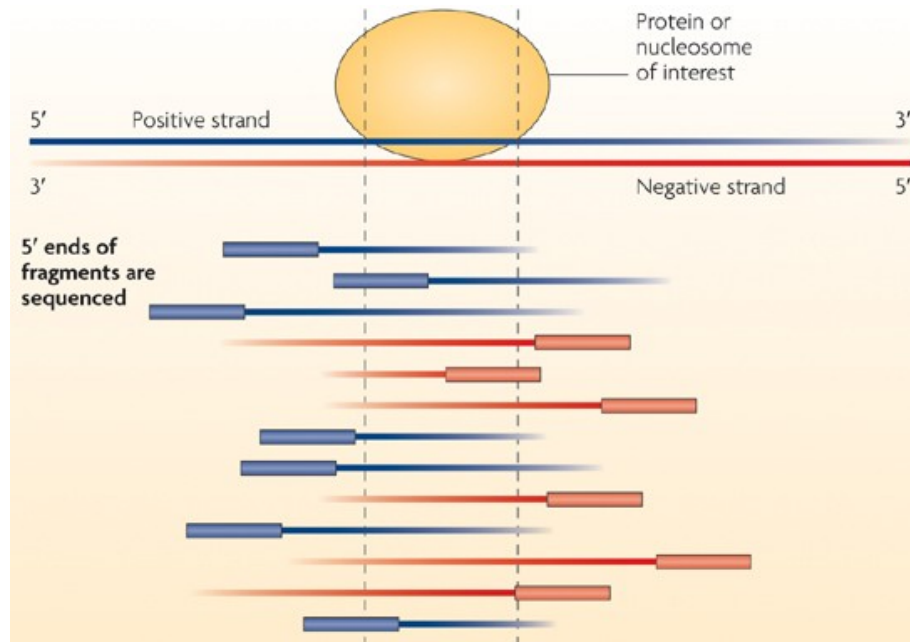
CAAAATACTGATATATACAACATGAACGAATGTCAGACAGTACATTGAAGGACAGAAGCCCGACAAAAATGAGCACATAATGTATGATTCCCC  
GTGTAGTGGCACGATCTTGGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCACCTCCCGAATAGCTGGGATTACAG  
GGGGATTACACGTTGGCCACGCTGGTCTGGAACCTCCTATCCTCAAGTAATCCGCCCCGCCTCGGCCTCCCAAAGTGCAGGCGTGAGCCAC  
AAAATGGTTATGGAGATCAAAATAAAGGTGGGGTCGGGAATCGACTGGGAAGAGACGTGATGAAACGTTTCTGGGACGATGAAAAGGGTCTC

# Peak calling





# Peak calling



Park, 2009

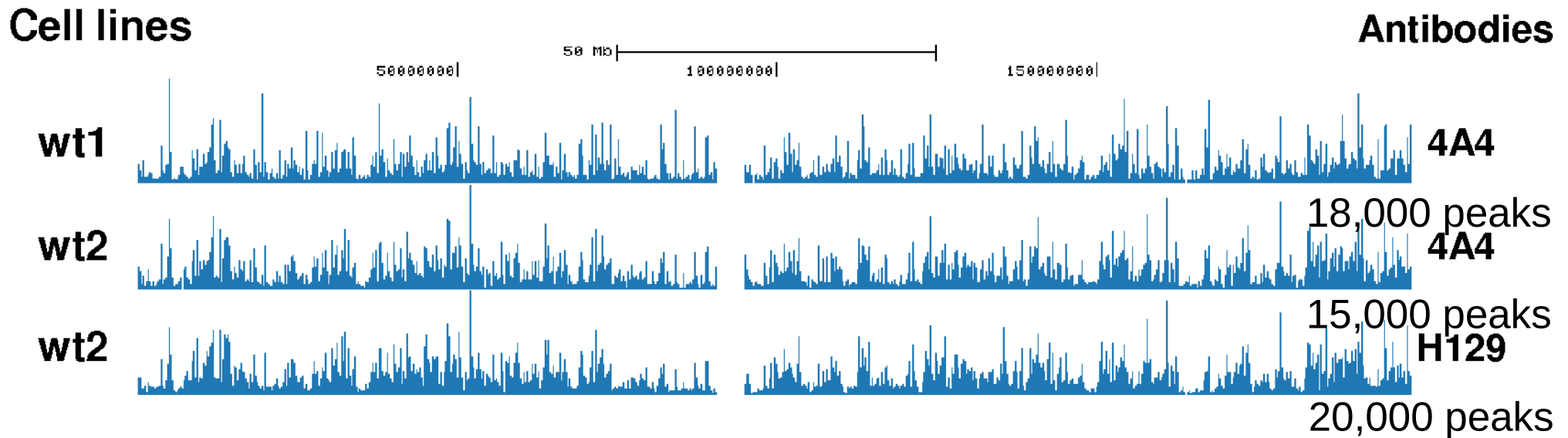
# ENCODE guidelines

- Antibody specificity
- At least 2 biological replicates
- Control per cell line / stage / condition
  - IgG, Input DNA
- Read depth
  - “Point” versus “broad”

Landt, 2012

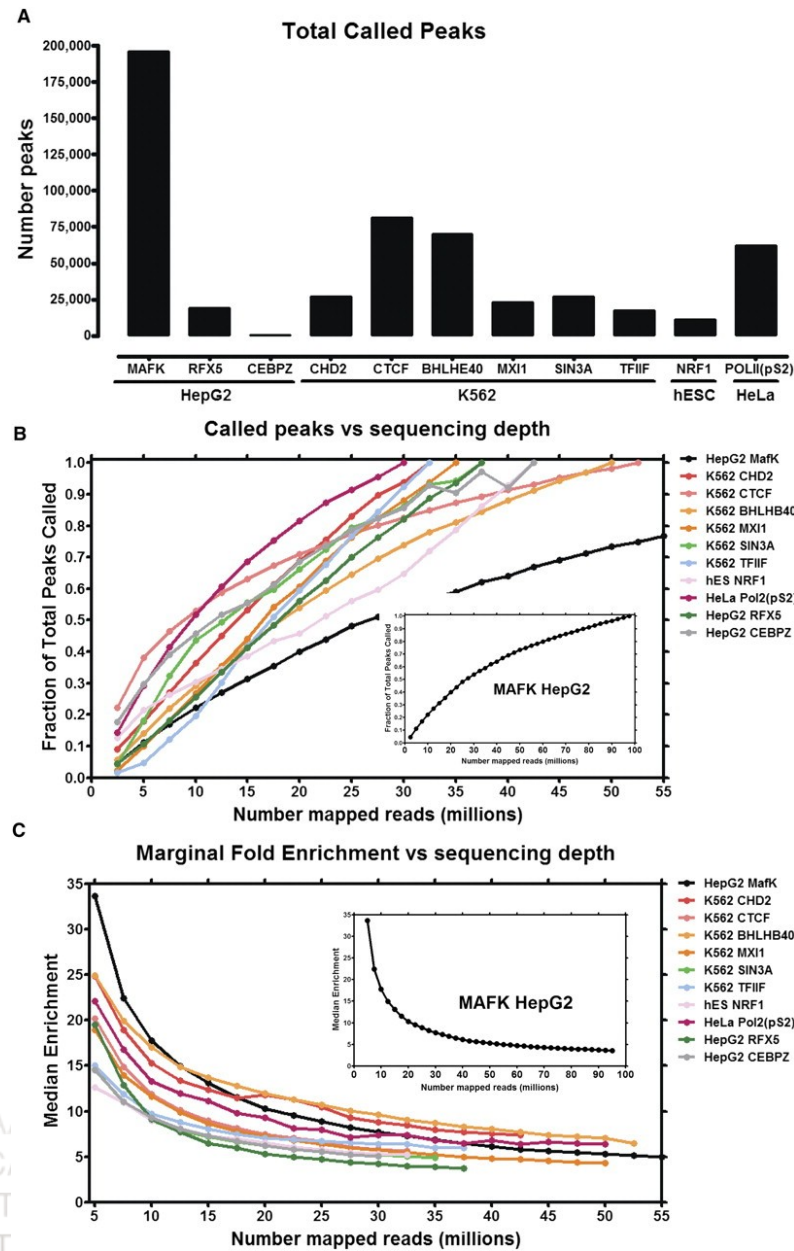
# Experimental design: Replicates

- ChIP-seq of p63 in primary human keratinocytes



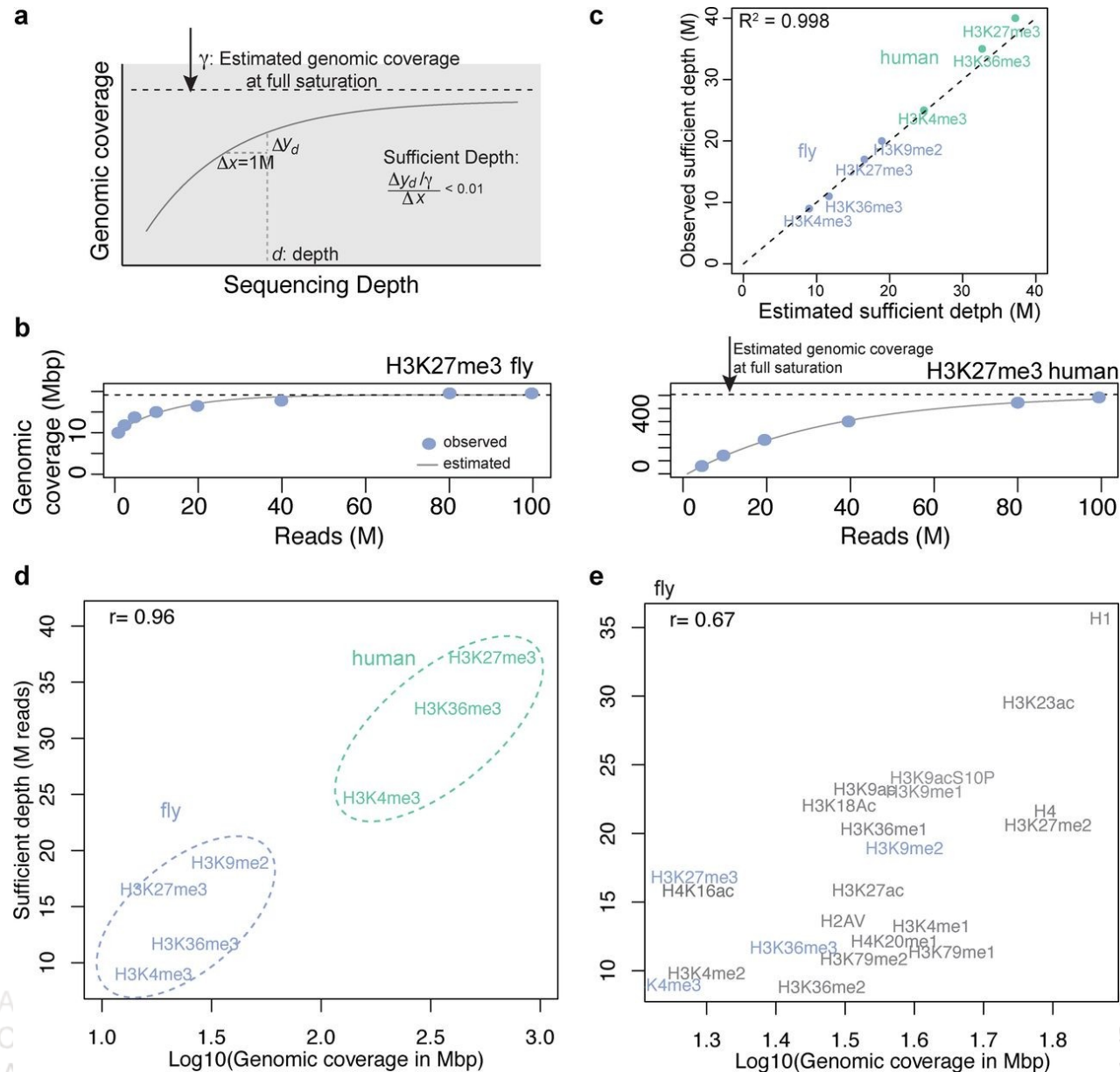
- ~20,000 peaks, ~11,000 shared
- 99% of the peaks were present in third replicate!

# Read depth, saturation



Landt, 2012

# Read depth, saturation



Jung, 2014

# Server IP addresses

- 54.80.155.122
- 54.162.253.10
- 54.167.247.172
- 54.205.72.202
- 54.82.190.179
- 54.162.100.112
- 54.242.87.148