CaptureC “peak finding”

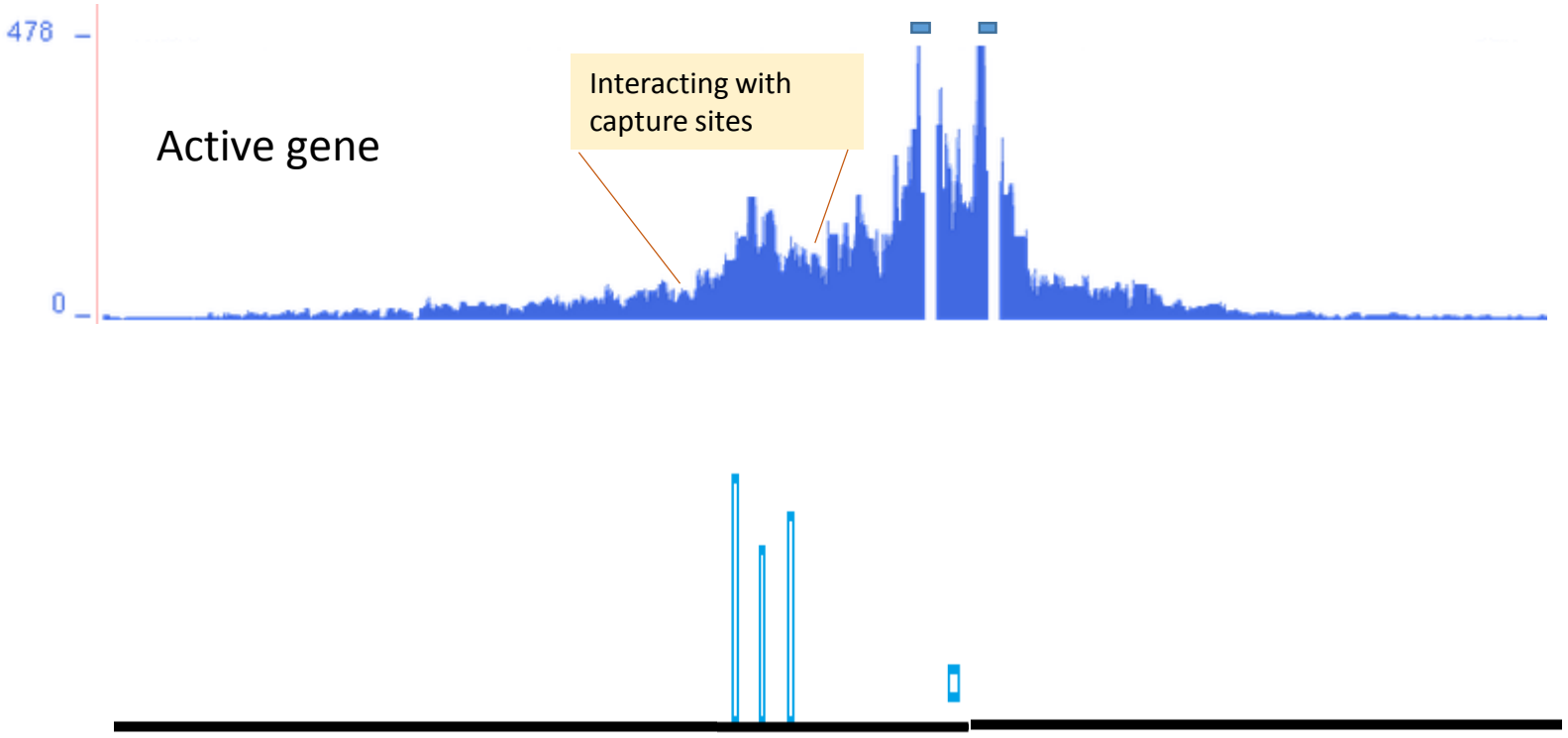
peakC, Chicago, r3Cseq, FourCseq

overview to “status quo” - Jelena 22 Jan 2016

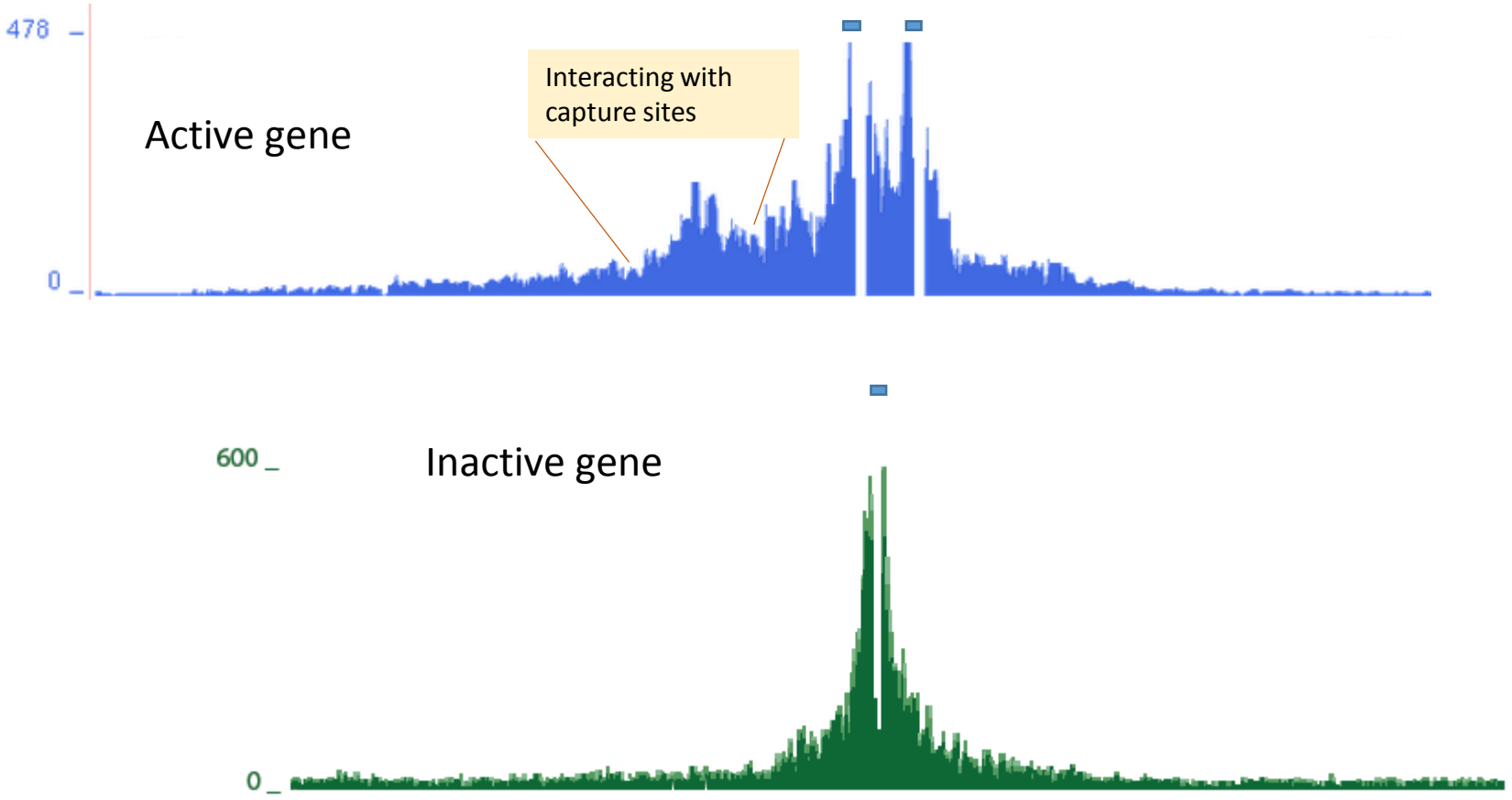
CaptureC “peak finding”

- INTRODUCTION
  - captureC signal / noise / statistics
- PART 1
  - fourCSeq, r3Cseq, peakC
- PART 2
  - The promise of Chicago

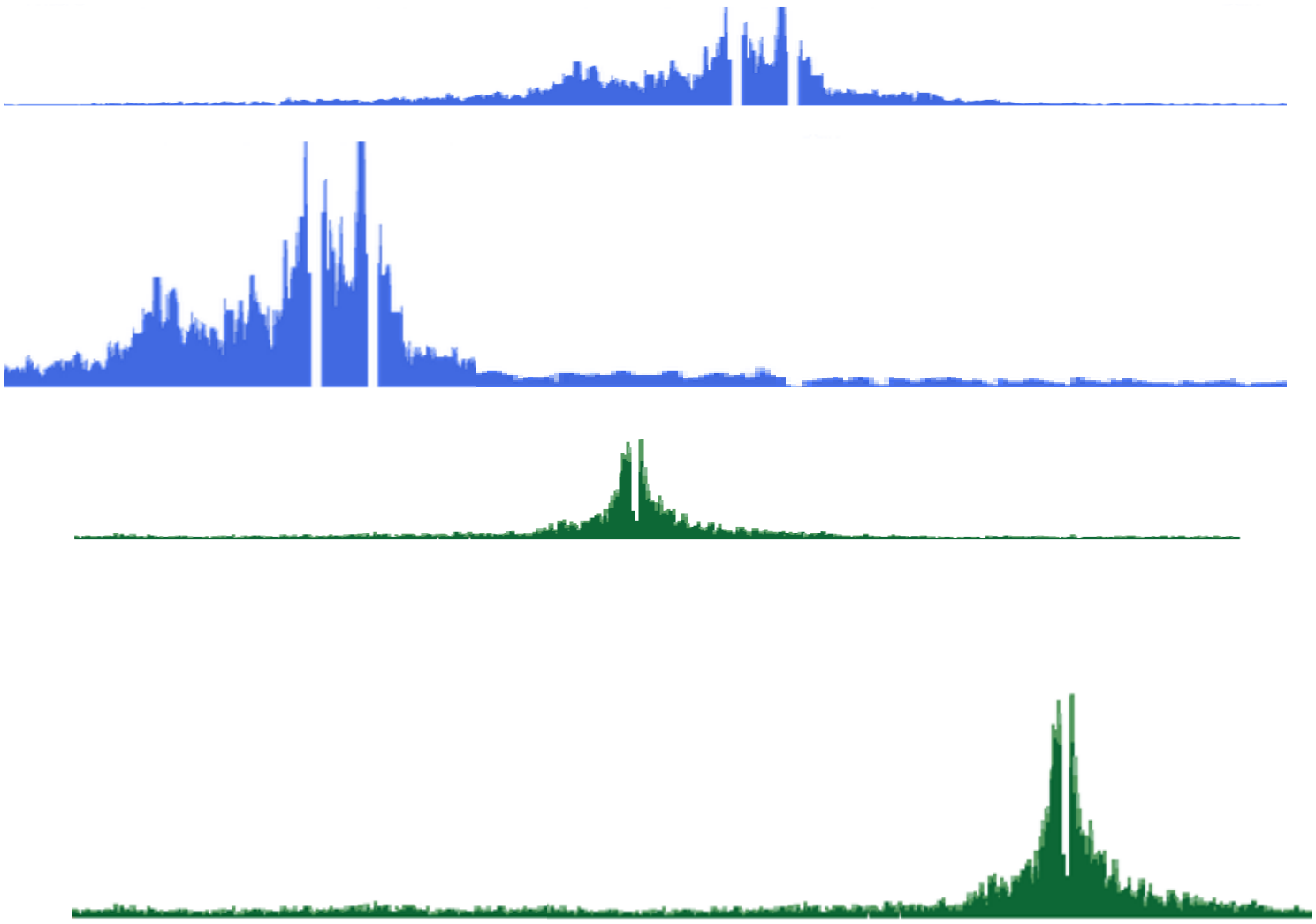
CaptureC “signal” vs “noise”



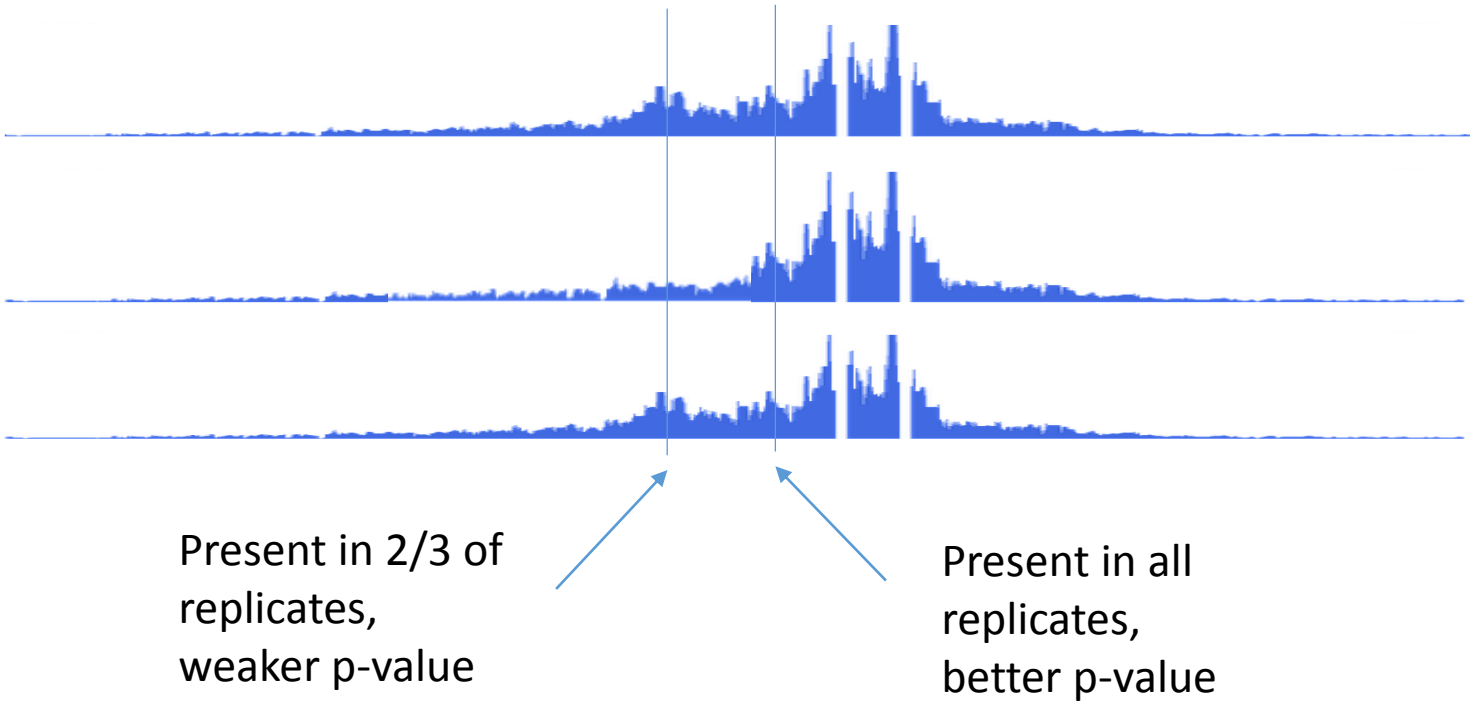
CaptureC “signal” vs “brownian noise”



CaptureC “signal” vs “technical noise”



p-values – which peaks are “real” ?

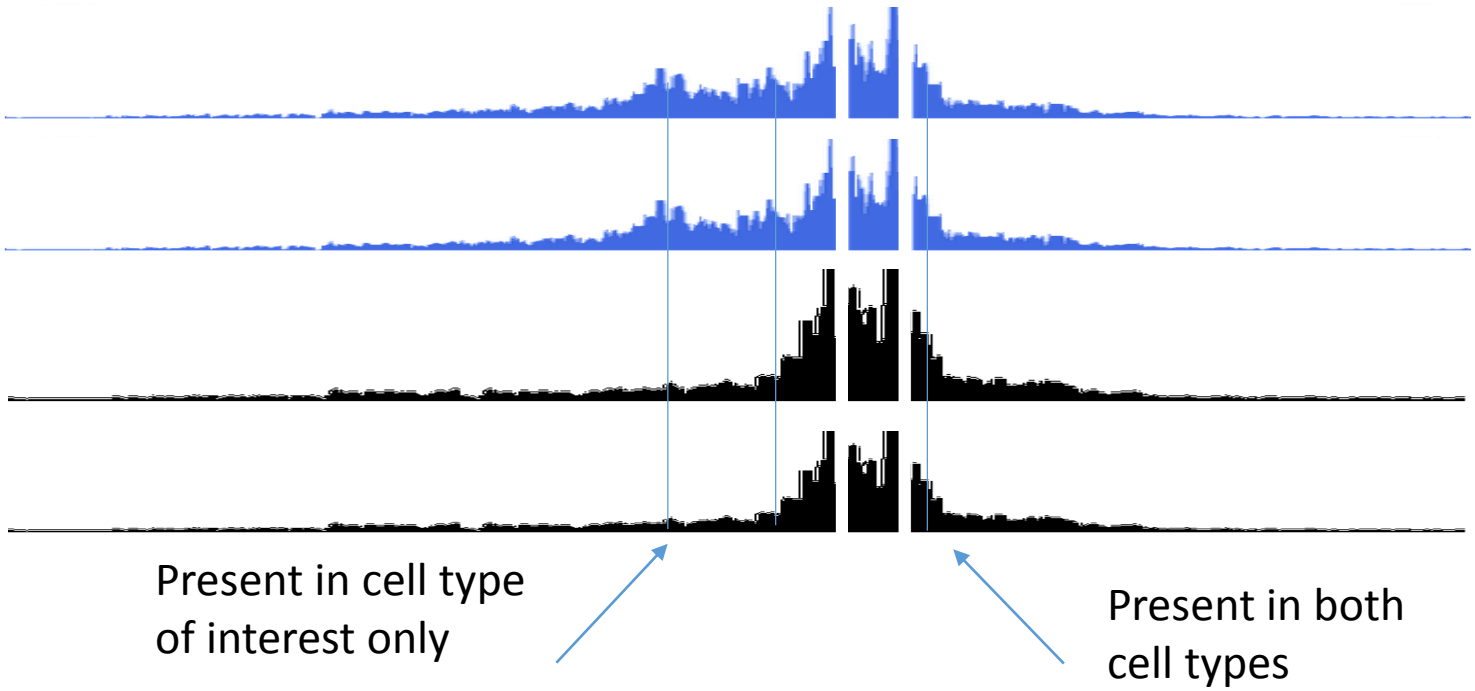


P- values : “lower is better”

cutoff examples : $p < 0.01$, $p < 10^{-5}$ etc…

“Statistically significant interactions (p<0.01 within 3 replicates data set)”

Fold changes, and their p-values – which peaks are “different” ?



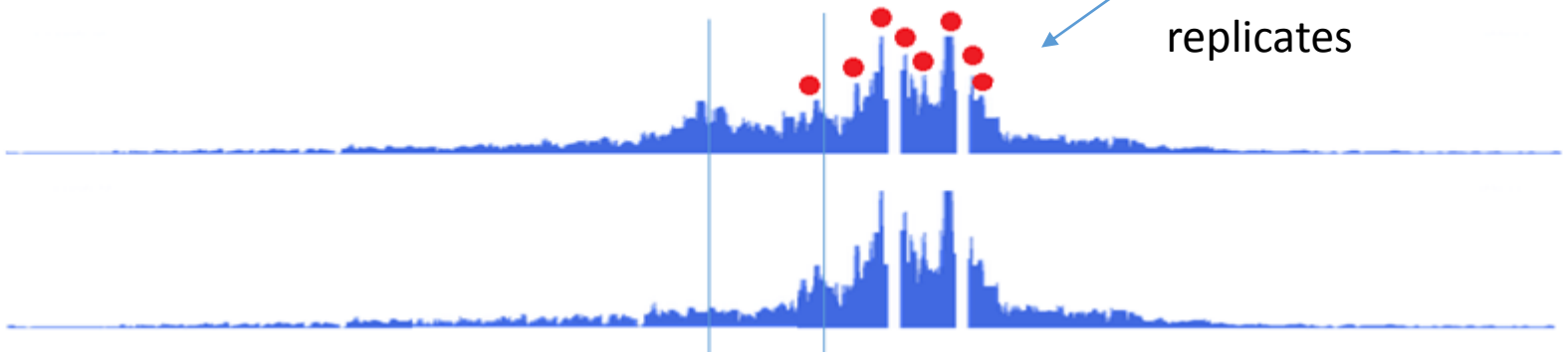
Fold changes : higher is better.

cutoff examples : $\log_2$fold change > 1 (sample at least 2X output)
$\log_2$fold change > 2 (sample at least 4X output)

“Statistically significant interactions ($\log_2$fold change > 2 , p< 0.01)”

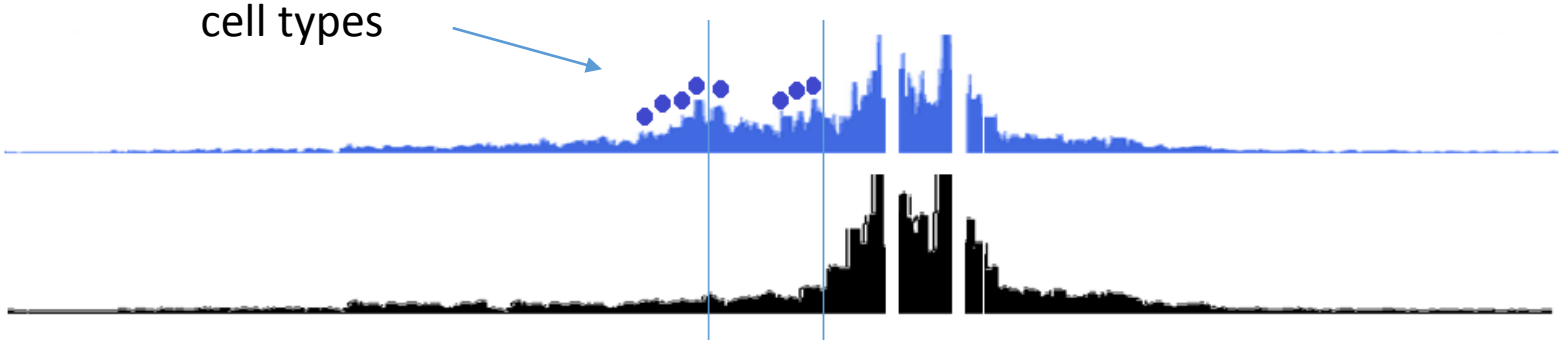
Peak in cell type of interest = red

Present in both replicates

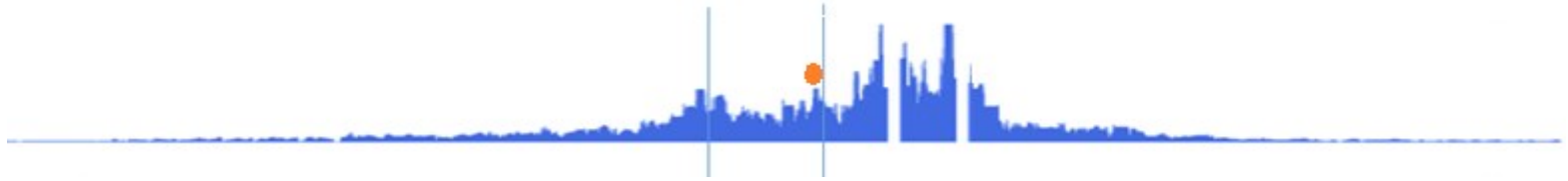


Different between cell types

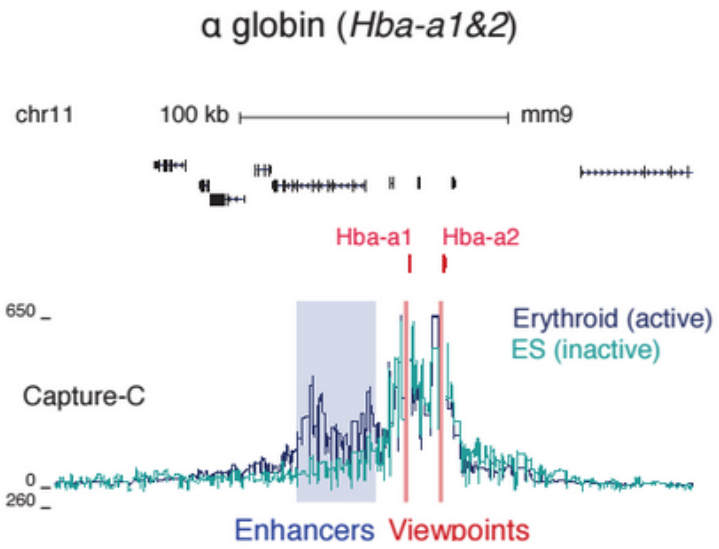
Different between cell types = blue



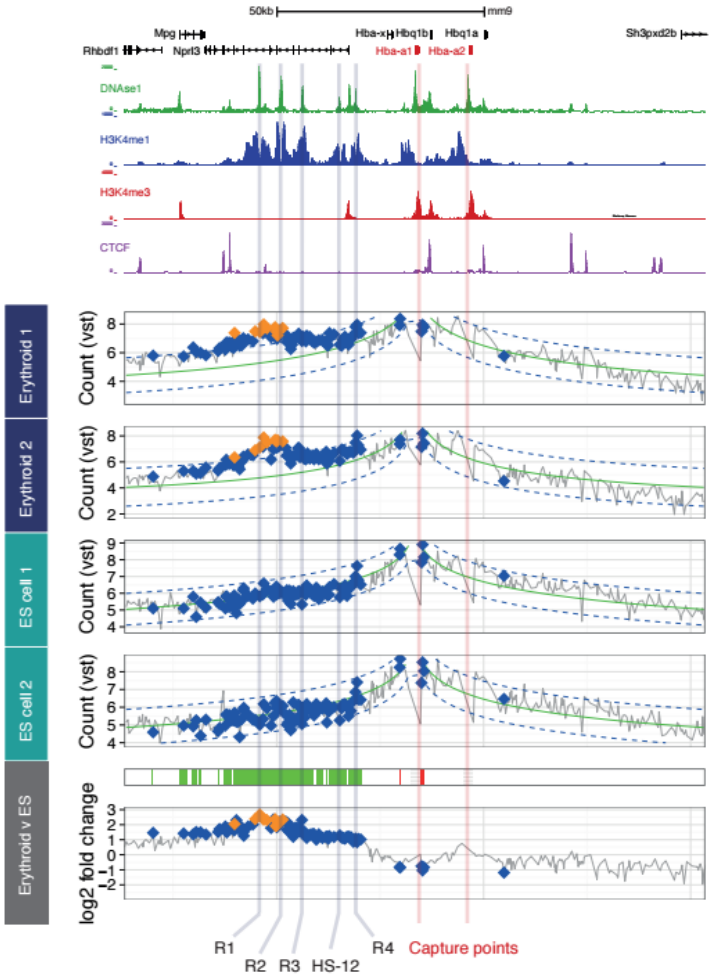
Peak in cell type of interest AND different between cell types = orange

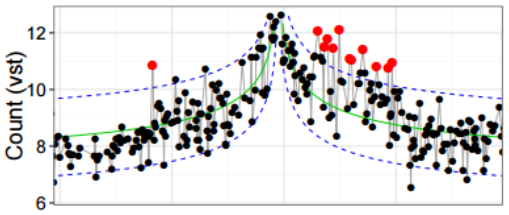


FourCSeq



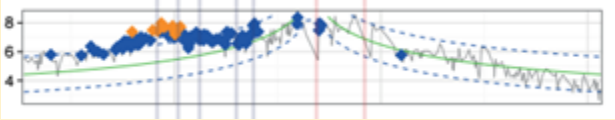
Supplementary Figure 21





Hard to get "enough" red dots here

fourCSeq – summary

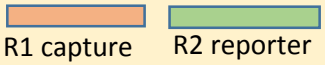


User experience

(+) Good tutorial
(-) Code not available
(-) Complicated R object

Input

(-) HiCup type PE bam



(+) Pipeline support

Nicolas Servant
Equipe NGS Analyse

Institut Curie, Plateforme de Bioinformatique
Unité 900 : Institut Curie - Inserm - Mines ParisTech

Output

(-) No UCSC-loadable tracks
(+) R object → bedgraph

Peak calls

(-) VERY close to viewpoint
(-) Weak distant peaks
(+) “Spot on” otherwise

(+) “Sweet spot” for CaptureC analysis (parameters)

Performance

(+) Plots (red, blue, orange)

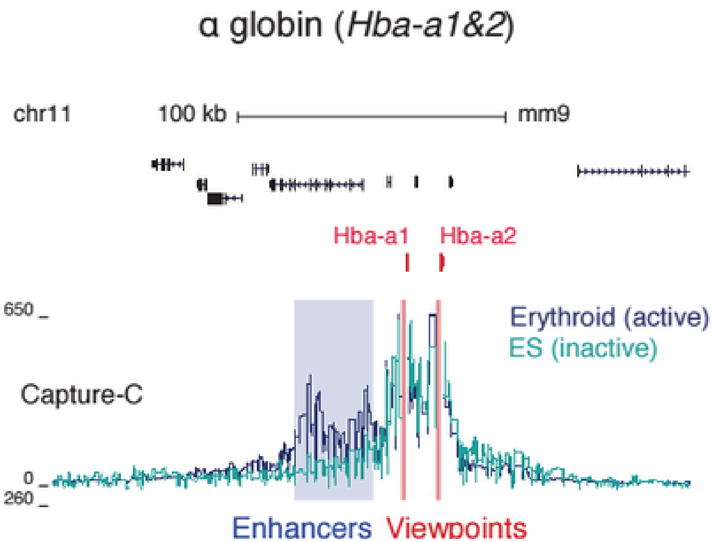
(-) “Trans” analysis broken ?

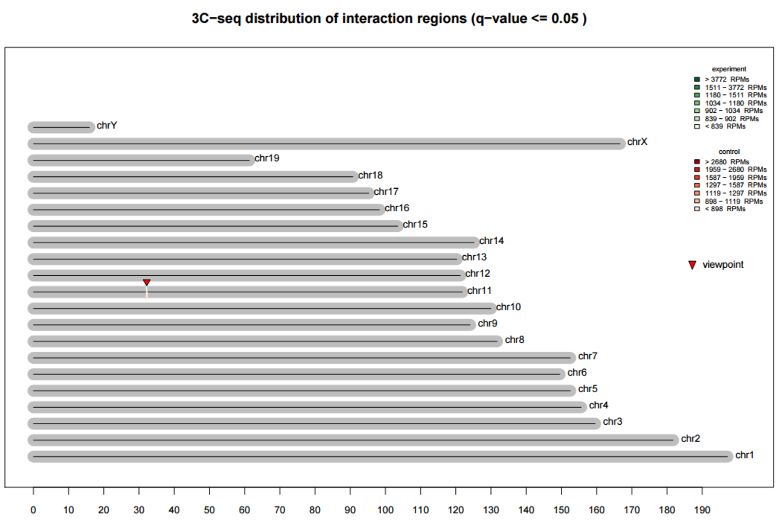
Properties

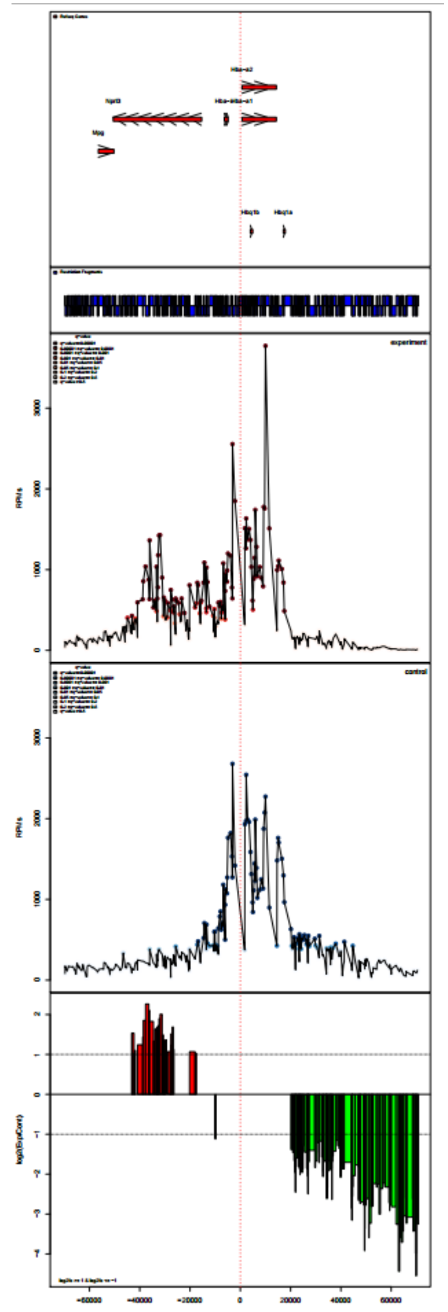
(+) Replicates
(+) Comparing cell types

(-) No trans chromosomes

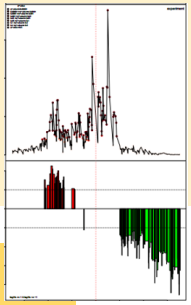
r3Cseq

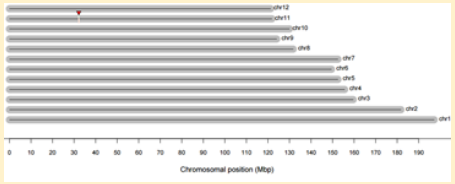

α globin (*Hba-a1&2*)
chr11
100 kb
mm9
Hba-a1
Hba-a2
650
Erythroid (active)
ES (inactive)
Capture-C
0
260
Enhancers
Viewpoints


3C−seq distribution of interaction regions (q−value <= 0.05)
experiment
control
viewpoint
chrY
chrX
chr19
chr18
chr17
chr16
chr15
chr14
chr13
chr12
chr11
chr10
chr9
chr8
chr7
chr6
chr5
chr4
chr3
chr2
chr1
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190
Chromosomal position (Mbp)


experiment
control

r3Cseq – summary





User experience

(+) Good tutorial
(+) Code available
(-) Complicated R object

(+) We know the developer

Supat Thongjuea

MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine

Input

(+) Relatively easy input

(+) Pipeline support

Output

(+) Auto-generates UCSC tracks
(+) Rest of tracks easy to parse from output

Peak calls

(-) VERY close to viewpoint
(-) Weak distant peaks
(+) “Spot on” otherwise

(-) Calls too wide regions
(-) No finetuning

Performance

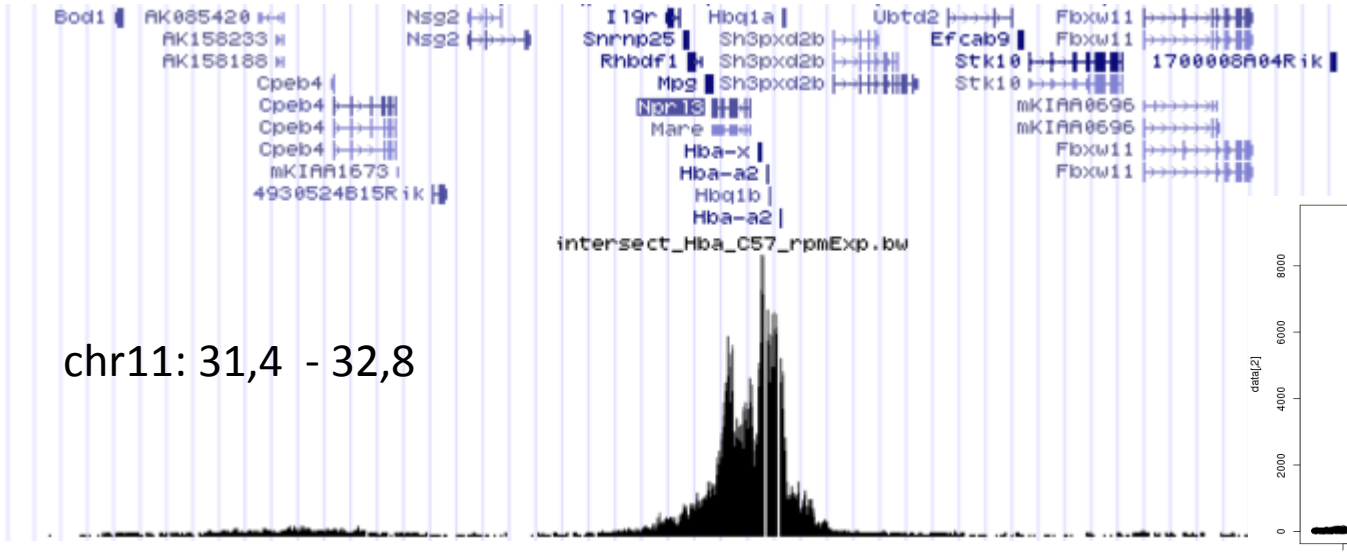
(+) Trans and long range cis

(+) Tested for globin genes

Properties

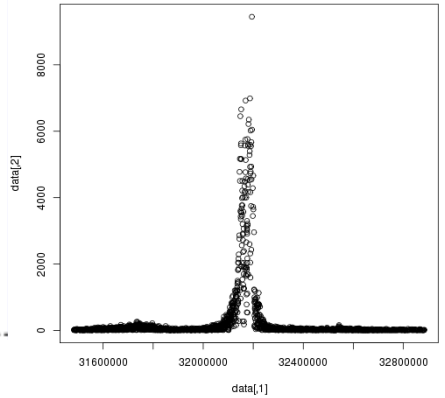
(-) Replicates

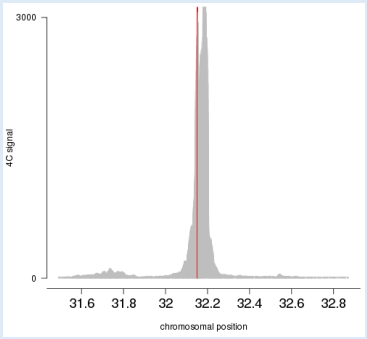
(+) Comparing cell types
(+) Trans chromosomes



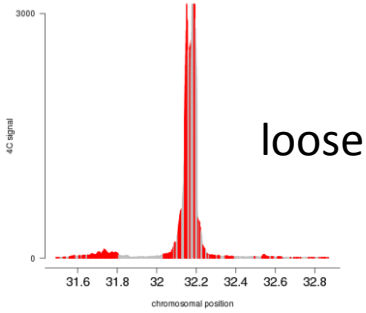
chr11: 31,4 - 32,8

peakC

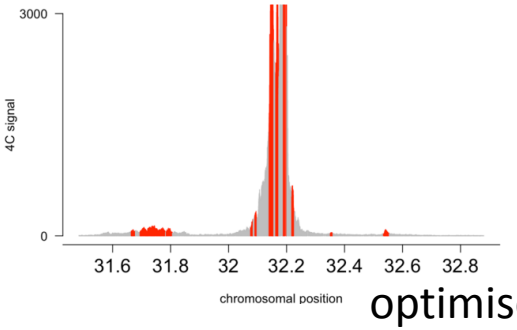




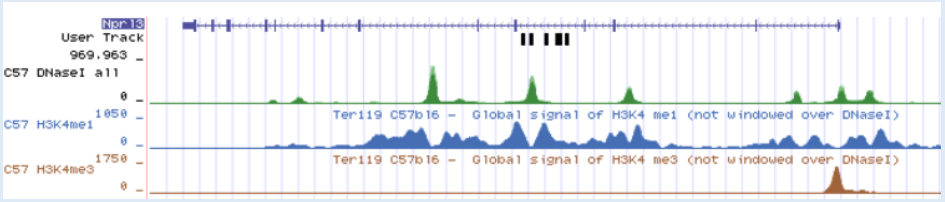
default

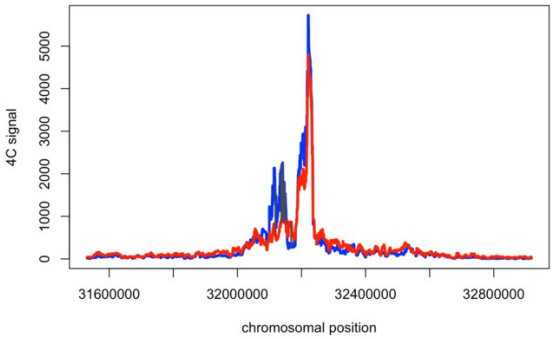


loose

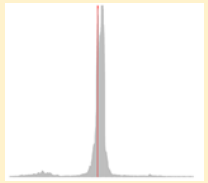


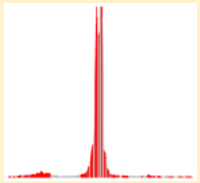
optimised

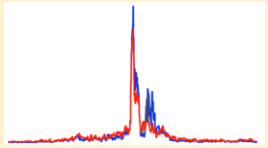




peakC – summary







User experience

(+) Code available
(+) Code can be edited
(+) Simple R object

(+) We know the developer

Elzo de Wit group
Netherlands Cancer Institute

(-) Code not published

Input

(+) Easy input

(+) Pipeline support

Output

(+) Easy output

Peak calls

(-) Default parameters not always very good
(+) Play with your peaks (easy)

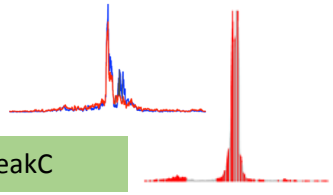
Performance

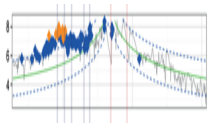
(+) Developed for CaptureC
(+) Easy to understand and modify

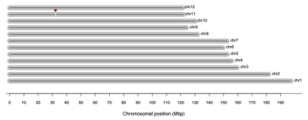
Properties

(+) Replicates
(+) Comparing cell types
(+) Trans chromosomes

SUMMARY

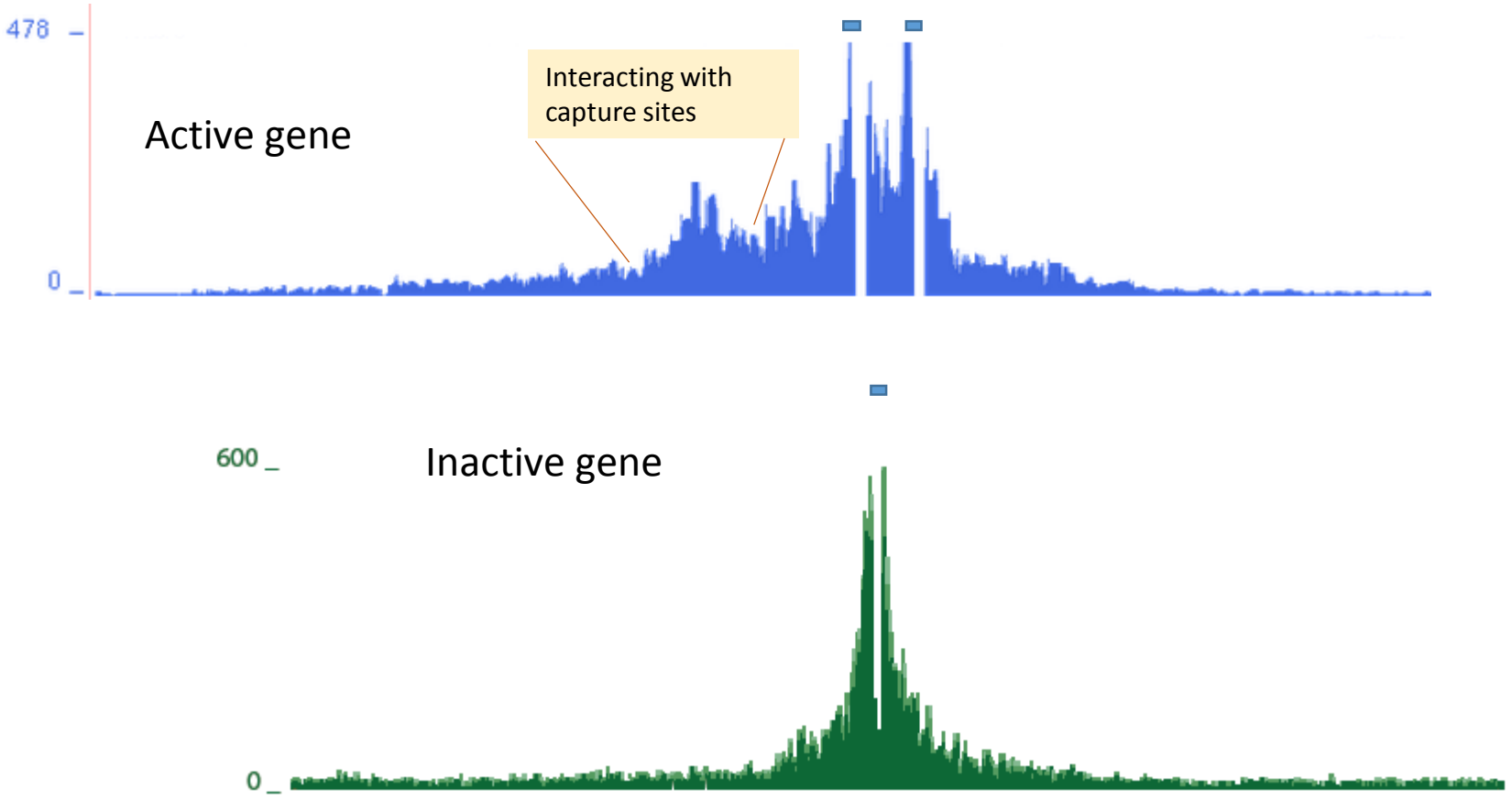




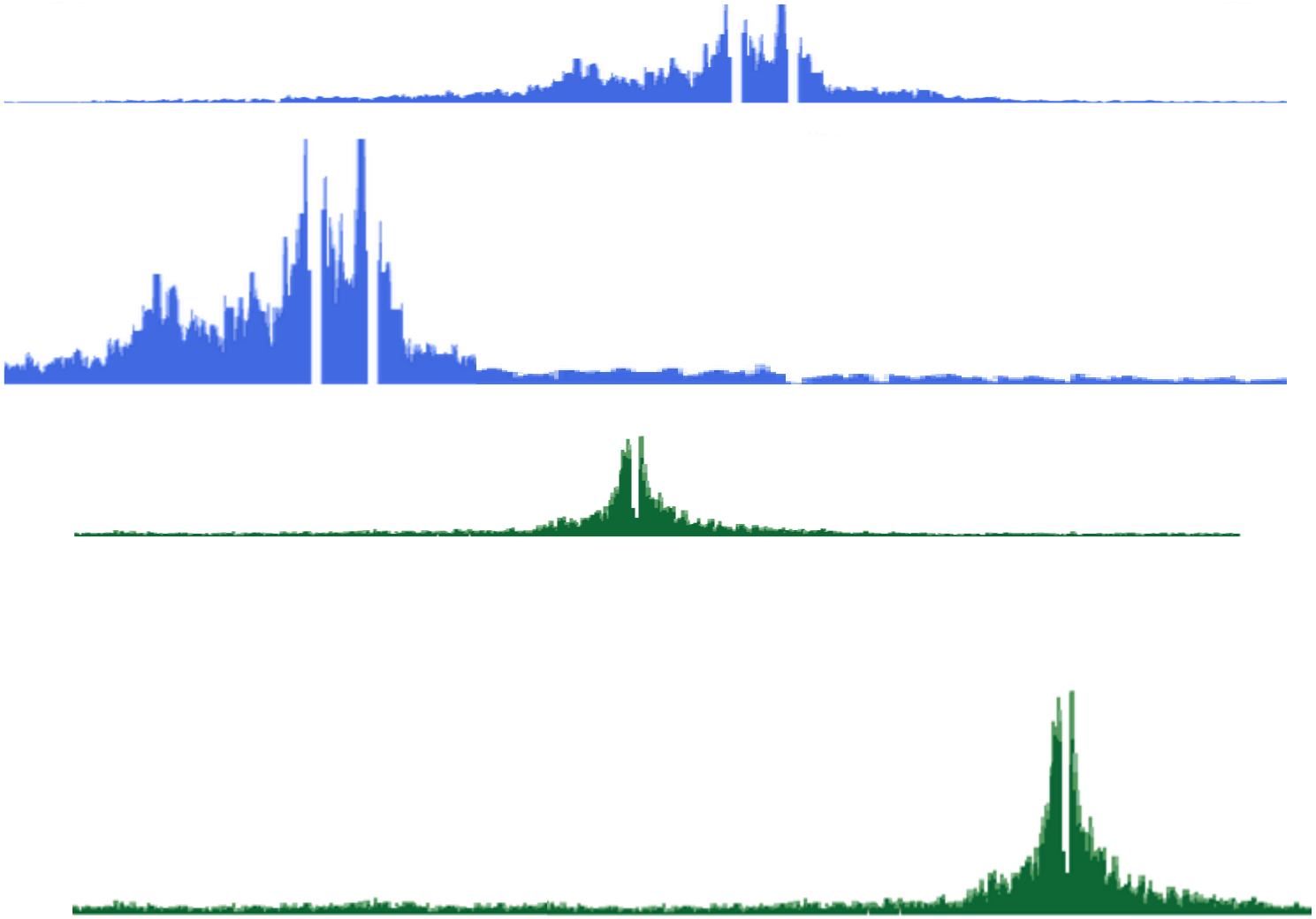


	peakC	FourCSeq	r3Cseq
INPUT	Pipeline support (columns 5-6 from the gff)	Pipeline support (bam via custom script)	Pipeline support (ploidy+blat filtered bam)
REPLICATES	YES	YES	NO
COMPARISON between cell types	YES	YES	YES
TRANS chromosomes	NO	NO (yes ?)	YES
Distance correction ?	NO / YES (optional)	YES	YES
Best part	“playground” for capture data !!	Red-blue-orange plot (cis)	Trans chromosome map
Worst part	Cannot be automated (default parameters ?)	Complicated R object	Too wide peaks (cannot be finetuned)
Worth to ask CBRG to add as R module ?	YES (once released)	YES	YES

The promise of Chicago : Brownian noise



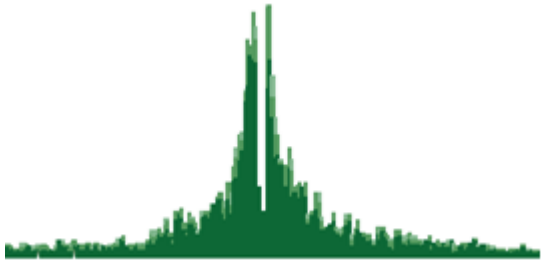
The promise of Chicago : “signal” vs “technical noise”



The promise of Chicago

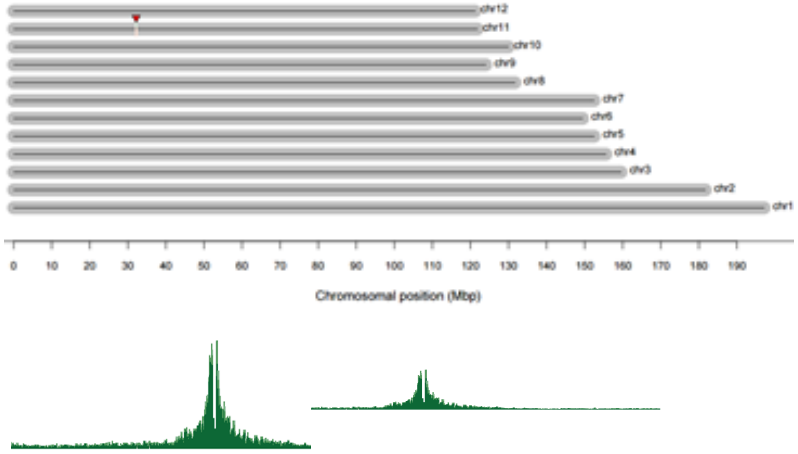
noise = Brownian noise + technical noise

Brownian noise dominates
CLOSE to capture oligo

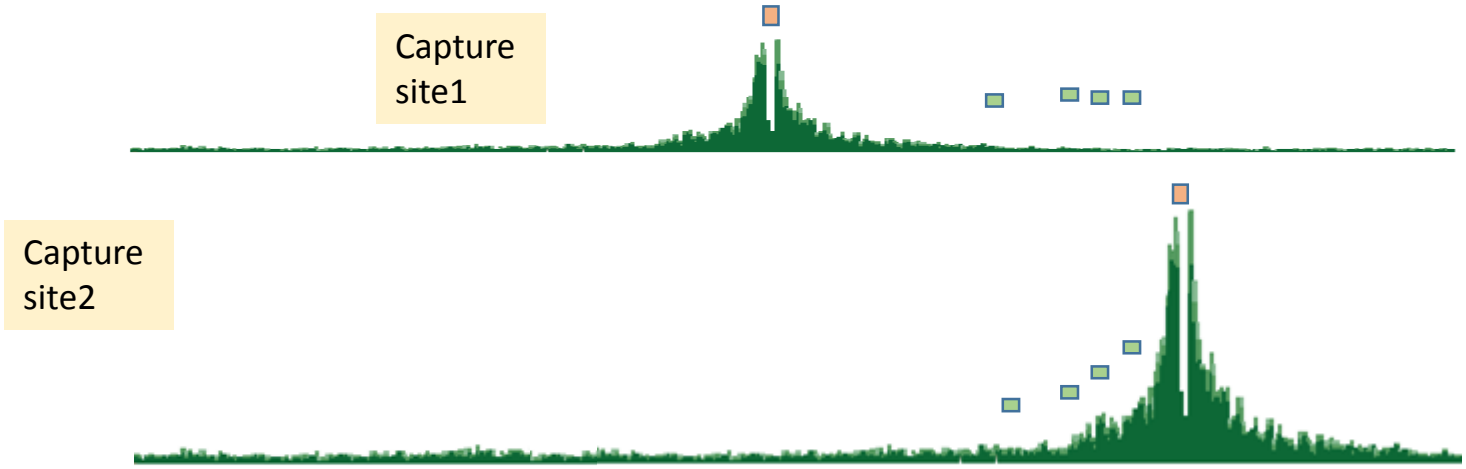


Technical noise dominates
in

- TRANS interactions
- Sequence-specific counts



The promise of Chicago : “signal” vs “technical noise” (1)



“NORMALISING CAPTURE SITES”

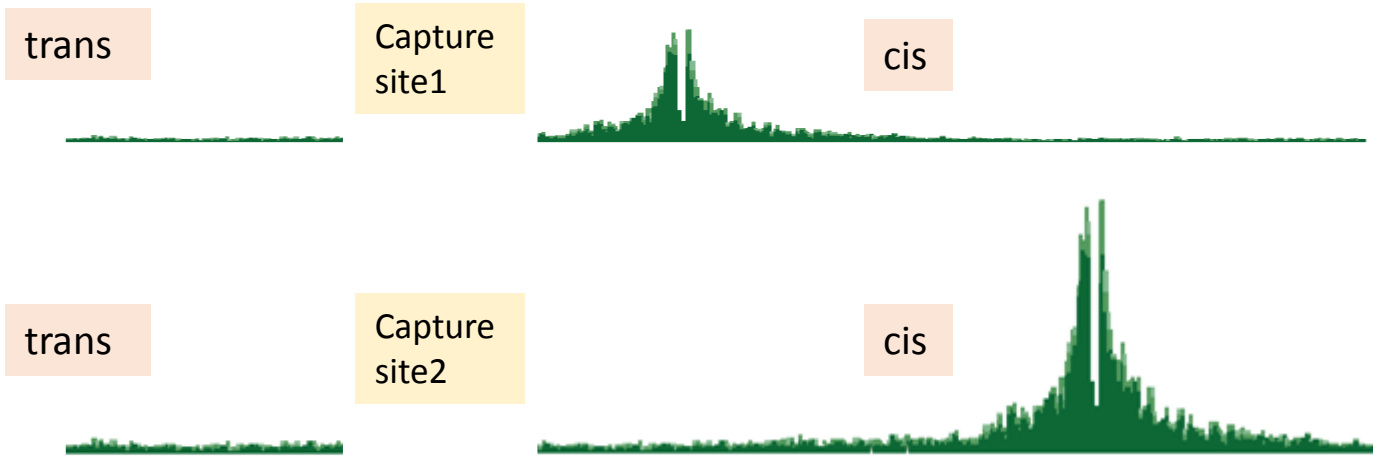
→ Use relative abundance of each CAPTURE to estimate the relative “strengths” of capture sites

→ Use relative abundance of each REPORTER to estimate the relative “strengths” of reporter sites

Capture2 fragment

reporter fragment

The promise of Chicago : “signal” vs “technical noise” (2)



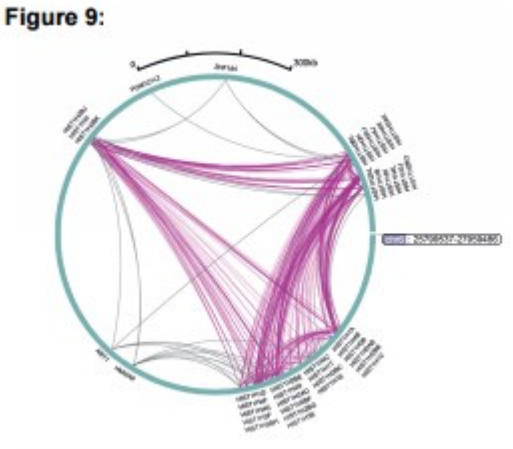
“SUBTRACTING RANDOM BACKGROUND”

Technical noise dominates in TRANS
→ Use the count of trans reads as estimate of “random background” for each oligo

Capture2 fragment

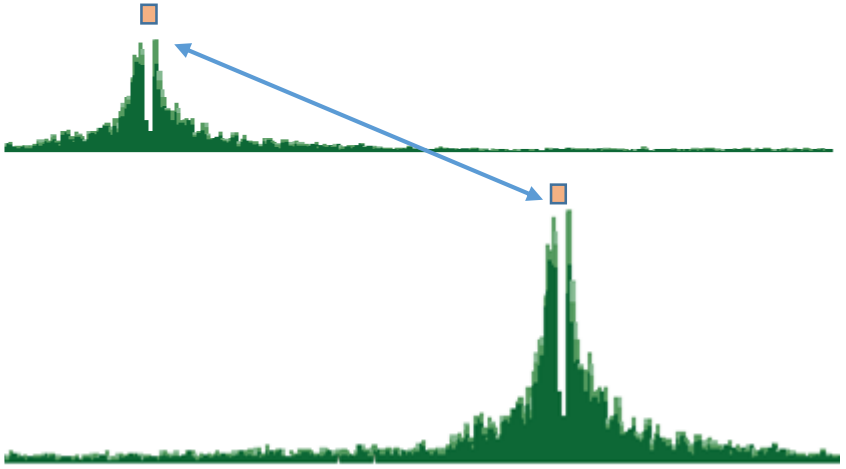
reporter fragment

The promise of Chicago : chromatin organisation

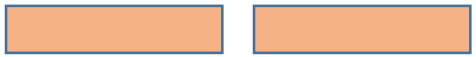


Capture site1

Capture site2



Contacts between capture oligos
→ capture-capture fragments show, if for example, promoters of “one kind of genes” cluster together in cellular space

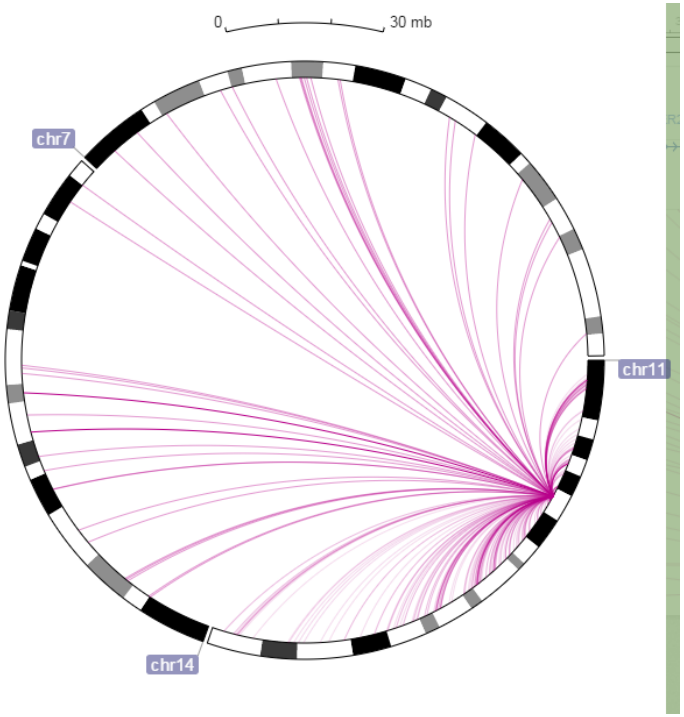


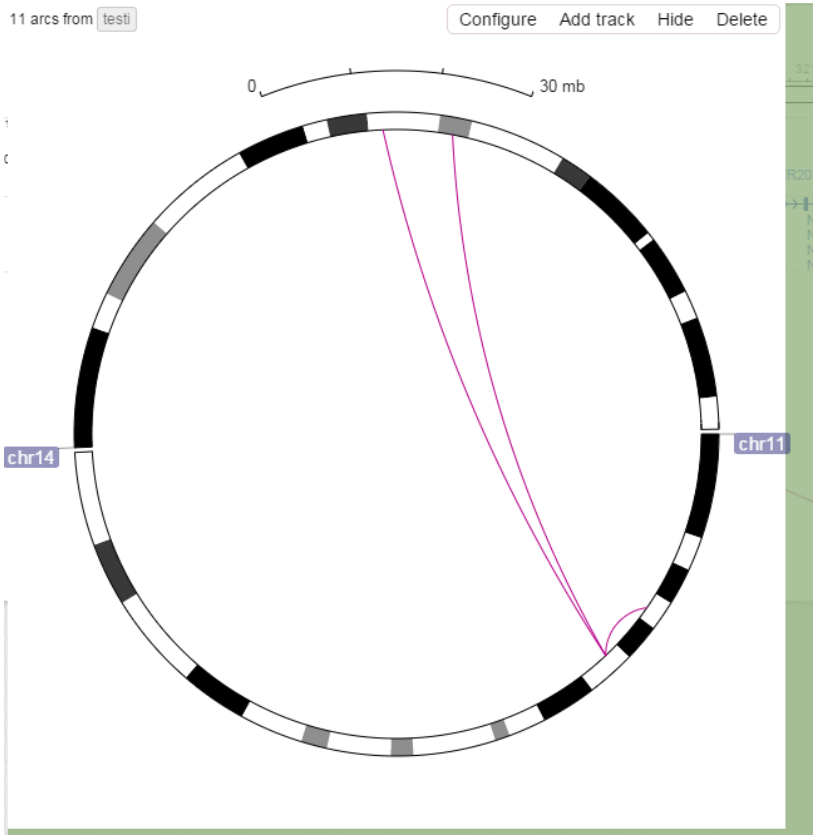
Capture2 fragment

Capture 1 fragment

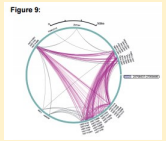

chr11:32084444-32205735
One pixel spans 72 bp
Tracks
Apps
qA1
qA2
qA3.1
qA3.2
qA3.3
qA4
qA5
qB1.1
qB1.3
qB2
qB3
qB4
qB5
qC
qD
qE1
qE2
1452
qA4
chr11
RepeatMasker
1-divergence%
RefSeq genes
testi
PB1
PB1D10
B1_Mm
B4A
B1_Mur1
MER20
B1_Mur3
ID_B1
B3
Snrnp25
Rhbdf1
Mpg
Nprl3
Hba-x
Hba-a1
Hba-a2
Hbq1b
Hbq1a
add terms

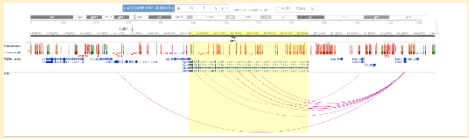
R1
R2
R3
HS-12
R4
Capture points


0
30 mb
chr7
chr11
chr14


11 arcs from testi
Configure
Add track
Hide
Delete
0
30 mb
chr14
chr11

chicago – summary





User experience

(-) Pre-release (bugs)
(-) Code not available
(-) Complicated R object

(+) Actively developed

Input

(-) HiCup type PE bam
(+) Code available
(-) Changes to CCanalyser.pl

Output

(?) Not tested

Performance

(-) Bugs prevent running our data
(+) Test data set runs fine

(+) Actively developed
(-) Cryptic error messages

Properties

(+) replicates

(-) comparing cell types
(+) trans chromosomes

(+) normalise between capture sites