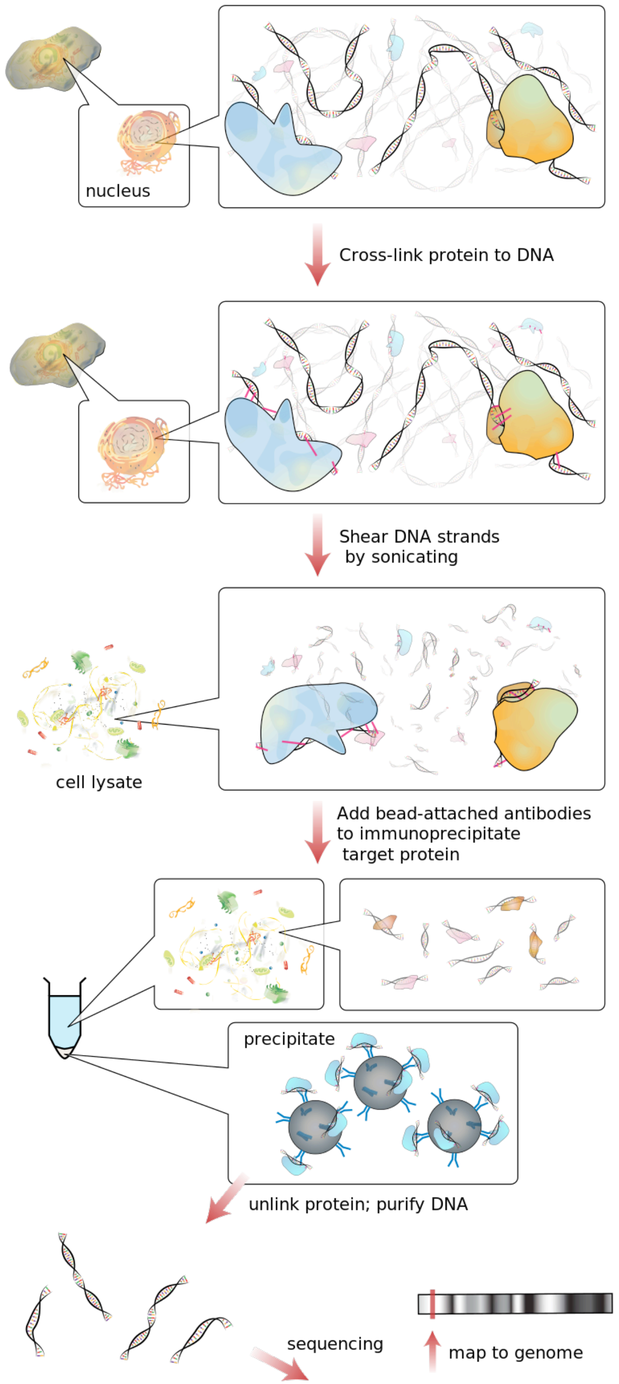
# Calibrated Chip-Seq

## (Quantitative)

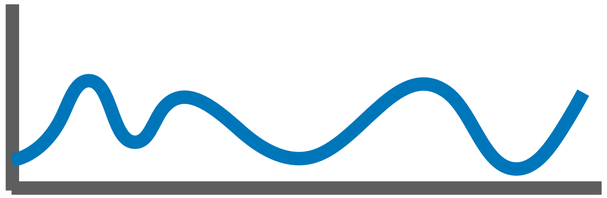
**Laurent Modolo 2021/04/19**

# Chip-Seq

1. cross-linking
2. cell breakage
3. DNA sonication
4. immunoprecipitation (for a given number of epitopes)
5. persistence of DNAs within the immunoprecipitates during the washing procedure
6. release from the immunoprecipitation beads
7. de-crosslinking
8. subsequent DNA purification
9. library construction
10. library amplification by PCR
11. sequencing reaction.



# Chip-Seq

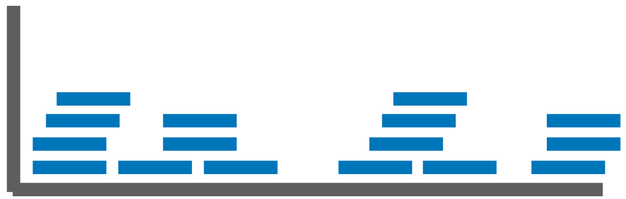


**O**ccupancy, **O**(t) is the probability that the IP protein is at position t

**N**umber of cells

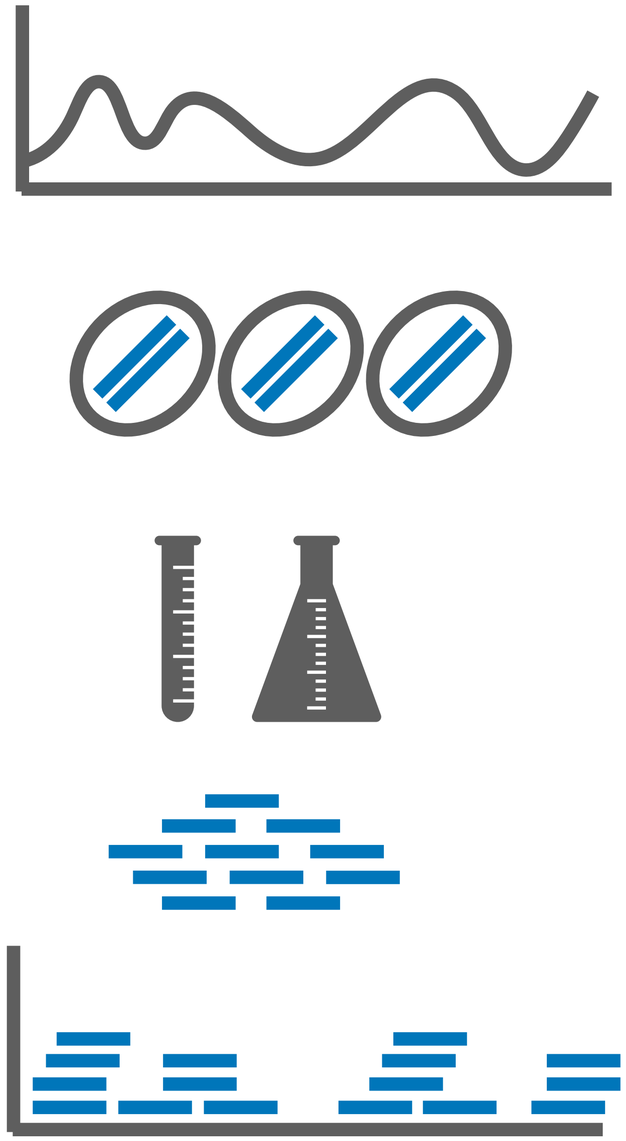
**E**fficiency

**IP** reads number



**IP**(t) number of reads at position t

# Chip-Seq



**O**ccupancy

**N**umber of cells

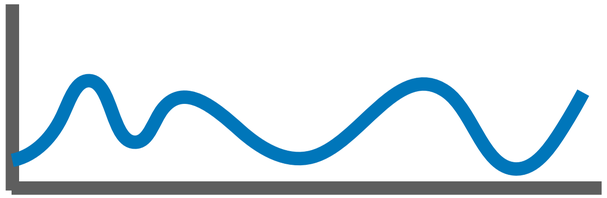
**E**fficiency

**IP** reads number

1. cross-linking
2. cell breakage
3. DNA sonication
4. immunoprecipitation (for a given number of epitopes)
5. persistence of DNAs within the immunoprecipitates during the washing procedure
6. release from the immunoprecipitation beads
7. de-crosslinking
8. subsequent DNA purification
9. library construction
10. library amplification by PCR
11. sequencing reaction.

**IP**(t) number of reads at position t

# Chip-Seq



**O**ccupancy

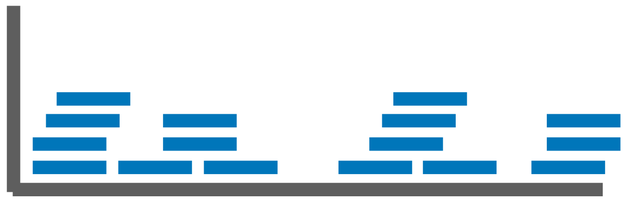
**N**umber of cells

**E**fficiency

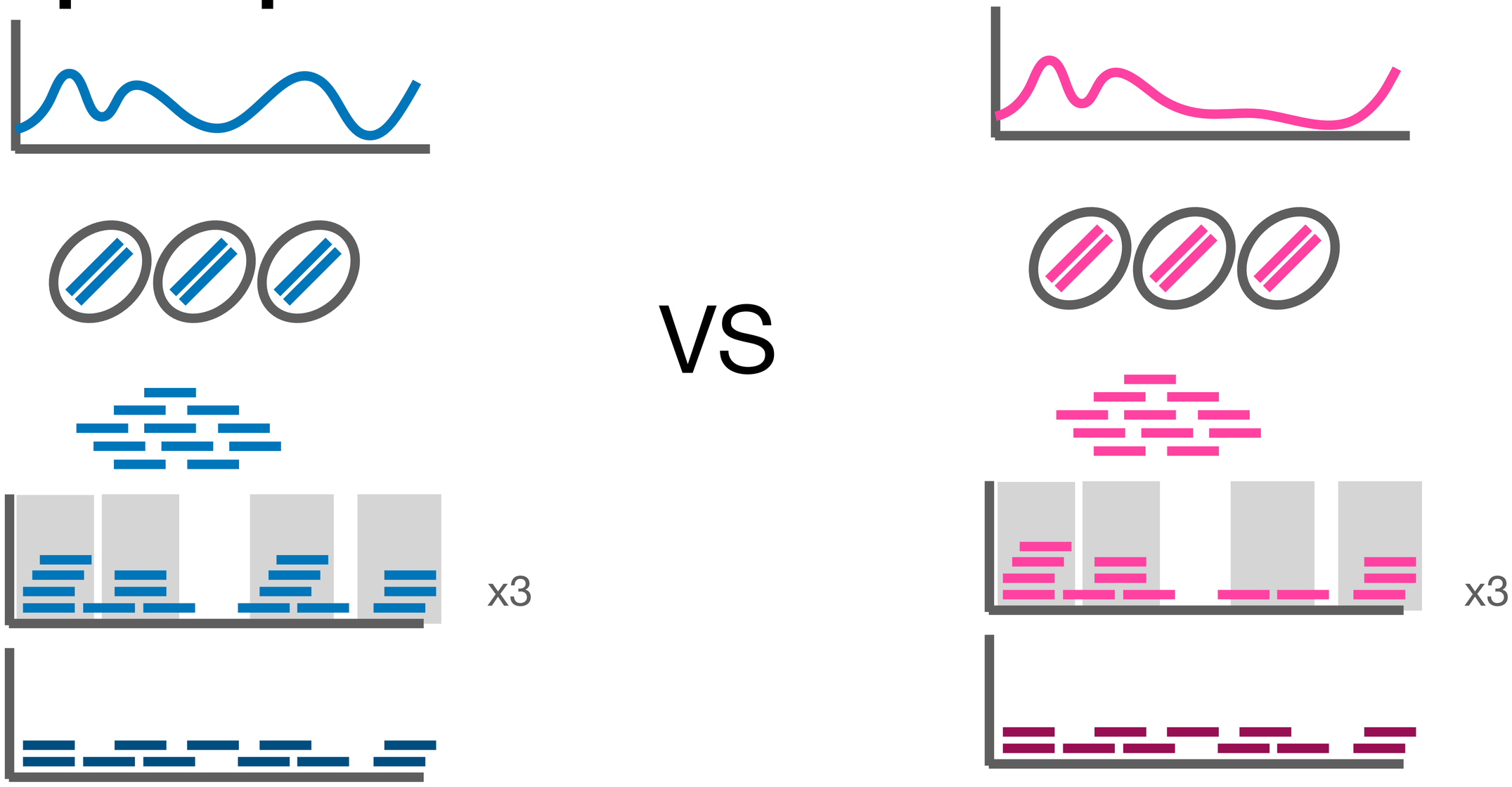
Without IP

**IP** reads number

**W**hole **C**ell **E**xtract

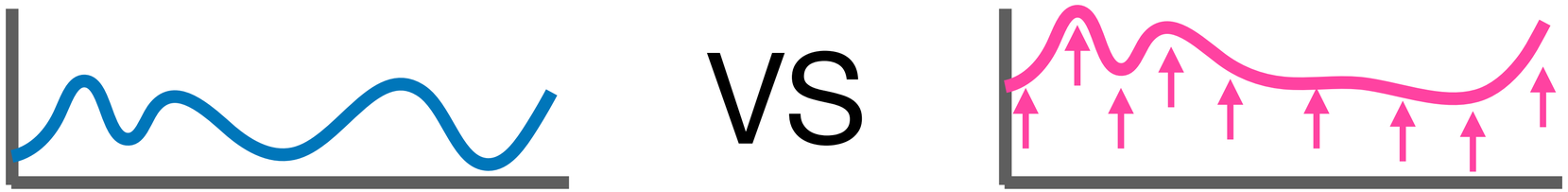


# Chip-Seq



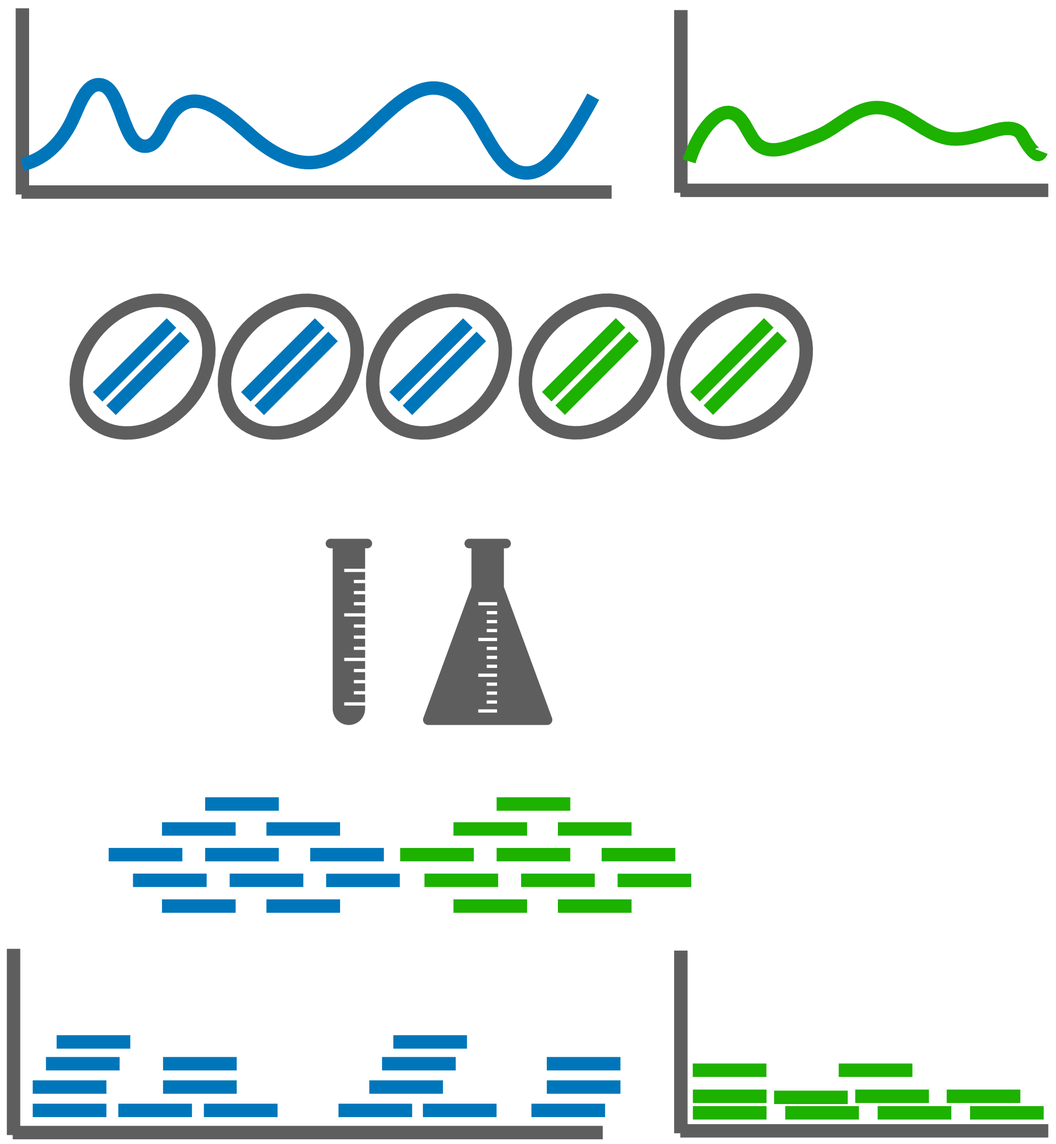
# Chip-Seq limitation

- Formaldehyde cross-links single stranded DNA to proteins much more efficiently than it does double stranded DNA.
- ChIP-seq merely reveals an estimate of the distribution of a protein across a genome.
- If occupancy were reduced or increased at all loci throughout the genome in a similar manner, then conventional ChIP-seq would not reveal any change.




Hu et. All NAR 2015

# Calibrated Chip-Seq



Experiment Calibration

**O**X **O**C

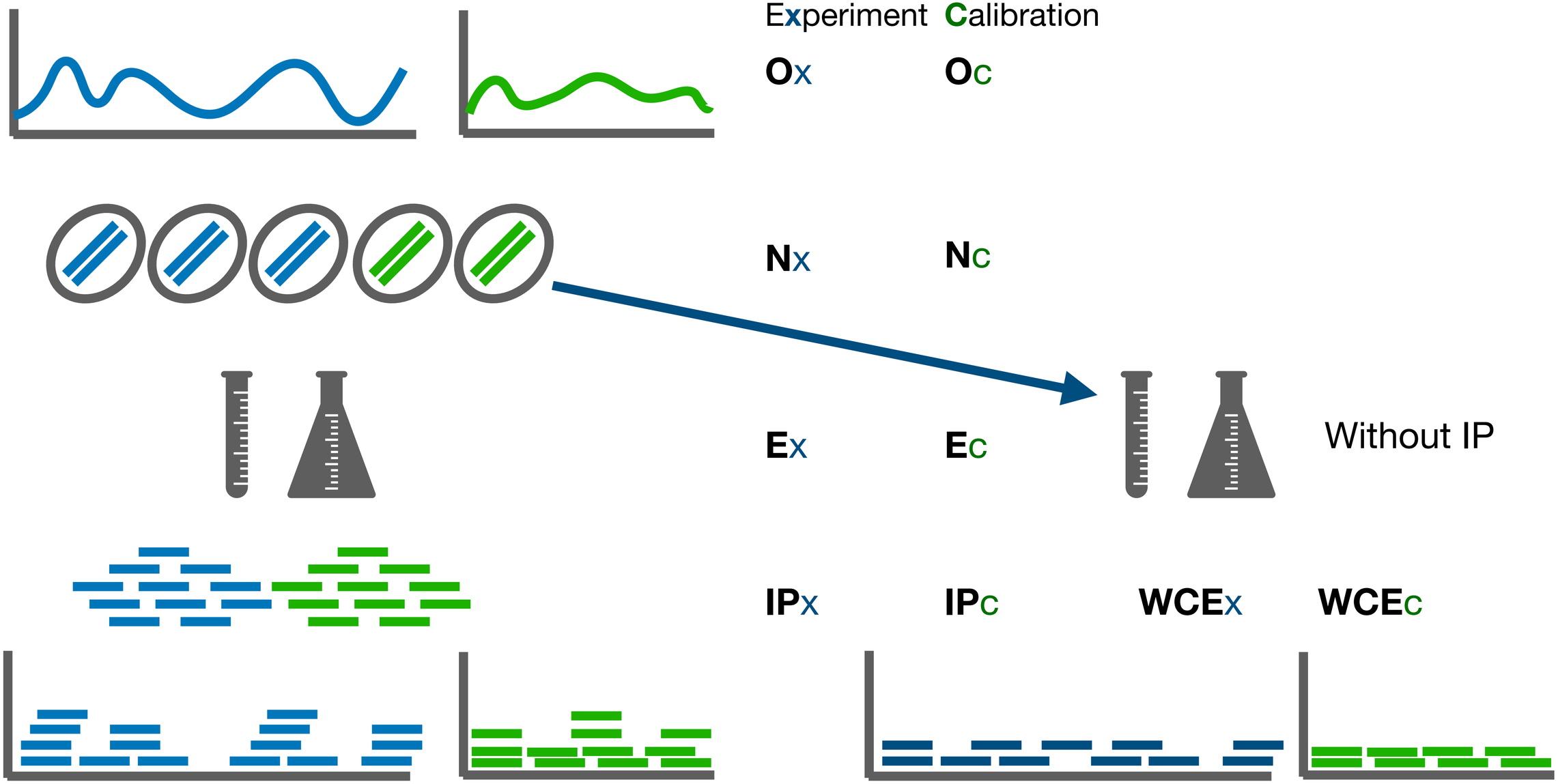
**N**X **N**C

**E**X **E**C

**IP**X **IP**C

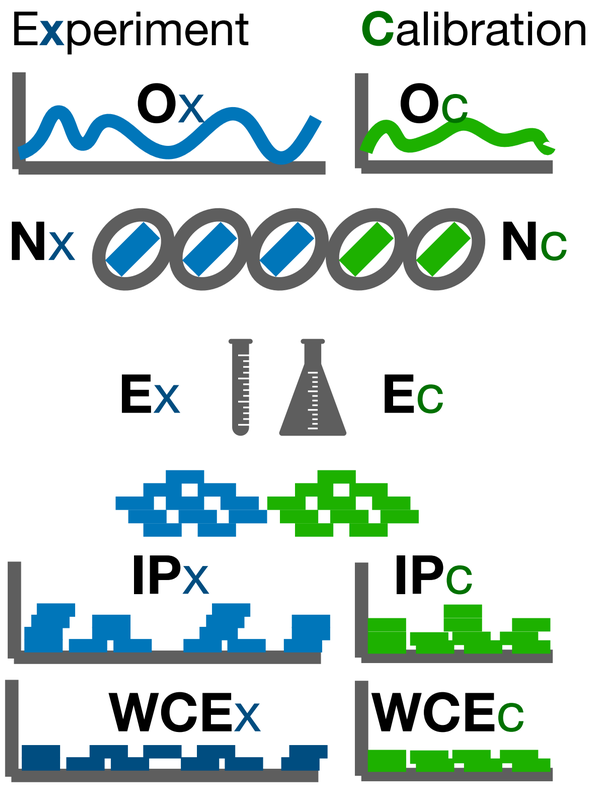
Hu et. All NAR 2015

# Calibrated Chip-Seq




Hu et. All NAR 2015

# Calibrated Chip-Seq




IP = N x O x E

$$F_x = \frac{IP_x}{IP_x + IP_c} = \frac{N_x O_x E_x}{N_x O_x E_x + N_c O_c E_c}$$

$$\frac{O_x E_x}{O_c E_c} = \frac{N_c Fx}{N_x(1 - F_x)}$$

$$\alpha = O_c \frac{E_c}{E_x}$$

$$O_x = \alpha \frac{N_c F_x}{N_x(1 - F_x)}$$

$$OR_x = \frac{N_c F_x}{N_x(1 - F_x)}$$

The **O**ccupancy **R**atio

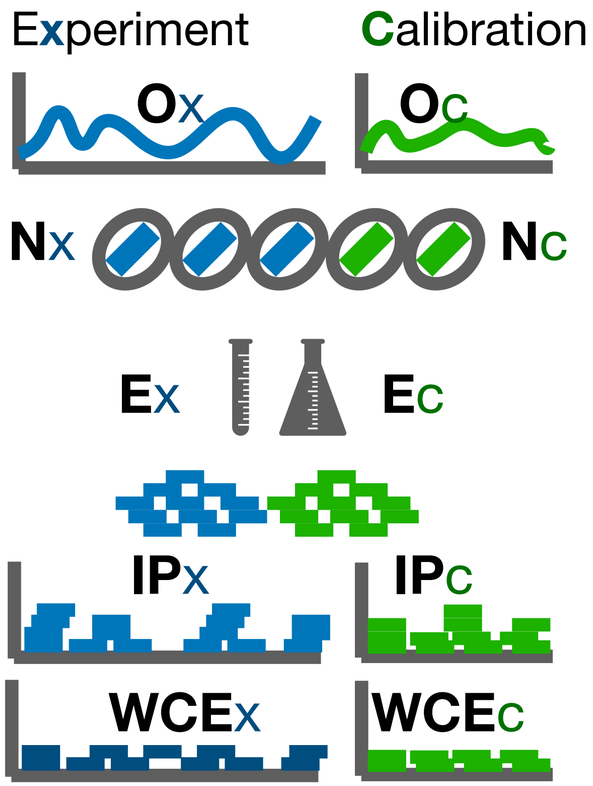
$$O = \frac{1}{T}\sum_{t=1}^{T} O(t)$$

$$IP = \sum_{t=1}^{T} IP(t)$$

$$WCE = \sum_{t=1}^{T} WCE(t)$$

Hu et. All NAR 2015

# Calibrated Chip-Seq




IP = N x O x E

$$F_x = \frac{IP_x}{IP_x + IP_c} = \frac{N_x O_x E_x}{N_x O_x E_x + N_c O_c E_c}$$

$$\frac{O_x E_x}{O_c E_c} = \frac{N_c Fx}{N_x(1 - F_x)} \qquad \alpha = O_c \frac{E_c}{E_x}$$

$$O_x = \alpha \frac{N_c F_x}{N_x(1 - F_x)}$$

$$OR_x = \frac{N_c F_x}{N_x(1 - F_x)}$$

The **O**ccupancy **R**atio

$$\widehat{OR}_x = \frac{WCE_c F_x}{WCE_x(1 - F_x)}$$

$$\widehat{OR}_x = \frac{WCE_c IP_x}{WCE_x IP_c}$$

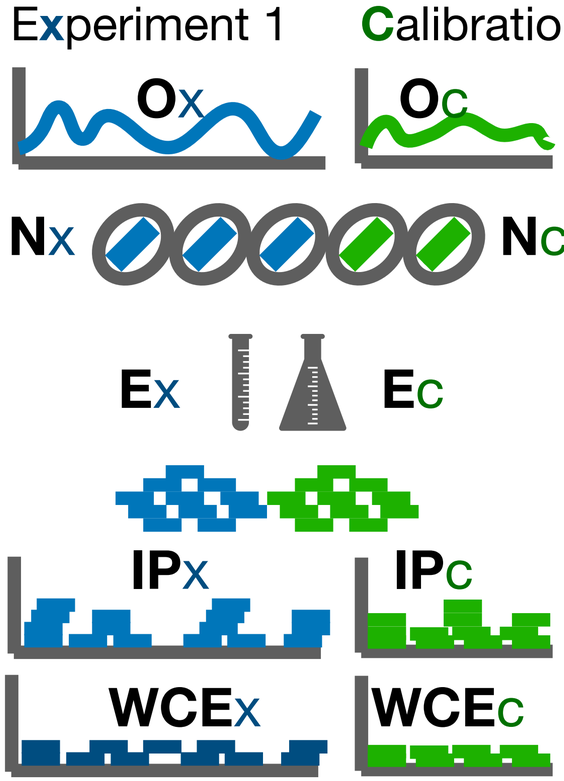
$$O = \frac{1}{T}\sum_{t=1}^{T} O(t)$$

$$IP = \sum_{t=1}^{T} IP(t)$$

$$WCE = \sum_{t=1}^{T} WCE(t)$$

Hu et. All NAR 2015

# Calibrated Chip-Seq

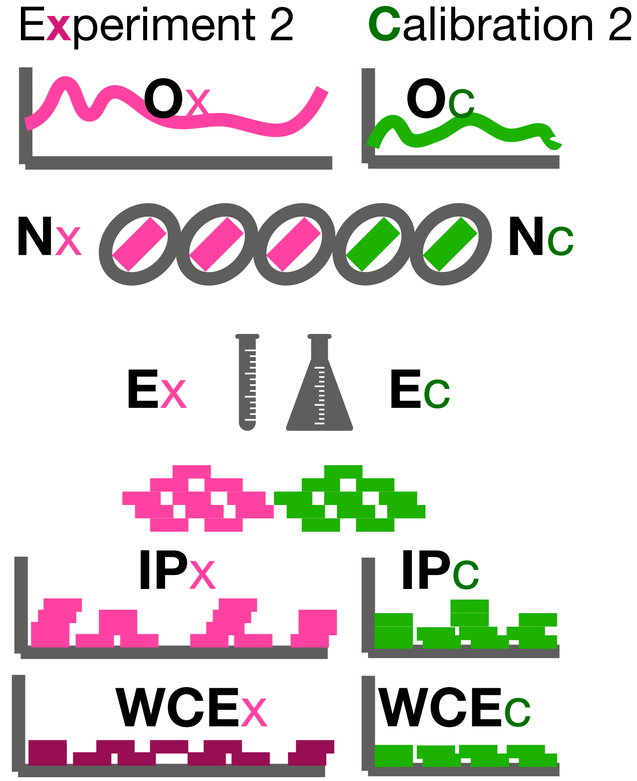



$$O_{c1}\frac{E_{c1}}{E_{x1}} = O_{c2}\frac{E_{c2}}{E_{x2}}$$

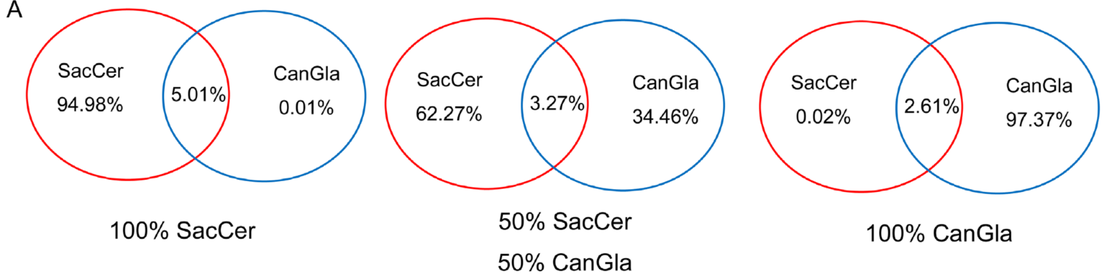
$$\widehat{OR}_{x1} = \frac{WCE_{c1}IP_{x1}}{WCE_{x1}IP_{c1}} \qquad \widehat{OR}_{x2} = \frac{WCE_{c2}IP_{x2}}{WCE_{x2}IP_{c2}}$$

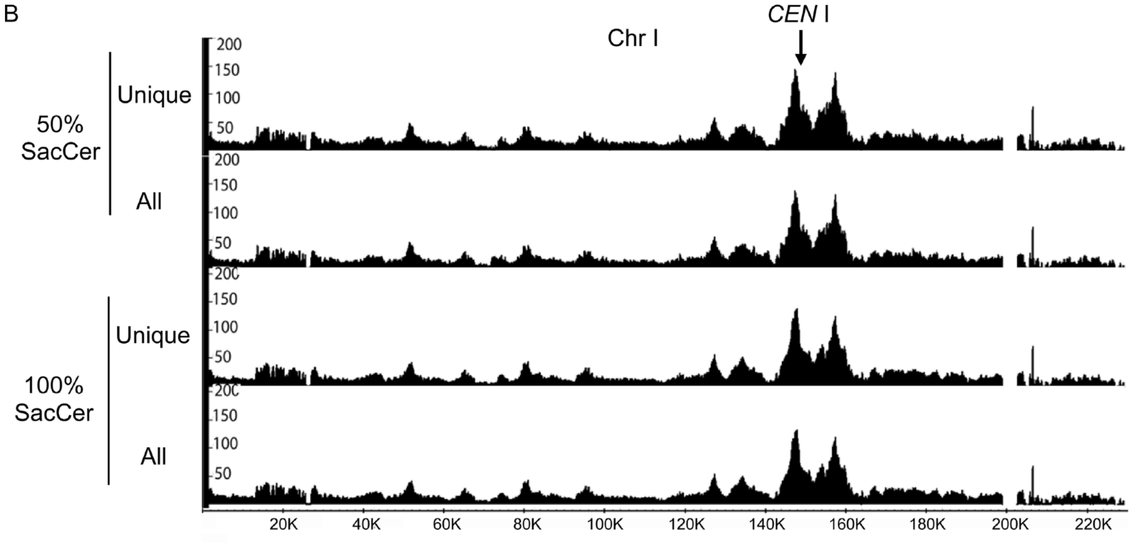
We can compare

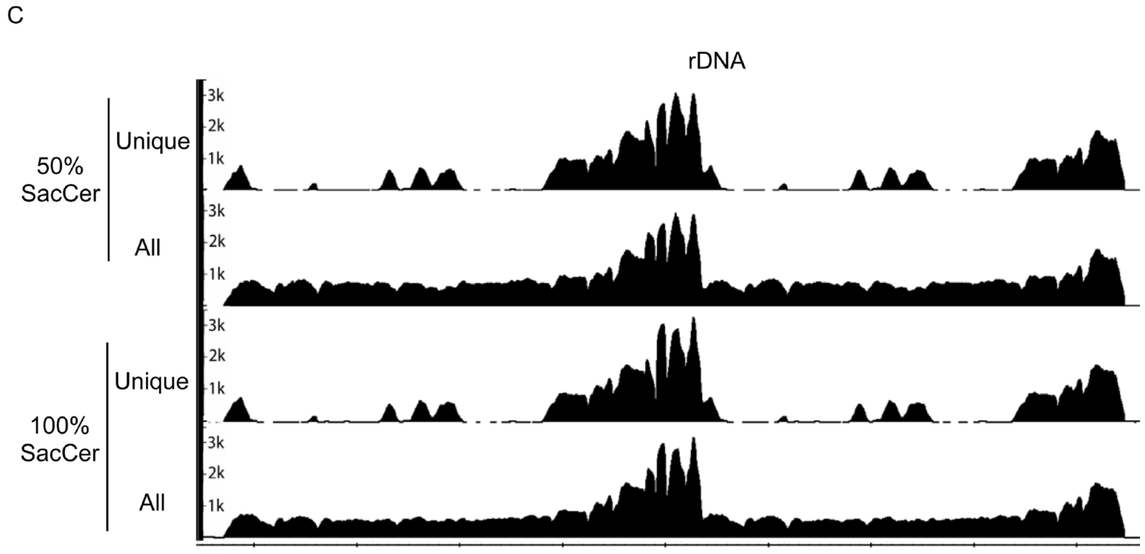
$\widehat{OR}_{x1}IP_{x1}(t)$ and $\widehat{OR}_{x2}IP_{x2}(t)$




Hu et. All NAR 2015



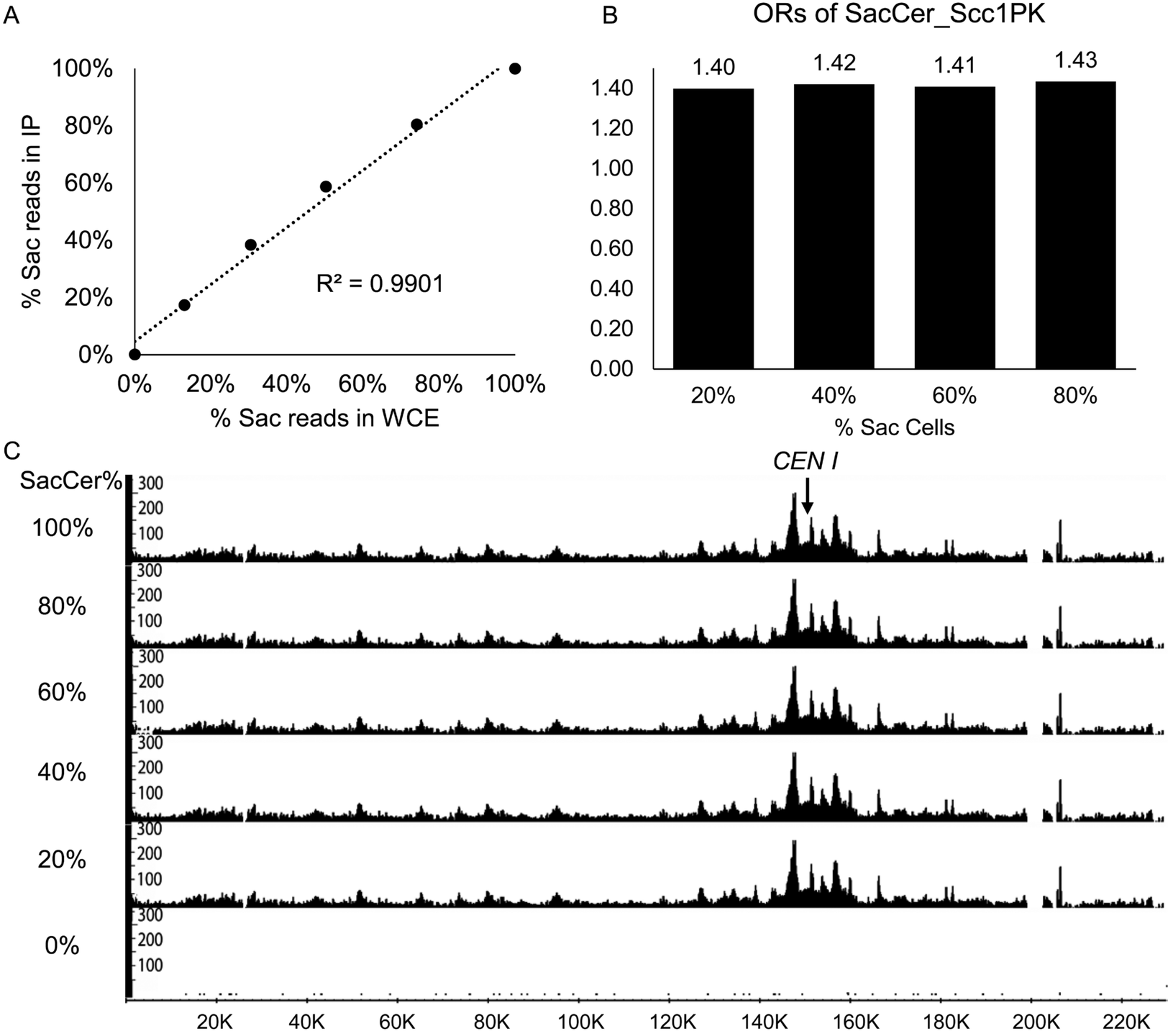




IP of cohesin's Scc1 subunit

- *Saccharomyces cerevisiae* (SacCer)
- *Candida glabrata* (CanGla)

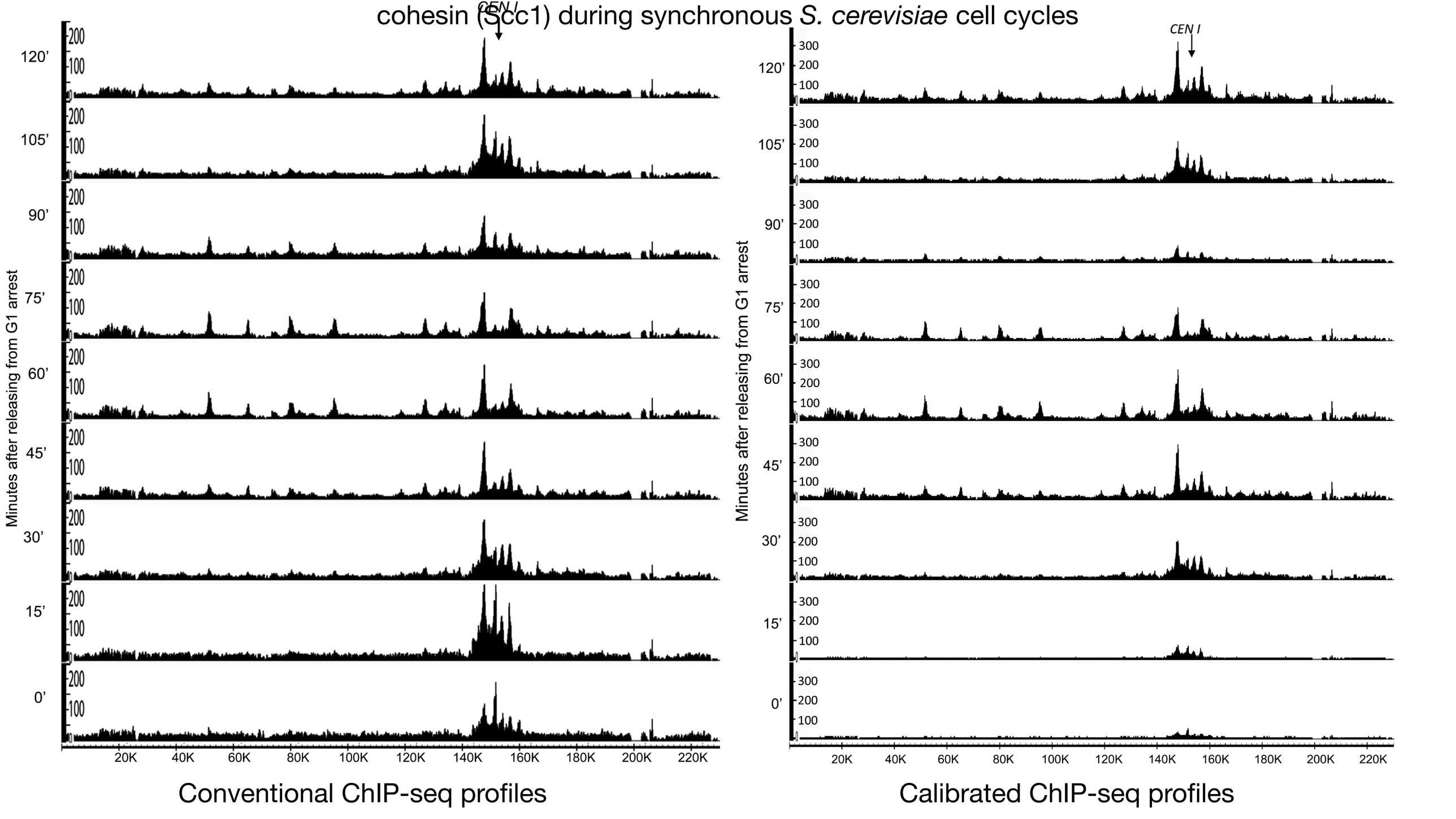
ChIP-seq profiles are unaffected by reference cells.

Hu et. All NAR 2015




- Correlation between percentages of reads aligning to *S. cerevisiae* genome from IP and whole cell extract (W) samples.
- ORs of mixtures with the indicated ratios.
- ChIP-seq distributions of SacCer Scc1 on chromosome I from mixtures with different ratios of *S. cerevisiae* and *C. glabrata* cells.

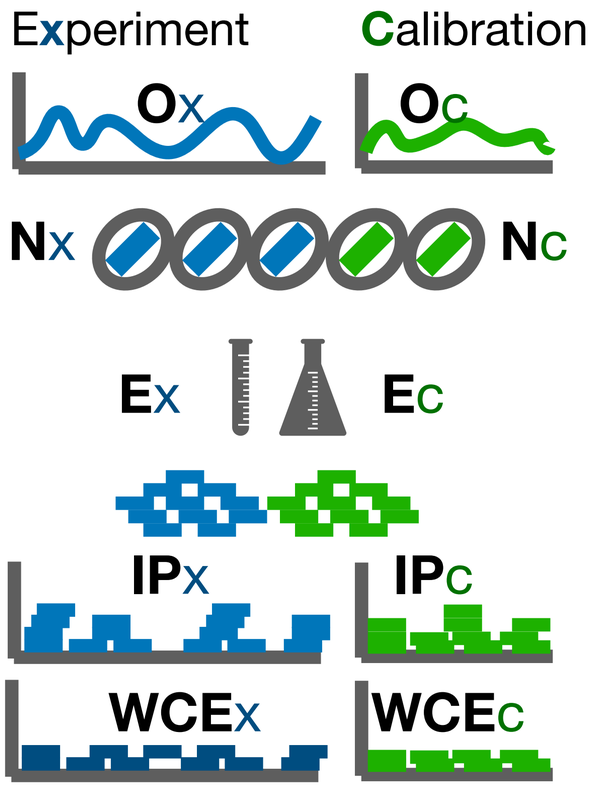
Hu et. All NAR 2015


cohesin (Scc1) during synchronous S. cerevisiae cell cycles
CEN I
CEN I
Minutes after releasing from G1 arrest
120'
105'
90'
75'
60'
45'
30'
15'
0'
200
100
300
20K
40K
60K
80K
100K
120K
140K
160K
180K
200K
220K
Conventional ChIP-seq profiles
Calibrated ChIP-seq profiles

# https://gitbio.ens-lyon.fr/LBMC/Bernard/quantitative-nucleosome-analysis

```
Genome              : data/PB_2020/fasta/S_pombe_ASM294v226.fa
Genome calibration  : data/PB_2020/fasta_calib/S288C_reference_sequence_R64-2-1_20150113.fasta
CSV_Path            : data/PB_2020/datafromPATH_run1_single.csv
CSV_SRA             : Not supplied
Remove Duplicate    : True
Peak Calling        : True
MACS2 Genome Size   : 1.25E7
Config Profile      : singularity
-------------------------------------------------
loading local csv files
executor >  local (1)
[3d/17e046] process > fastp:fastp_default (IP_4483_L3_R1)                                                [  0%] 0 of 10
executor >  local (1)
[3d/17e046] process > fastp:fastp_default (IP_4483_L3_R1)                                                [  0%] 0 of 9
[c5/5db1dd] process > mapping:concatenate_genome (S_pombe_ASM294v226 S288C_reference_sequence_R64-2-1_20150113) [100%] 1 of 1, cached: 1 ✔
[55/e3fa3f] process > mapping:index_fasta (S_pombe_ASM294v226)                                           [100%] 1 of 1, cached: 1 ✔
[-        ] process > mapping:mapping_fastq                                                             -
[-        ] process > mapping:sort_bam                                                                  -
[-        ] process > mapping:filter_bam_mapped                                                         -
[-        ] process > mapping:filter_bam_multimapped                                                    -
[-        ] process > mapping:sort_bam_monomapped                                                       -
[-        ] process > mapping:sort_bam_multimapped                                                      -
[-        ] process > mapping:filter_bam                                                                -
[-        ] process > mapping:filter_bam_calib                                                          -
[-        ] process > mapping:rename_calib_bam                                                          -
[-        ] process > mapping:mark_duplicate                                                            -
[-        ] process > mapping:mark_duplicate_calib                                                      -
[-        ] process > mapping:index_bam                                                                 -
[-        ] process > mapping:rename_indexed_bam                                                        -
[-        ] process > mapping:group_indexed_bam                                                         -
[-        ] process > coverage_analysis:stats_bam                                                       -
[-        ] process > coverage_analysis:rename_mapping_stats                                            -
[-        ] process > coverage_analysis:mapping_stats                                                   -
[-        ] process > coverage_analysis:bam_to_bedgraph                                                 -
[-        ] process > coverage_analysis:coverage_normalization                                          -
[-        ] process > coverage_analysis:bedgraph_to_bigwig                                              -
[-        ] process > coverage_analysis:normalization_stats                                             -
[-        ] process > peak_calling:peak_calling_macs2                                                   -
[-        ] process > peak_calling:merge_peaks                                                          -
[-        ] process > peak_calling:peak_calling_quantification                                          -
[-        ] process > flagstat_2_multiqc                                                                -
[-        ] process > multiqc                                                                           -
```

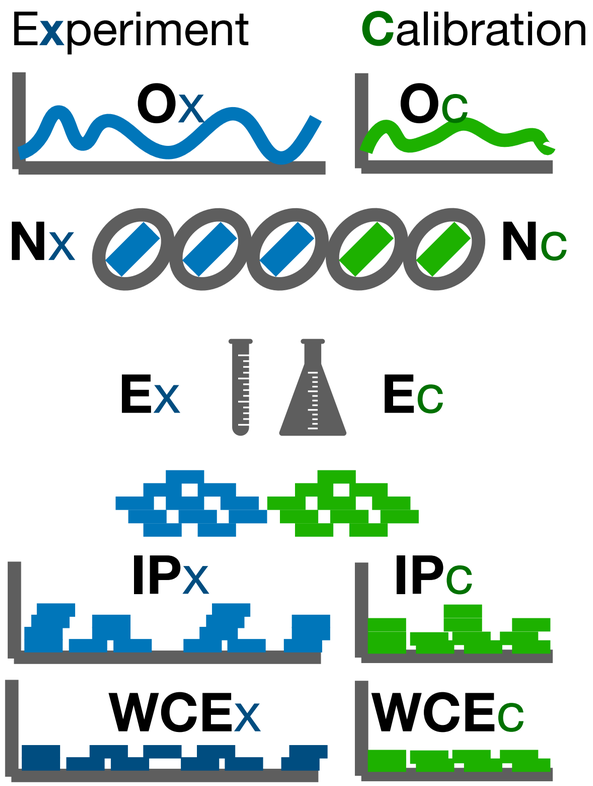
# Calibrated Chip-Seq




$$\widehat{OR}_{xi} = \frac{WCE_{ci} IP_{xi}}{WCE_{xi} IP_{ci}}$$

$$\widehat{OR}_{xi} = \frac{\frac{IP_{xi}}{WCE_{xi}}}{\frac{IP_{ci}}{WCE_{ci}}}$$

# Calibrated Chip-Seq




$$\frac{IP_{xi}(t_{xi})}{WCE_{xi}(t_{xi})} \qquad \frac{IP_{ci}(t_{ci})}{WCE_{ci}(t_{ci})}$$

$$\widehat{OR}_{xi} = \frac{WCE_{ci} IP_{xi}}{WCE_{xi} IP_{ci}}$$

$$\widehat{OR}_{xi} = \frac{\frac{IP_{xi}}{WCE_{xi}}}{\frac{IP_{ci}}{WCE_{ci}}}$$

$$\widehat{OR}_{xi}(t_{xi}) = \frac{\frac{IP_{xi}(t_{xi})}{WCE_{xi}(t_{xi})}}{\frac{IP_{ci}(t_{ci})}{WCE_{ci}(t_{ci})}}$$