ChIP seq

Tingwen Chen (陳亭妏)
Bioinformatics center
CGU

5.4.2012

Part I

DNA and Proteins

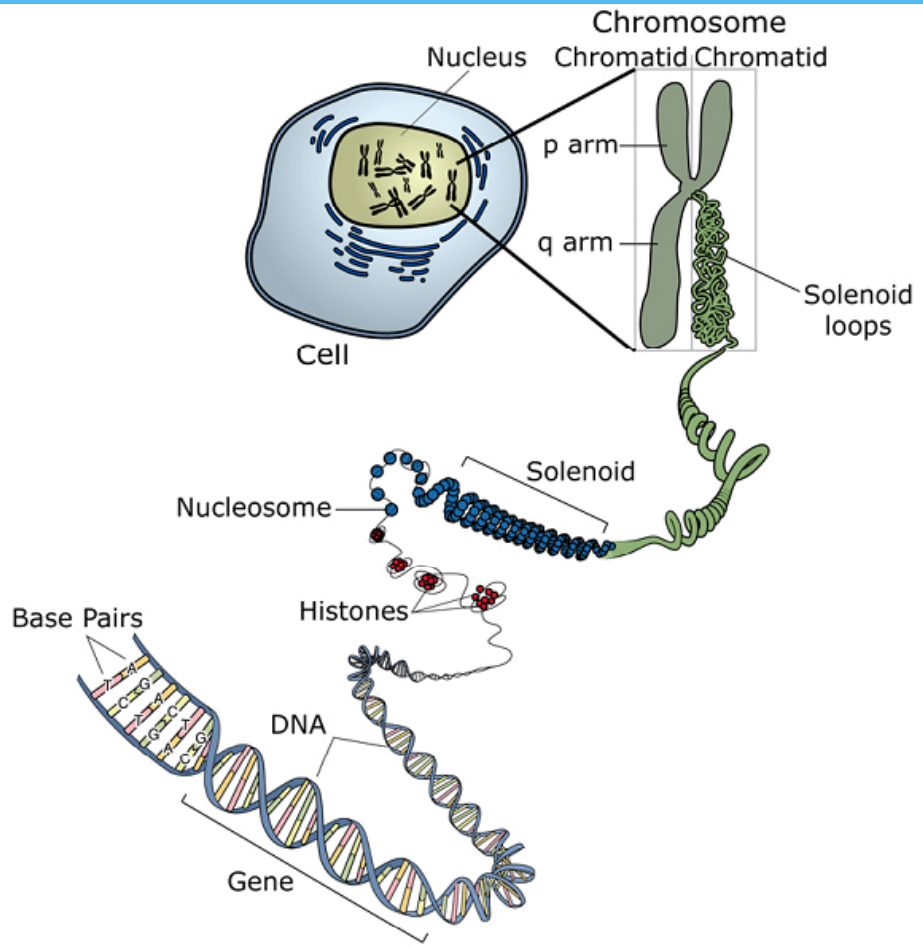
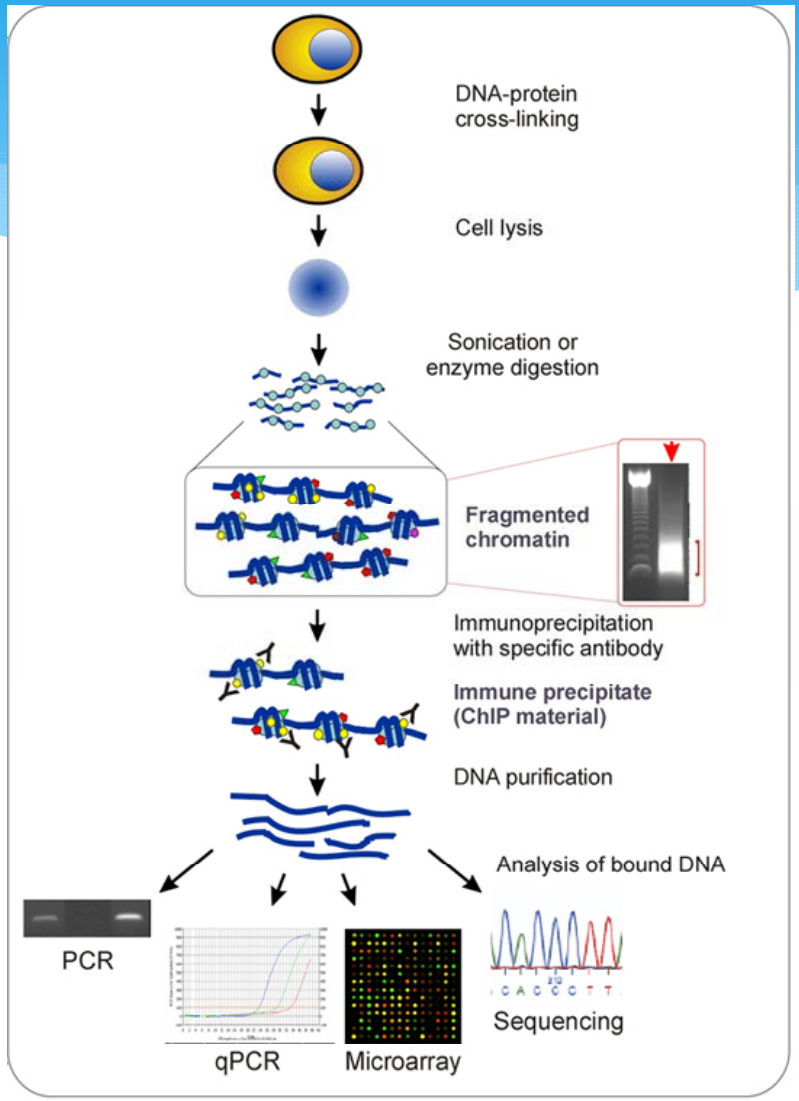


Image adapted from: National Human Genome Research Institute.

* Histone
* Histone acetylases
* Histone deacetylases
* Chromosome remodelers
* Transcription factor
* Meyhlases
* …

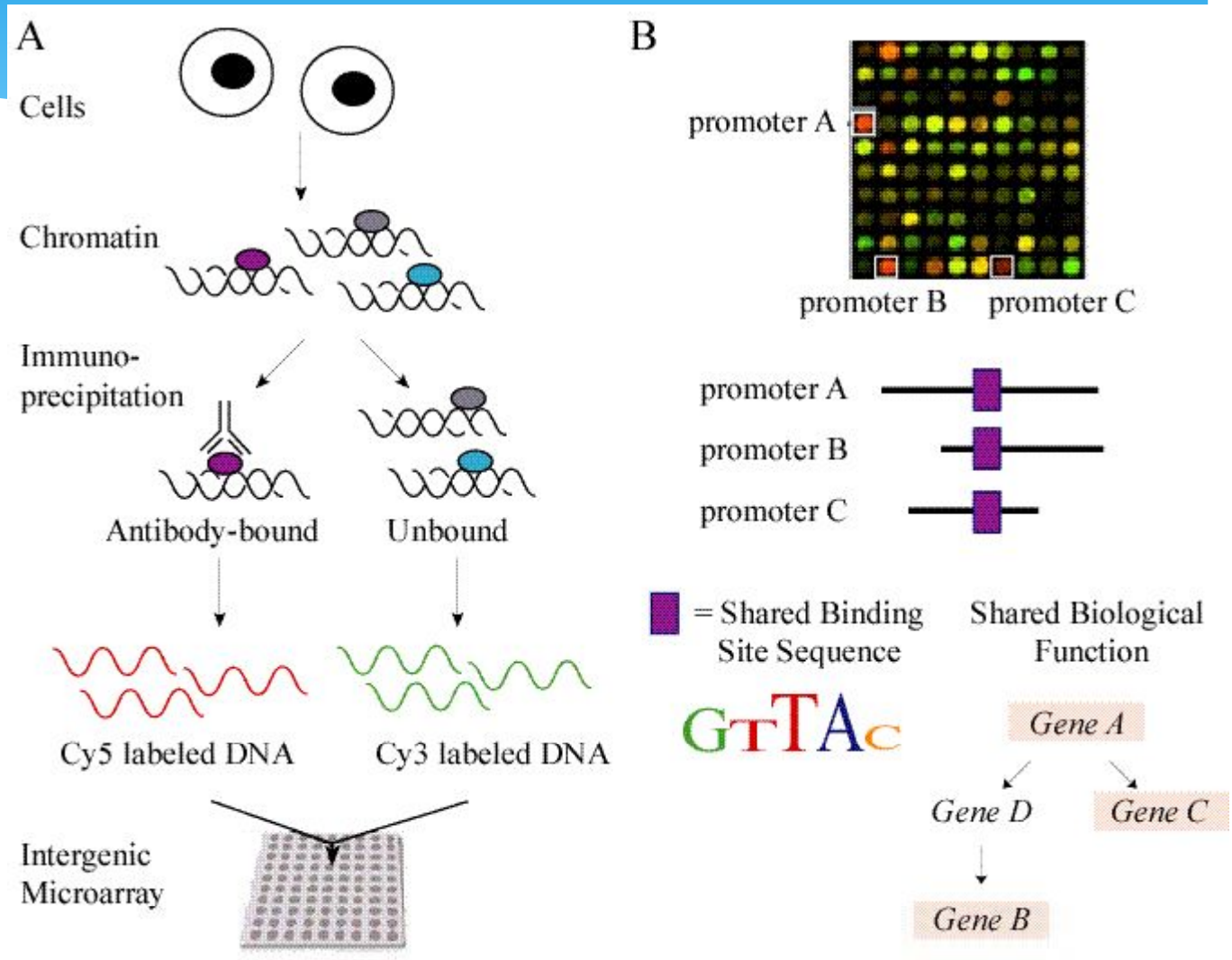
What is ChIP

* Chromatin immunoprecipitation
* Technique used to investigate the interaction between proteins and DNA in the cell



http://www.bioscience.org/2008/v13/af/2733/fulltext.asp?bframe=figures.htm&doi=yes

ChIP chip

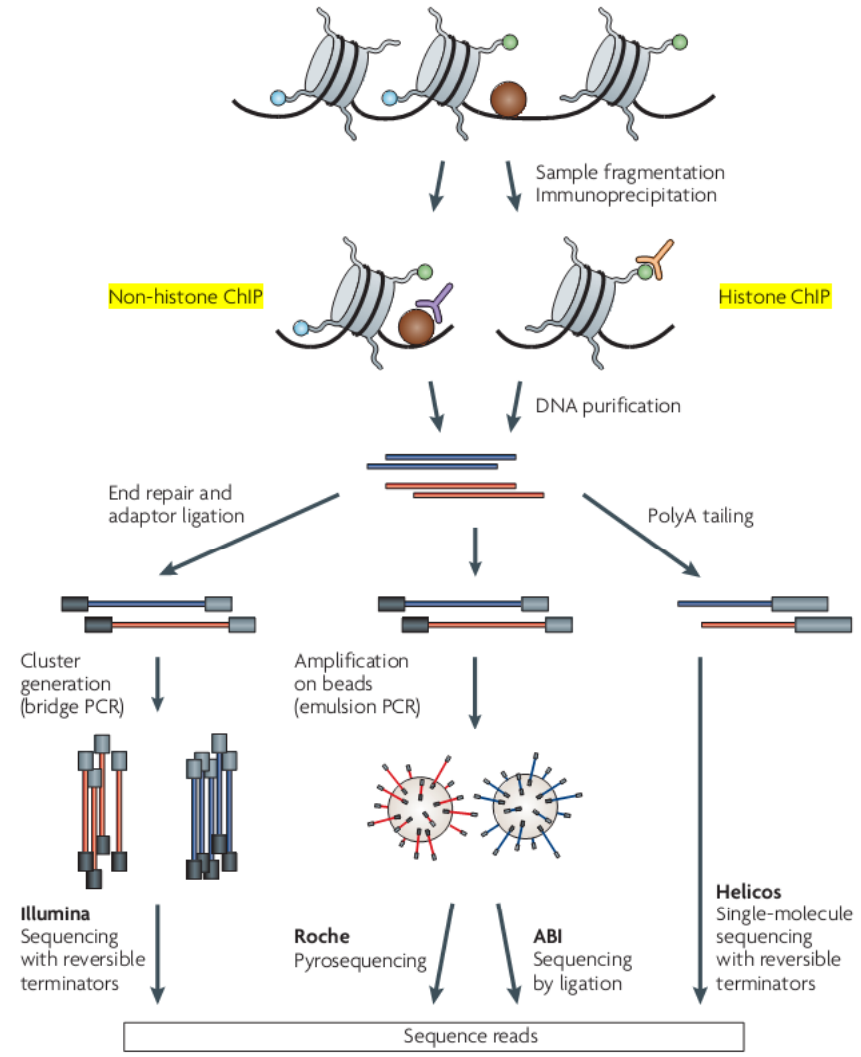


(Wong and Chang, 2005)

What is ChIP-Sequencing?

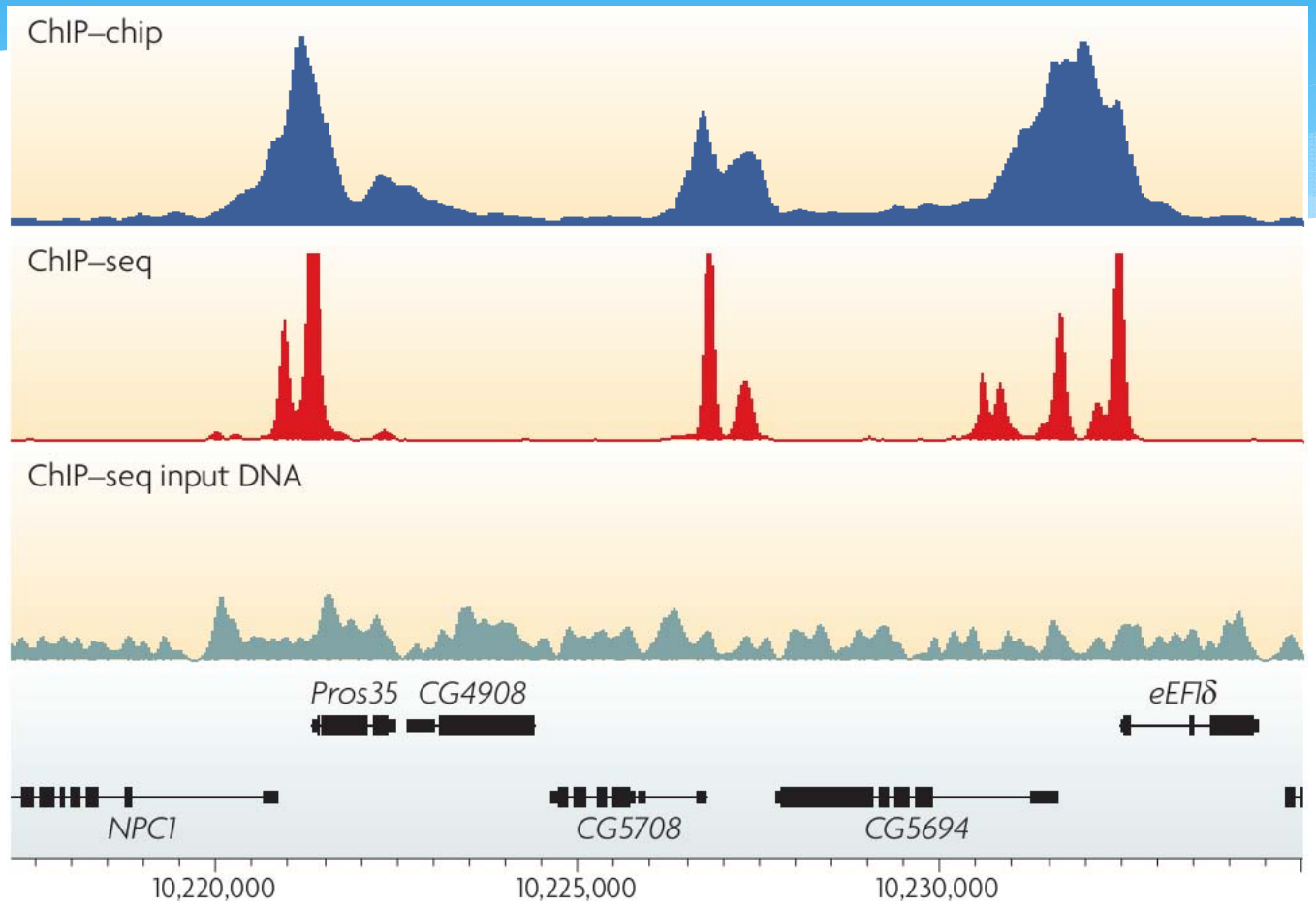
* ChIP-Sequencing is a new frontier technology to analyze protein interactions with DNA.
* ChIP-Seq
  * Combination of chromatin immunoprecipitation (ChIP) with ultra high-throughput massively parallel sequencing
  * Allow mapping of protein–DNA interactions in-vivo on a genome scale

ChIP seq



(2009, Park)

resolution



(Park, 2009)

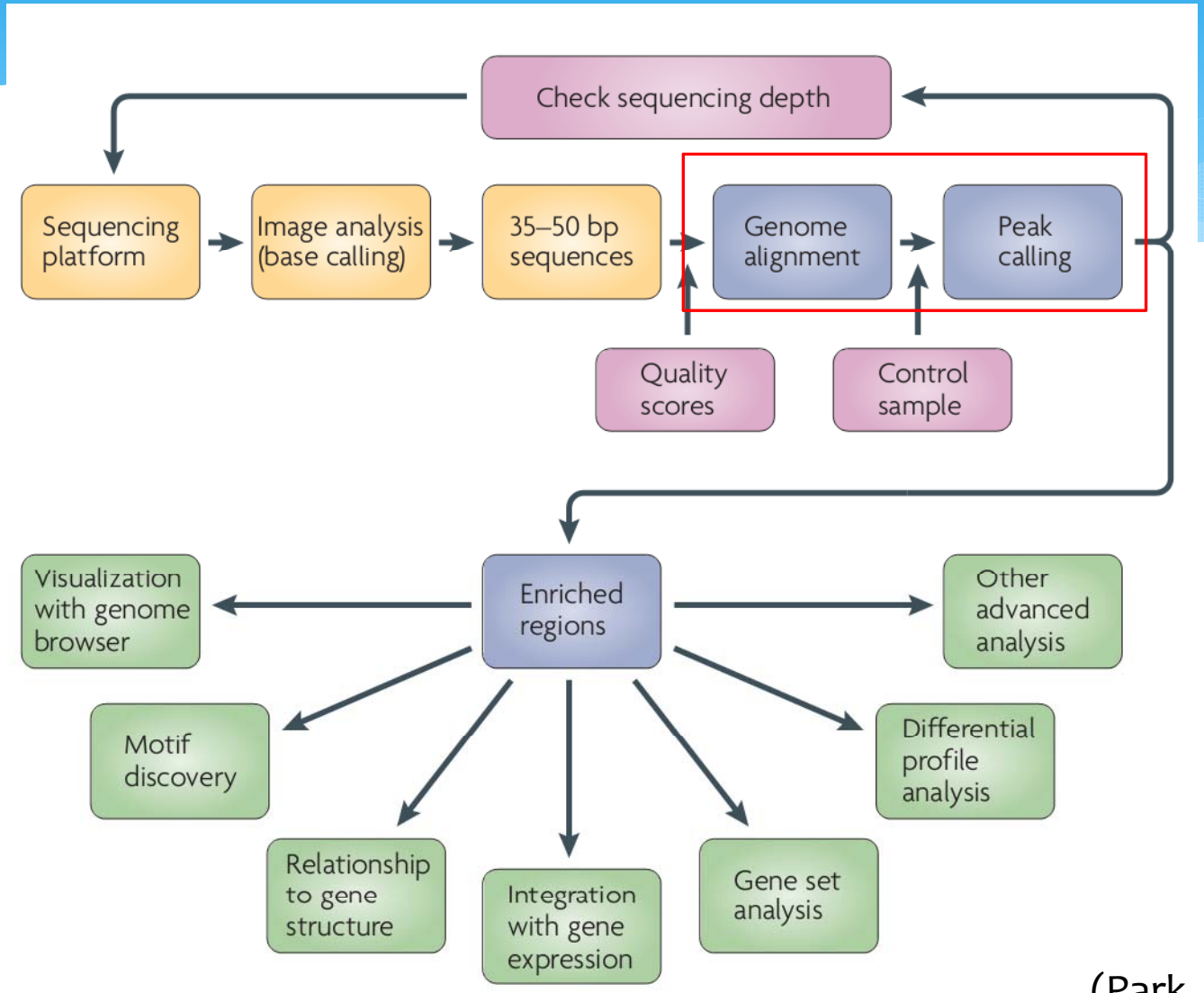
comparison

Table 1 | **Comparison of ChIP–chip and ChIP–seq**

	ChIP–chip	ChIP–seq
Maximum resolution	Array-specific, generally 30–100 bp	Single nucleotide
Coverage	Limited by sequences on the array; repetitive regions are usually masked out	Limited only by alignability of reads to the genome; increases with read length; many repetitive regions can be covered
Cost	US$400–800 per array (1–6 million probes); multiple arrays may be needed for large genomes	Currently US$1,000–2,000 per lane (using the Illumina Genome Analyzer); 6–15 million reads before alignment
Source of platform noise	Cross-hybridization between probes and nonspecific targets	Some GC bias can be present
Experimental design	Single- or double-channel, depending on the platform	Single channel
Cost-effective cases	Profiling of selected regions; when a large fraction of the genome is enriched for the modification or protein of interest (broad binding)	Large genomes; when a small fraction of the genome is enriched for the modification or protein of interest (sharp binding)
Required amount of ChIP DNA	High (a few micrograms) 10-100 ng => > 2 µg	Low (10–50 ng)
Dynamic range	Lower detection limit; saturation at high signal	Not limited
Amplification	More required	Less required; single-molecule sequencing without amplification is available
Multiplexing	Not possible	Possible

(Park, 2009)

For exam-ple, only 48% of the human genome is non-repetitive, but 80% is mappable with 30 bp reads and 89% is mappable with 70 bp reads.

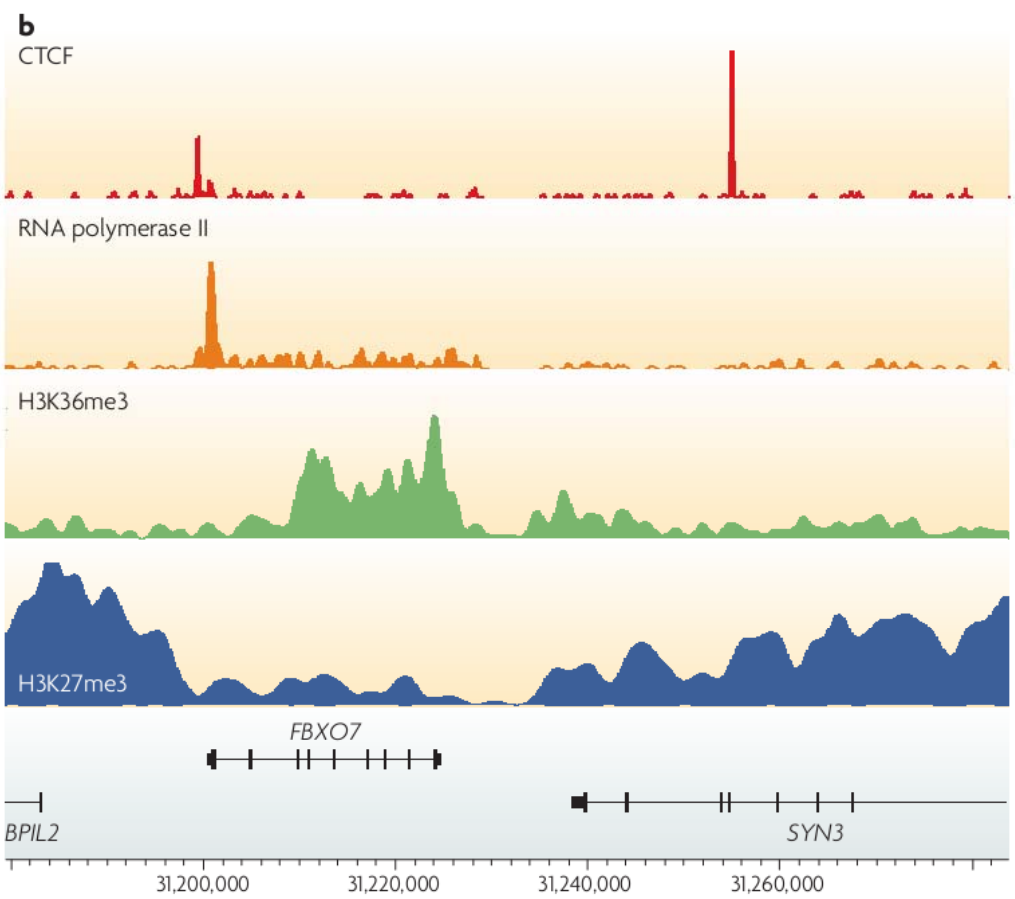

Check sequencing depth
Sequencing platform
Image analysis (base calling)
35–50 bp sequences
Genome alignment
Peak calling
Quality scores
Control sample
Visualization with genome browser
Enriched regions
Other advanced analysis
Motif discovery
Differential profile analysis
Relationship to gene structure
Integration with gene expression
Gene set analysis

(Park, 2009)

Mapping Methods: Indexing the Oligonucleotide Reads

* ELAND (Cox, unpublished)
  * "Efficient Large-Scale Alignment of Nucleotide Databases" (Solexa Ltd.)
* SeqMap (Jiang, 2008)
  * "Mapping massive amount of oligonucleotides to the genome"
* RMAP (Smith, 2008)
  * "Using quality scores and longer reads improves accuracy of Solexa read mapping"
* MAQ (Li, 2008)
  * "Mapping short DNA sequencing reads and calling variants using mapping quality scores"

Peak calling



(Park, 2009)

Sharp (e.g. TF binding)

Mixture (e.g. polymerase binding)

Broad (e.g. histone modification)

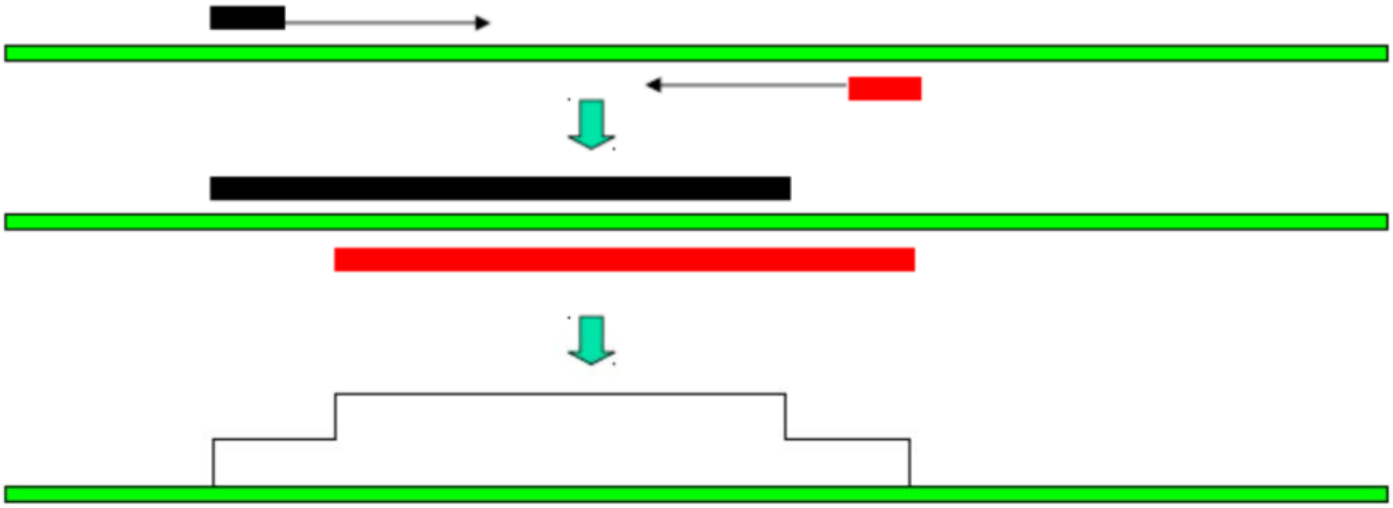
Region level Peak calling

* Usually a sliding-window approach is used
  * Typically, window size depends on the event size
  * Often overlapping/adjacent/nearby regions are merged
* More rarely, an island approach is used
  * Build regions out of overlapping (inferred) fragments or reads.
* Most of the time, enriched region is trimmed to give a higher resolution event location (this would be the actual peak)
* Sometimes, regions/peaks are split up in post-processing (multiple nearby events)

Base pair level peak calling

* Typically two strategies:
* Find the number of fragments (usually Not reads) overlapping that position
  * need to go from reads to fragments
* Find the number of reads (fragment ends) reported at that position (possibly, taking strandedness into account)
* Very large selection of tools and techniques:
* ERANGE, FindPeaks, MACS, QuEST, CisGenome , SISSRS, USeq, PeakSeq, SPP, ChIPSeqR , GLITR, ChIPDiff, T-PIC, BayesPeak, MOSAiCS, CCAT, CSAR

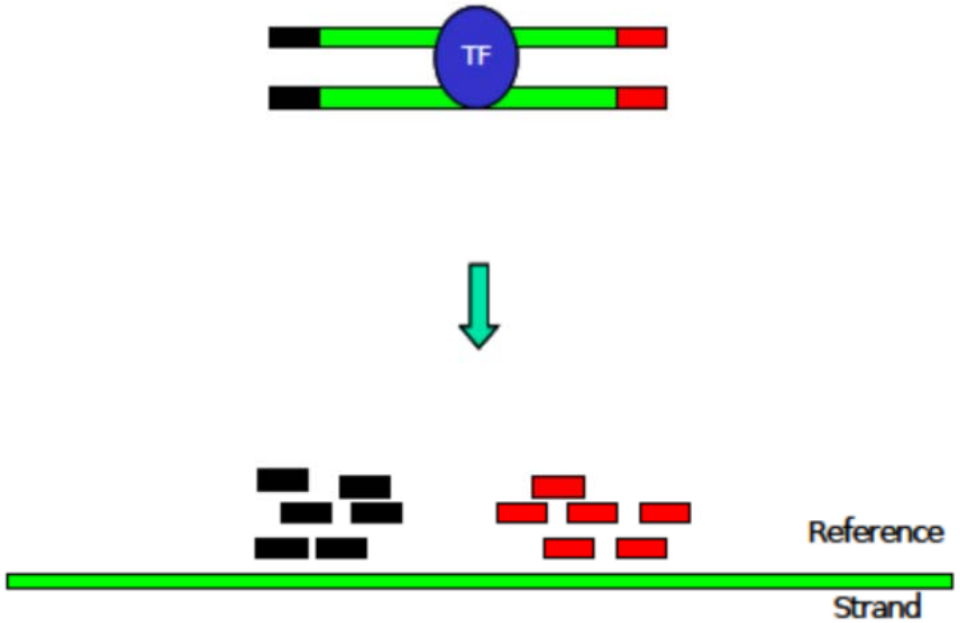
Fragments based



Number of fragments (usually NOT reads) overlapping that position

Slide modified from István Albert

Reads based



Slide modified from István Albert

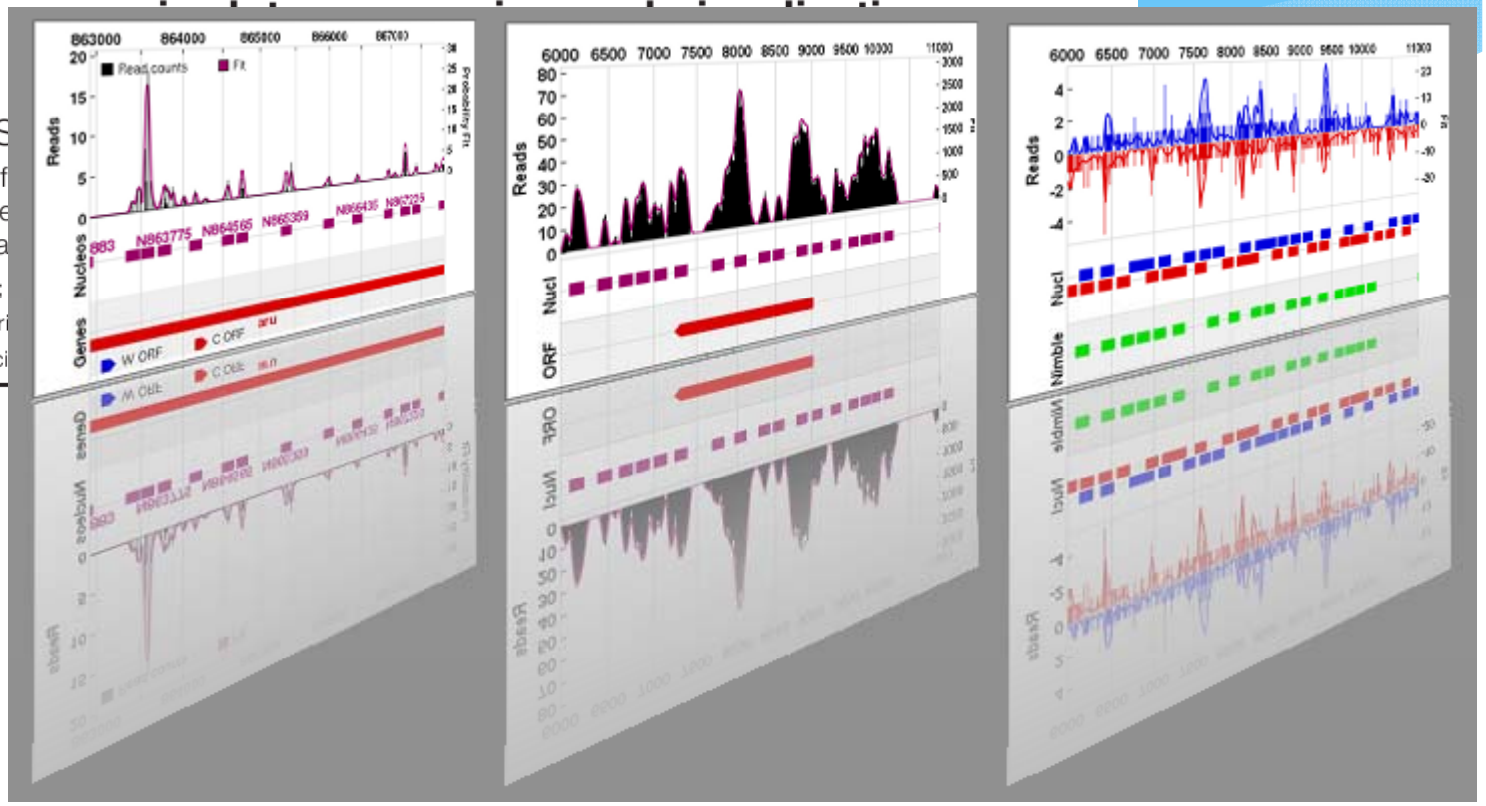
BIOINFORMATICS APPLICATIONS NOTE Vol. 24 no. 10 2008, pages 1305–1306
doi:10.1093/bioinformatics/btn119

Genome analysis

GeneTrack—a framework

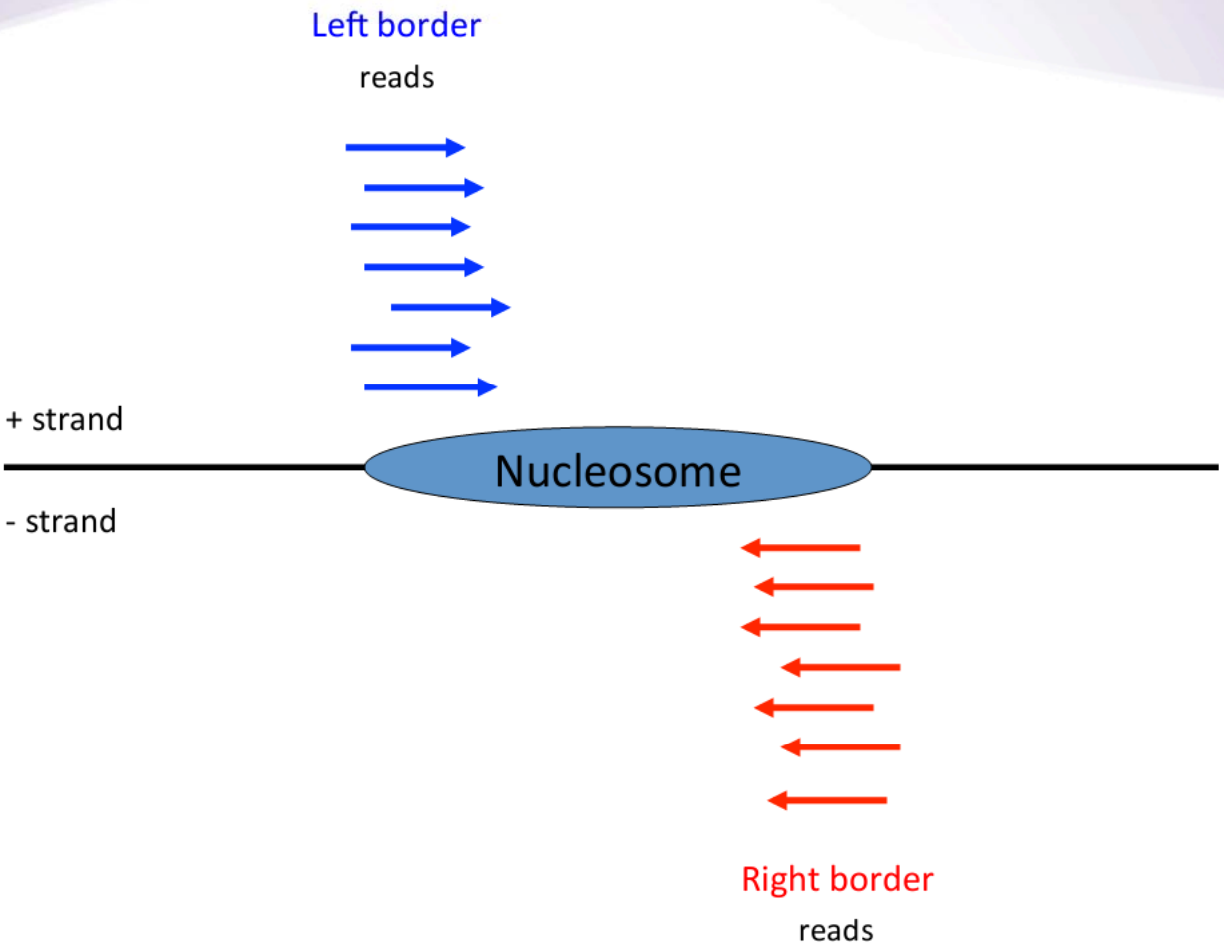
Istvan Albert[1,2,3,*], S
[1]Huck Institutes for the Lif
Regulation and [3]Departme
University Park, Pennsylva

Received on February 13, 2008;
Advance Access publication Apri
Associate Editor: Alfonso Valenci



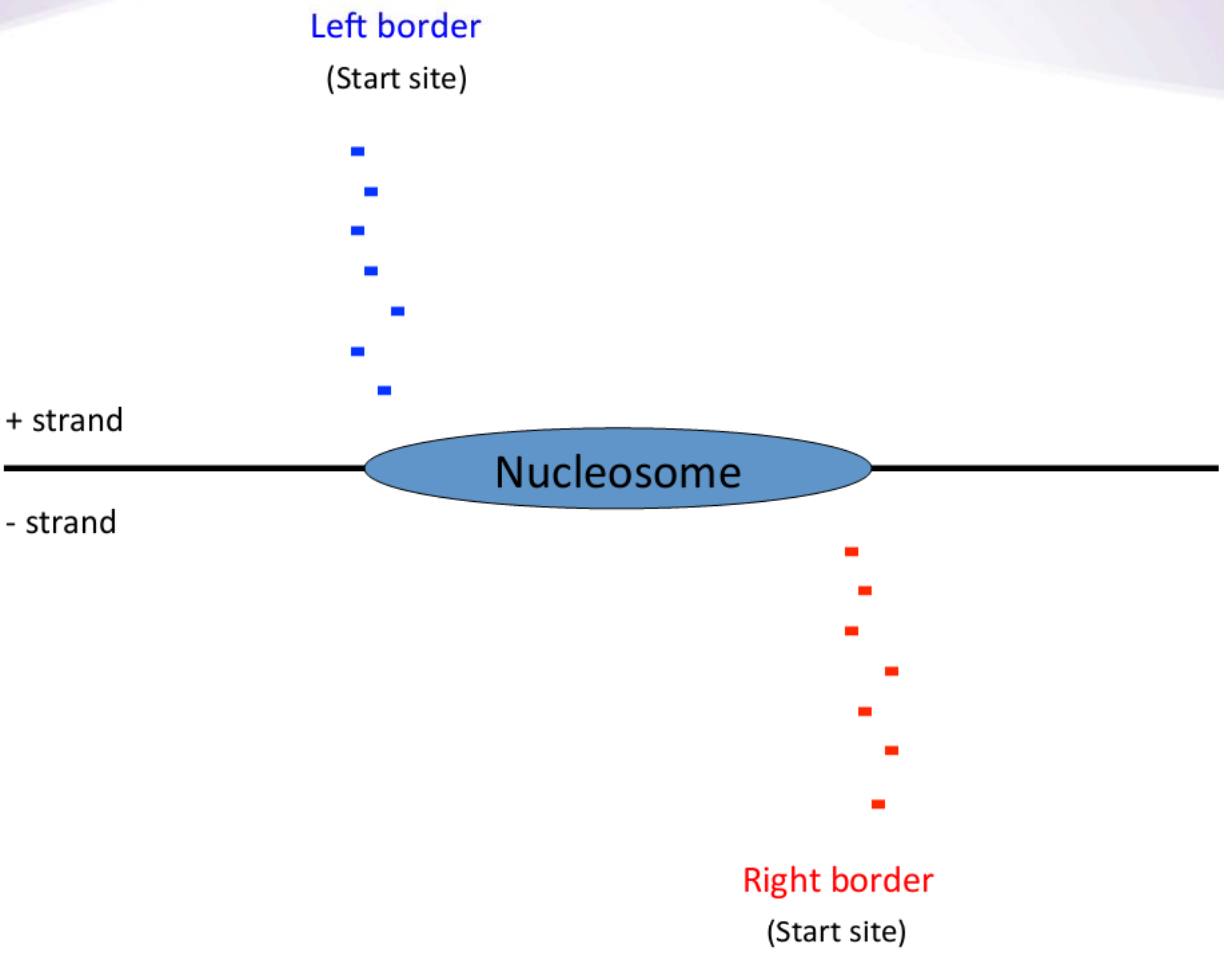
http://code.google.com/p/genetrack/

Aligning to reference genome



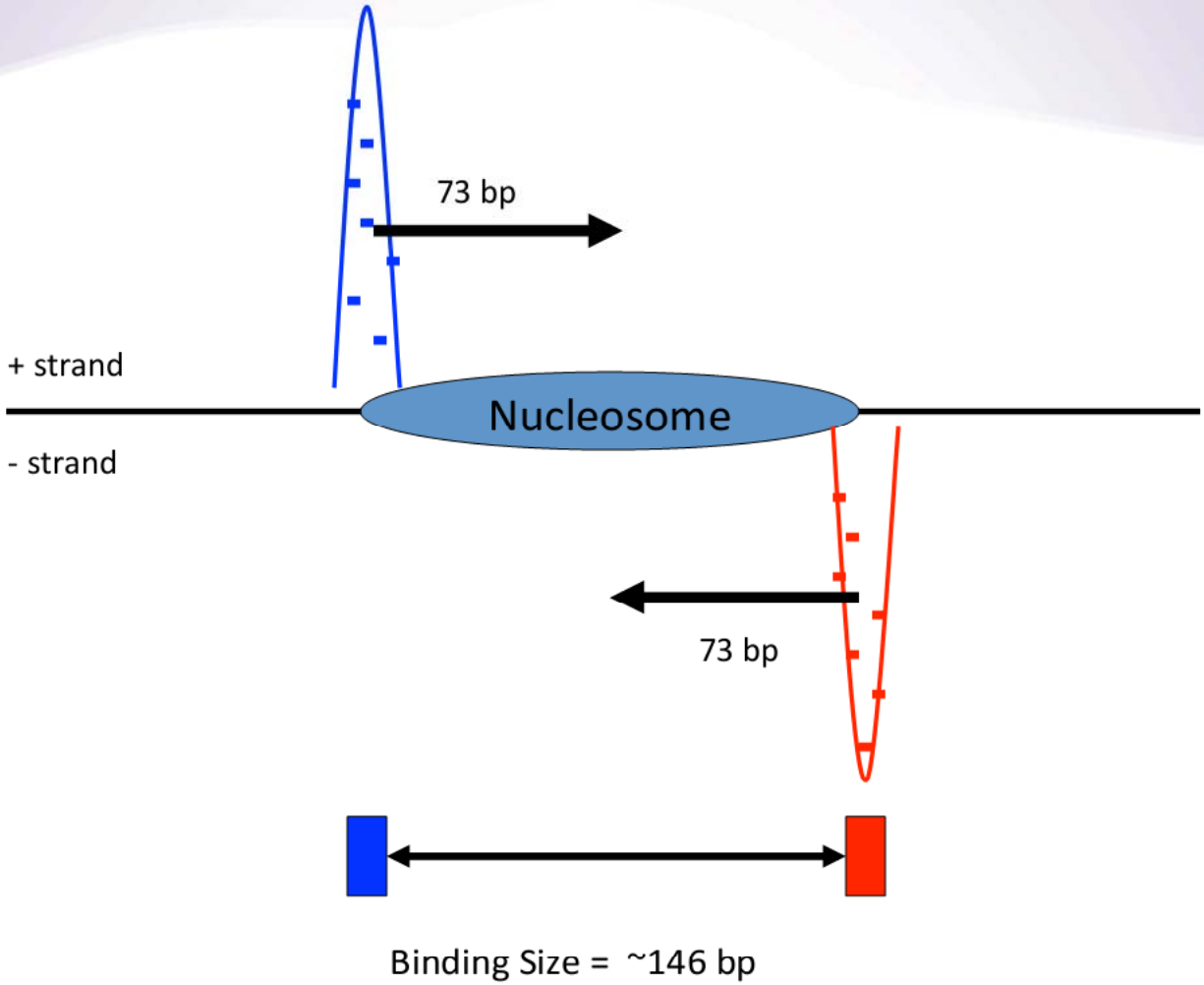
Slide modified from István Albert

Find the borders



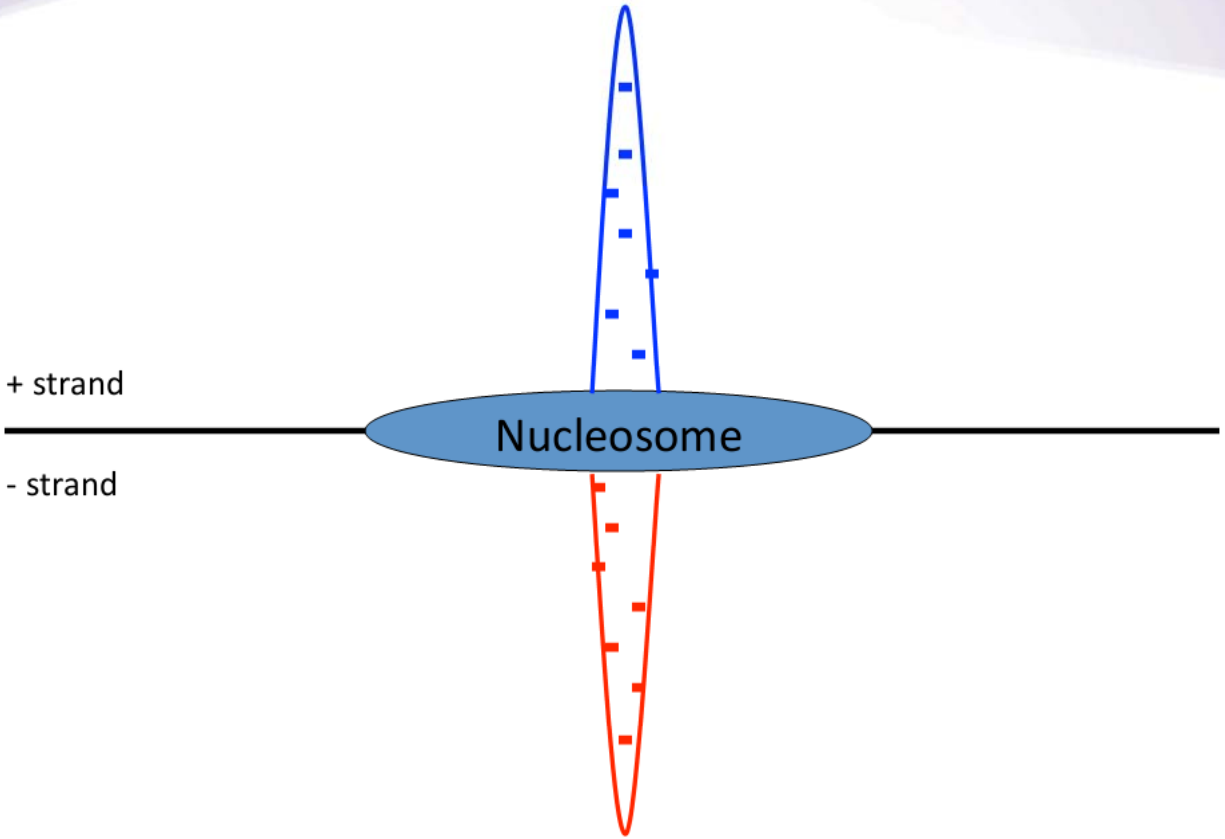
Slide modified from István Albert

Finding the binding sizes



Slide modified from István Albert

Shift to the center



Slide modified from István Albert

Enrichment measures

* Overlap approach: typically, the maximum overlap in the region is the measure
* Read count approach: typically, the total number of reads in the region is the measure
* Variation: calculate separate enrichment measures based on strand-specific reads.

Peak-Calling: Background

* No-model approach (no BG estimation)
  - Require enrichment > cutoff (user-specified)
  - E.g., number of reads in 1kb bin > 10 (arbitrary number).
  - Maybe use some other requirements (post-filtering)

  => No statistics can be done.

Peak-Calling: Background

* Model null distribution of enrichment values based on sample itself
  * Analytical
  * Empirical (simulation-based)

* Use significance measure (p-value, FDR) cutoff to retain regions

Peak-Calling: Background

* First assumption people made: the distribution of read/fragment start sites is uniform across genome (apart from event sites)
  * Poisson process with per-base rate = #(reads)/G
  * Variation: exclude non-mappable portion of genome from G (mappability depends on your alignment strategy, unresolved bases in genome assembly)
  * Variation: empirical null distribution based on simulations. This is more amenable to modifications
  * For any p-value/FDR, it is straightforward to calculate enrichment significance cutoffs for both count-based and overlap-based measures
* There is a problem: the distribution of read/fragment start sites is far from uniform as also seen in control samples (samples lacking enrichment due to event of interest)

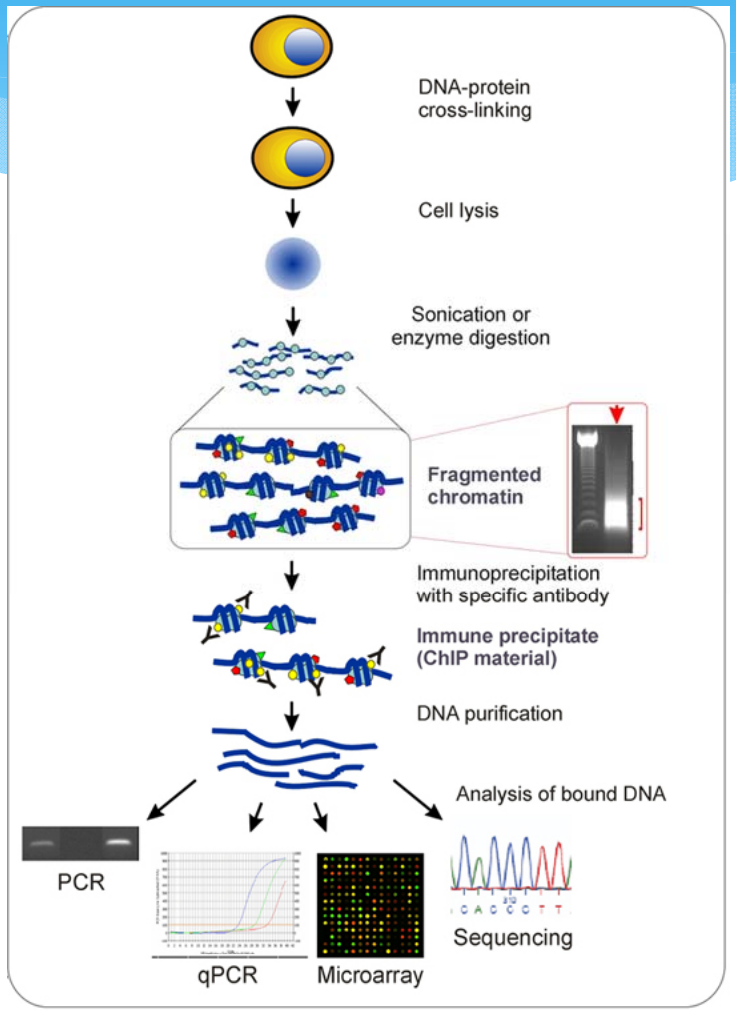
Non-Uniformity of ChIP Sample Background: Sequence features

* Some of this non-uniformity can be attributed to library prep/sequencing and alignment steps

* Mappability
  * Depending on alignment strategy, there can be structural 0's in data.
  * Paired-ends information helps mitigate this somewhat
  * Longer read lengths help to mitigate this too
* GC bias
  * Illumina-sequenced reads tend to be GC-rich
  * There are some protocol modifications that try to minimize this bias

negative controls

* Input DNA
* Non-specific antibody
* Different tissue



http://www.bioscience.org/2008/v13/af/2733/fulltext.asp?bframe=figures.htm&doi=yes

ChIP-Seq Negative Controls

- Used mostly to filter out false positives (high level noise), but can have other uses:
  - Idea: potential false positive will be enriched in both treatment and control.
- What kind of control to use?
  - 2 basic kinds of control:
    - One that undergoes IP (non-specific AB, different cell type)
    - One that doesn't (input DNA)
- Relying on input DNA control assumes that the IP step is bias-free (apart from bias towards regions of interest)

ChIP-Seq Negative Controls: Word of Caution

- Choice of control is extremely important, especially if its only purpose is filtering out false positives.
- A control will fail to filter out false positives if its enrichment profile is very different from the enrichment profile of false positive regions in the treatment sample
- A lot of thought should go into this.

Analysis with Control Sample

- Once you have obtained a control sample, several issues need to be addressed to incorporate it into analysis:
  - Normalization
  - Adjusting enrichment measures
  - Obtaining significance measures
- These are often not separate steps, but are intertwined together in their use of control sample

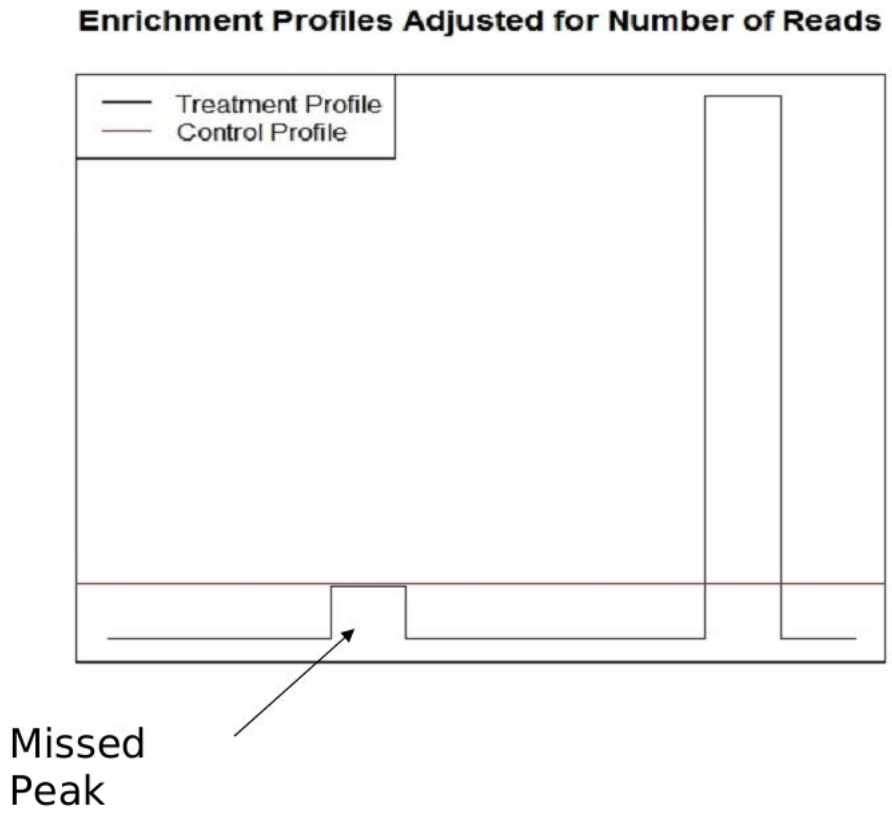
Analysis with Control Sample : Normalization

- Why normalize?
  - Usually, enrichment measures for region in treatment and control are compared and normalization is trying to adjust for differences between the samples due to effects other than enrichment for regions of interest in treatment sample
- Usual strategies:
  - Don’t normalize (not very common)
  - Adjust for difference in total lane counts
  - Adjust for difference in background profiles

Normalization: Adjusting for Total Lane Counts

- Reasoning: vast majority of treatment reads are from background, so enrichment profiles in 2 samples should scale with total number of reads.
  - This can perform poorly with treatment samples highly enriched for events of interest
- Approaches:
  - Subsample to have samples with equal numbers of reads
  - Use scaling ratio to adjust enrichment measure
    - E.g., scale read counts in regions before taking treatment/control ratio

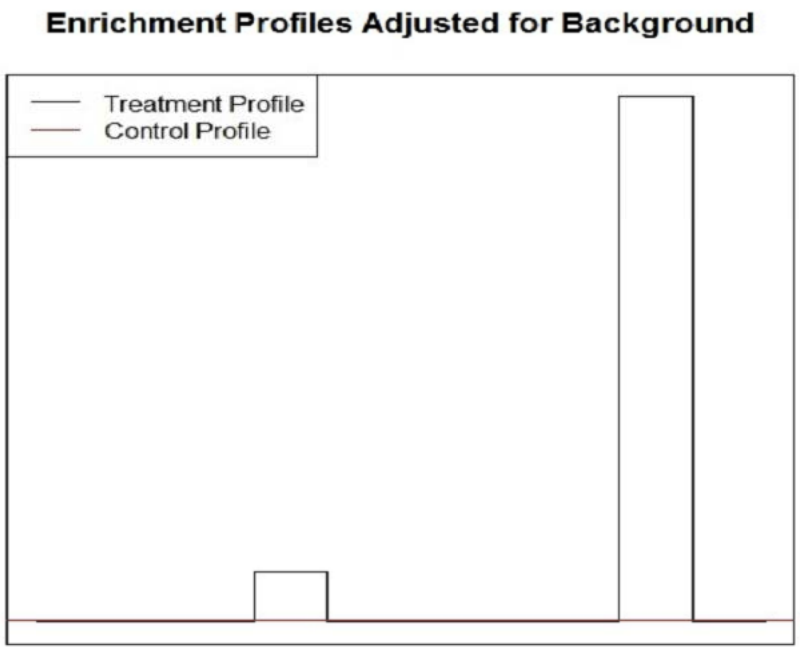
Total Lane Count Adjustment: Warning



Normalization: Adjusting for Difference in BG Profile

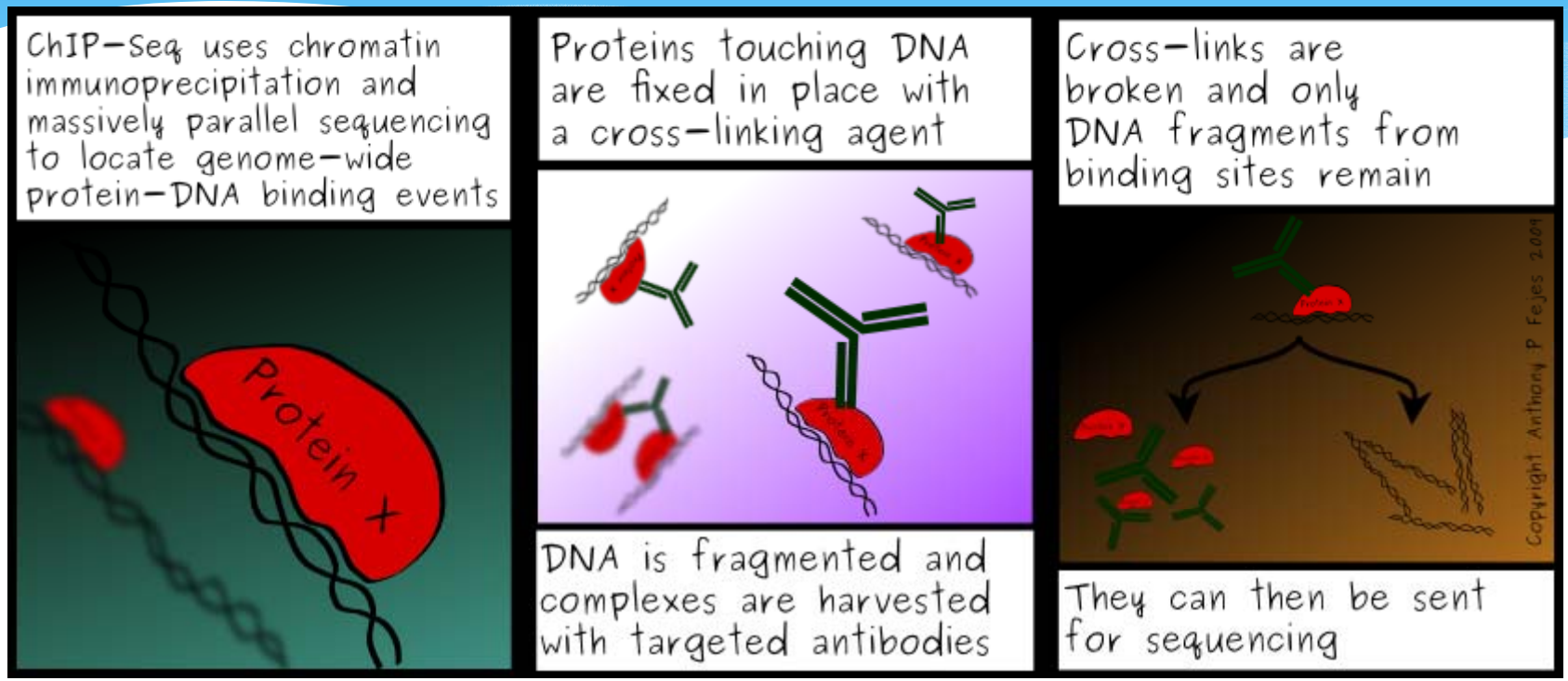
- Reasoning: instead of scaling based on total number of reads, which can lead to problems as seen above, scale BG profiles instead.
- Scale numbers of reads (either at bp level or region level) by scaling ratio to normalize
- Scaling ratio estimation:
  - Read ratios in LOW-COUNT regions
  - Linear Regression (omitting candidate peaks)
  - Heights of peaks in common
- Use scaling ratio in similar fashion as with total lane counts adjustments.

Difference in BG Profile Adjustment: Goal



Analysis with Control Sample: Obtaining Significance Measures

- There are several ways peak-callers use control sample information to assign significance to enrichment measures
- Enrichment measure itself can be a measure of significance (p-values, sometimes corrected for multiple hypothesis testing)
- Often, FDR is obtained as (estimated false positives)/(all positives) for each enrichment measure cutoff, where “all positives” are the peaks with enrichment measures exceeding the cutoff. What about “estimated false positives”?


ChIP-Seq uses chromatin immunoprecipitation and massively parallel sequencing to locate genome-wide protein-DNA binding events
Protein X
Proteins touching DNA are fixed in place with a cross-linking agent
DNA is fragmented and complexes are harvested with targeted antibodies
Cross-links are broken and only DNA fragments from binding sites remain
Copyright Anthony P Fejes 2009
They can then be sent for sequencing

Examples

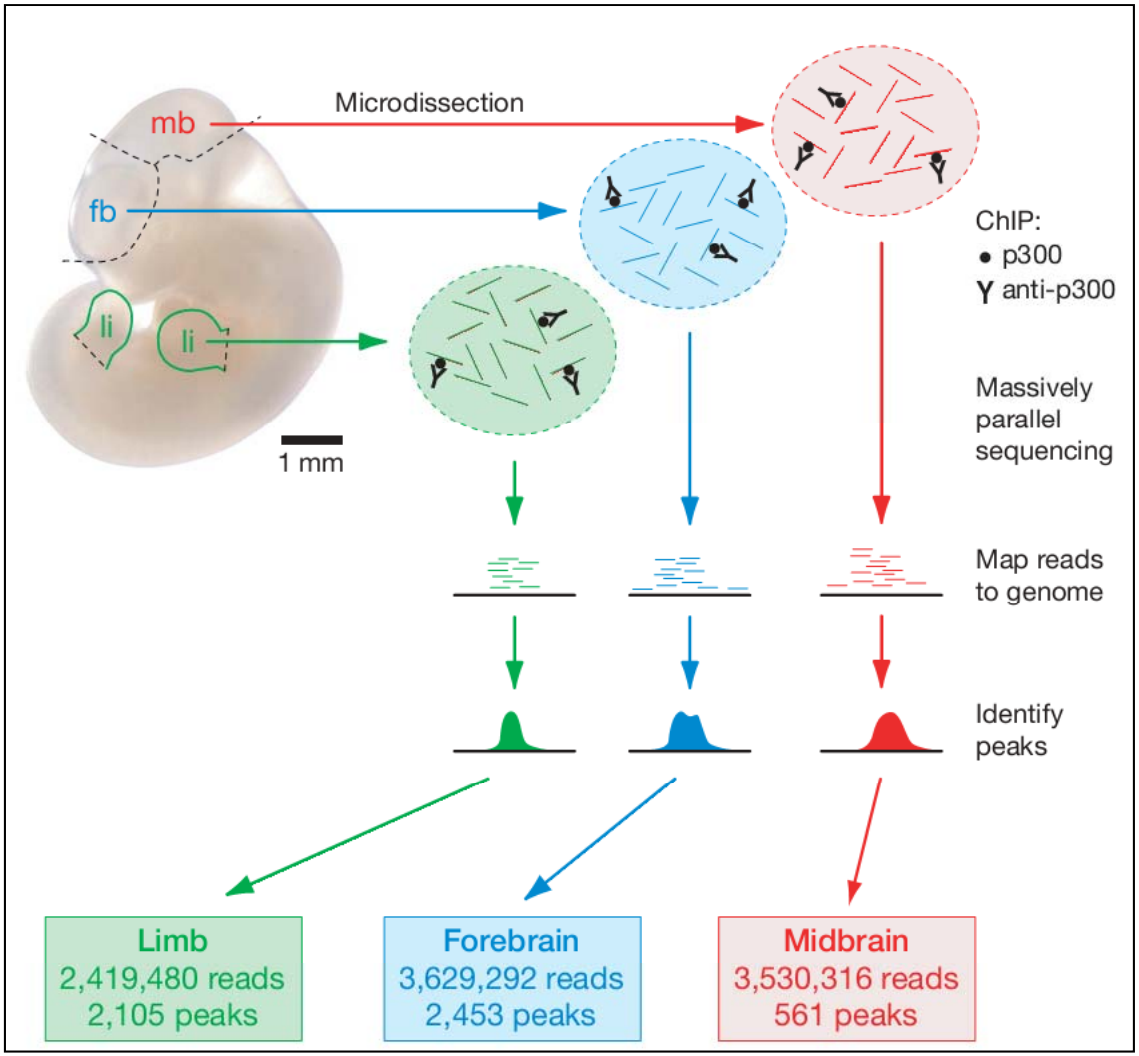
nature Vol 457|12 February 2009|doi:10.1038/nature07730

ARTICLES

ChIP-seq accurately predicts tissue-specific activity of enhancers

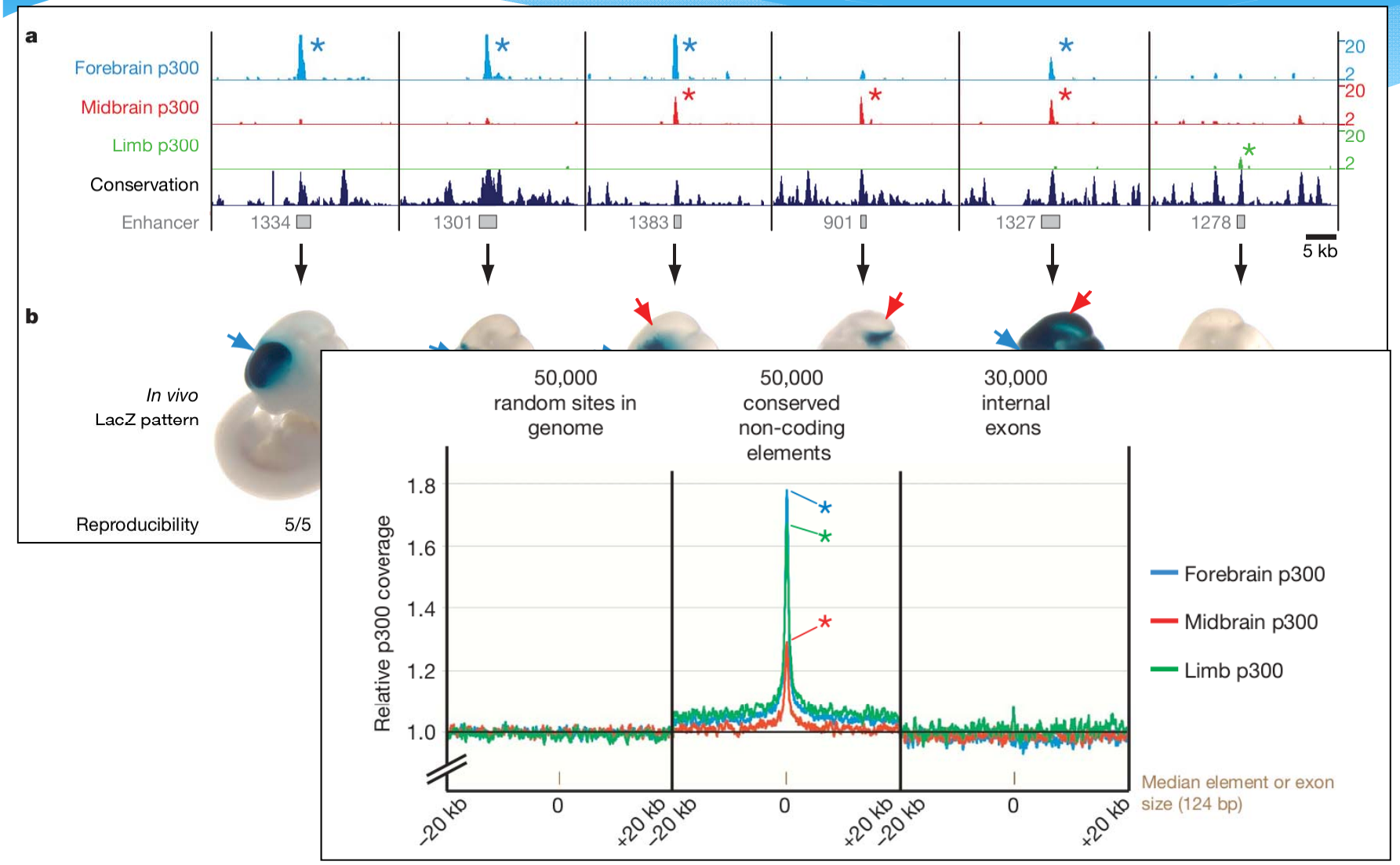
Axel Visel[1]*, Matthew J. Blow[1,2]*, Zirong Li[3], Tao Zhang[2], Jennifer A. Akiyama[1], Amy Holt[1], Ingrid Plajzer-Frick[1], Malak Shoukry[1], Crystal Wright[2], Feng Chen[2], Veena Afzal[1], Bing Ren[3], Edward M. Rubin[1,2] & Len A. Pennacchio[1,2]

A major yet unresolved quest in decoding the human genome is the identification of the regulatory sequences that control the spatial and temporal expression of genes. Distant-acting transcriptional enhancers are particularly challenging to uncover because they are scattered among the vast non-coding portion of the genome. Evolutionary sequence constraint can facilitate the discovery of enhancers, but fails to predict when and where they are active *in vivo*. Here we present the results of chromatin immunoprecipitation with the enhancer-associated protein p300 followed by massively parallel sequencing, and map several thousand *in vivo* binding sites of p300 in mouse embryonic forebrain, midbrain and limb tissue. We tested 86 of these sequences in a transgenic mouse assay, which in nearly all cases demonstrated reproducible enhancer activity in the tissues that were predicted by p300 binding. Our results indicate that *in vivo* mapping of p300 binding is a highly accurate means for identifying enhancers and their associated activities, and suggest that such data sets will be useful to study the role of tissue-specific enhancers in human biology and disease on a genome-wide scale.



fb, forebrain; li, limb; mb, midbrain

The acetyltransferase and transcriptional coactivator p300 is a near-ubiquitously expressed component of enhancer-associated protein assemblies and is critically required for embryonic development.


a
Forebrain p300
Midbrain p300
Limb p300
Conservation
Enhancer
1334
1301
1383
901
1327
1278
20
2
20
2
20
2
5 kb
b
In vivo
LacZ pattern
Reproducibility
5/5
50,000
random sites in
genome
50,000
conserved
non-coding
elements
30,000
internal
exons
Relative p300 coverage
1.8
1.6
1.4
1.2
1.0
-20 kb
0
+20 kb
-20 kb
0
+20 kb
-20 kb
0
+20 kb
Forebrain p300
Midbrain p300
Limb p300
Median element or exon
size (124 bp)

© 2008 Nature Publishing Group http://www.nature.com/naturemethods

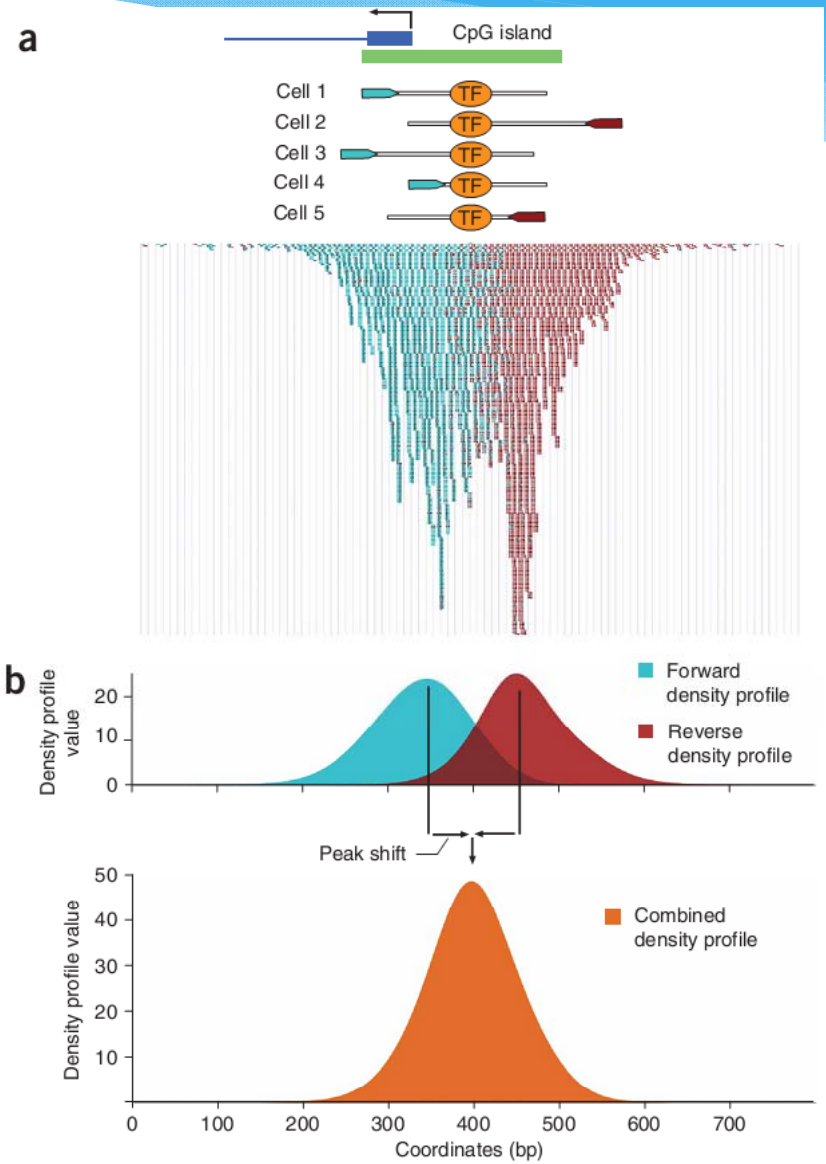
Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data

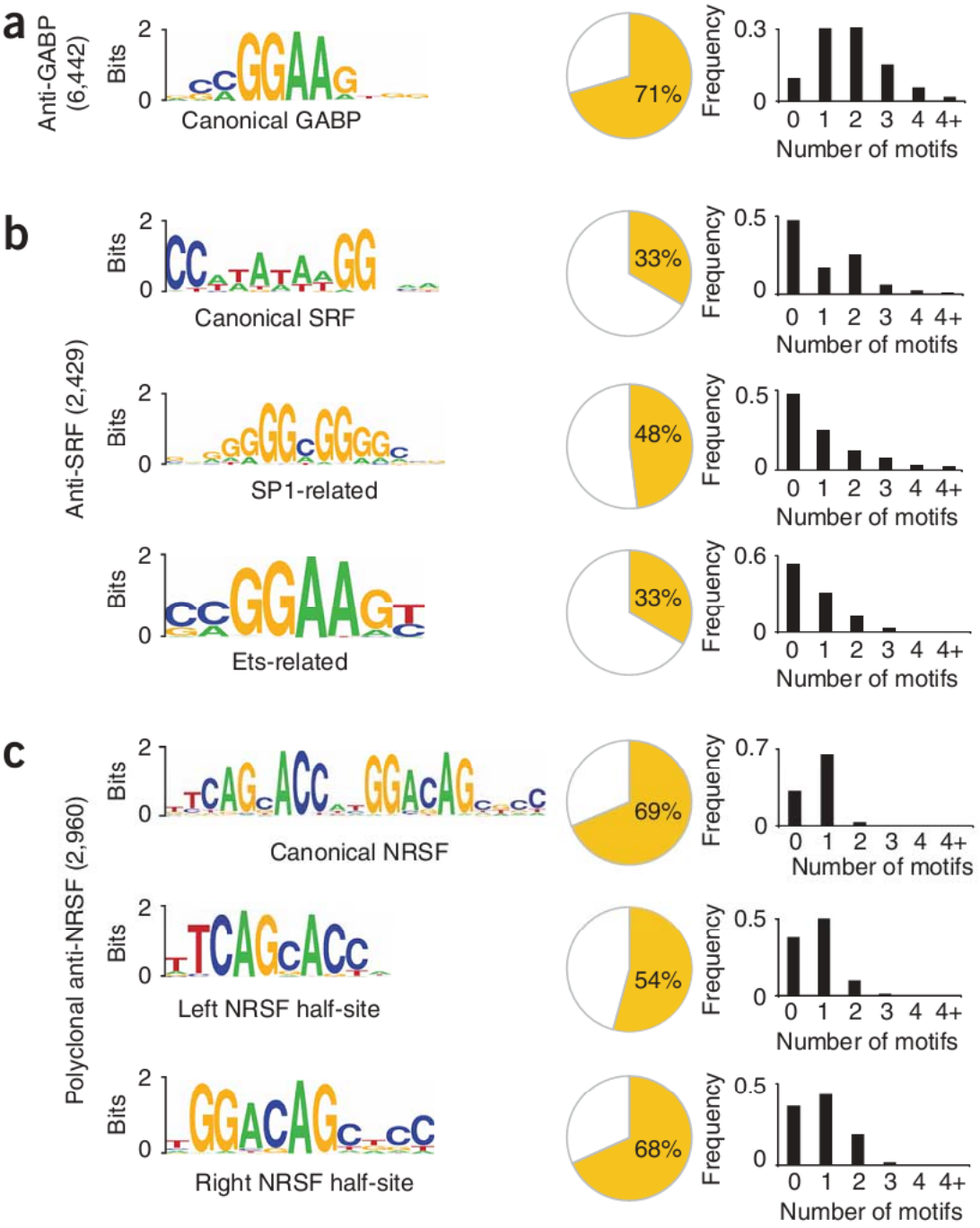
Anton Valouev[1,4], David S Johnson[2,4], Andreas Sundquist[3], Catherine Medina[2], Elizabeth Anton[2], Serafim Batzoglou[3], Richard M Myers[2] & Arend Sidow[1,2]

Molecular interactions between protein complexes and DNA mediate essential gene-regulatory functions. Uncovering such interactions by chromatin immunoprecipitation coupled with massively parallel sequencing (ChIP-Seq) has recently become the focus of intense interest. We here introduce quantitative enrichment of sequence tags (QuEST), a powerful statistical framework based on the kernel density estimation approach, which uses ChIP-Seq data to determine positions where protein complexes contact DNA. Using QuEST, we discovered several thousand binding sites for the human transcription factors SRF, GABP and NRSF at an average resolution of about 20 base pairs. MEME motif-discovery tool–based analyses of the QuEST-identified sequences revealed DNA binding by cofactors of SRF, providing evidence that cofactor binding specificity can be obtained from ChIP-Seq data. By combining QuEST analyses with Gene Ontology (GO) annotations and expression data, we illustrate how general functions of transcription factors can be inferred.

fragments. Regions of high sequencing read density are referred to as peaks. The output of software implementing peak-finding methodology is a list of 'peak calls' that comprises the genomic locations of sites inferred to be occupied by the protein. To date, studies that have presented ChIP-Seq data[5,6] used peak-finding methodology that heuristically quantifies read density but does not take full advantage of certain important properties of the data such as the directionality of sequencing reads. The growing importance of ChIP-Seq demands development of rigorous and transparent statistical approaches that fully leverage the inherent advantages of ChIP-Seq.

We here describe QuEST, a ChIP-Seq data analysis method that is based on realistic statistical modeling of the ChIP-Seq experimental approach. QuEST generates peak calls with substantial power and resolution by leveraging key attributes of the sequencing data such as directionality of reads and the size of fragments that were sequenced (which, notably, is estimated from the data themselves rather than provided by the user). QuEST achieves the desired balance between sensitivity and specificity by calculating false-


a
CpG island
Cell 1
Cell 2
Cell 3
Cell 4
Cell 5
TF
b
Density profile value
20
10
0
Forward density profile
Reverse density profile
Peak shift
50
40
30
20
10
Density profile value
Combined density profile
0
100
200
300
400
500
600
700
Coordinates (bp)



Growth-associated binding protein (GABP)

serum response factor (SRF)

neuron-restrictive silencer factor (NRSF)

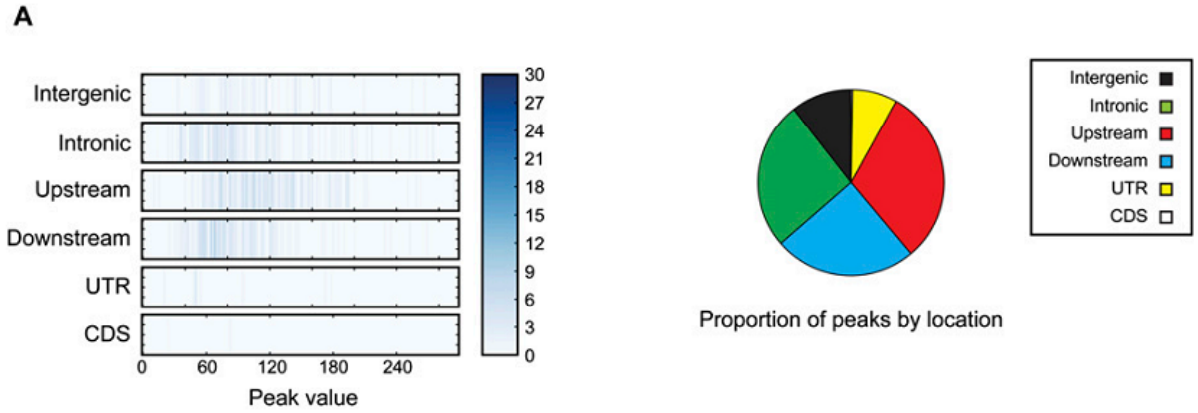
A ChIP-seq defined genome-wide map of vitamin D receptor binding: Associations with disease and evolution

Sreeram V. Ramagopalan,[1,2,3,6,7] Andreas Heger,[4,6] Antonio J. Berlanga,[1,2] Narelle J. Maugeri,[1] Matthew R. Lincoln,[1,2] Amy Burrell,[1,2] Lahiru Handunnetthi,[1,2] Adam E. Handel,[1,2] Giulio Disanto,[1,2] Sarah-Michelle Orton,[1,2] Corey T. Watson,[5] Julia M. Morahan,[1,2] Gavin Giovannoni,[3] Chris P. Ponting,[4] George C. Ebers,[1,2,7] and Julian C. Knight[1,7]

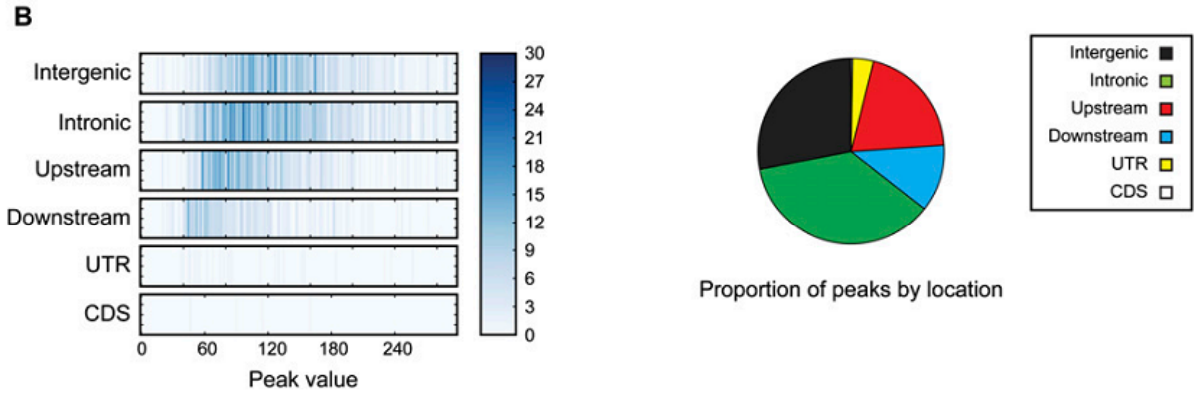
[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Headington, Oxford OX3 7BN, United Kingdom; [2]Department of Clinical Neurology, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, United Kingdom; [3]Blizard Institute of Cell and Molecular Science, Queen Mary University of London, Barts and The London School of Medicine and Dentistry, London E1 2AT, United Kingdom; [4]MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford OX1 3QX, United Kingdom; [5]Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

Initially thought to play a restricted role in calcium homeostasis, the pleiotropic actions of vitamin D in biology and their clinical significance are only now becoming apparent. However, the mode of action of vitamin D, through its cognate nuclear vitamin D receptor (VDR), and its contribution to diverse disorders, remain poorly understood. We determined VDR binding throughout the human genome using chromatin immunoprecipitation followed by massively parallel DNA sequencing (ChIP-seq). After calcitriol stimulation, we identified 2776 genomic positions occupied by the VDR and 229 genes with significant changes in expression in response to vitamin D. VDR binding sites were significantly enriched near autoimmune and cancer associated genes identified from genome-wide association (GWA) studies. Notable genes with VDR binding included *IRF8*, associated with MS, and *PTPN2* associated with Crohn's disease and TID. Furthermore, a number of single nucleotide polymorphism associations from GWA were located directly within VDR binding intervals, for example, rs13385731 associated with SLE and rs947474 associated with TID. We also observed significant enrichment of VDR intervals within regions of positive selection among individuals of Asian and European descent. ChIP-seq determination of transcription factor binding, in combination with GWA data, provides a powerful approach to further understanding the molecular bases of complex diseases.

20:1352–1360 © 2010 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/10; www.genome.org



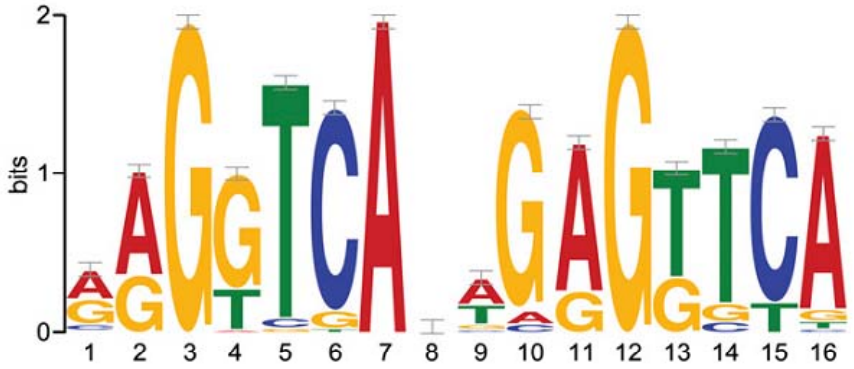
Unstimulated cells

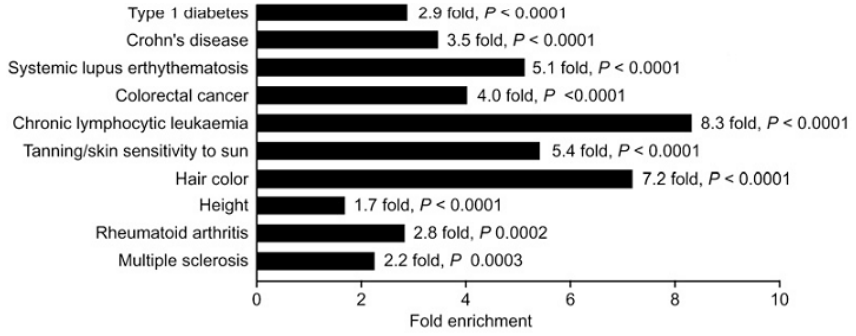


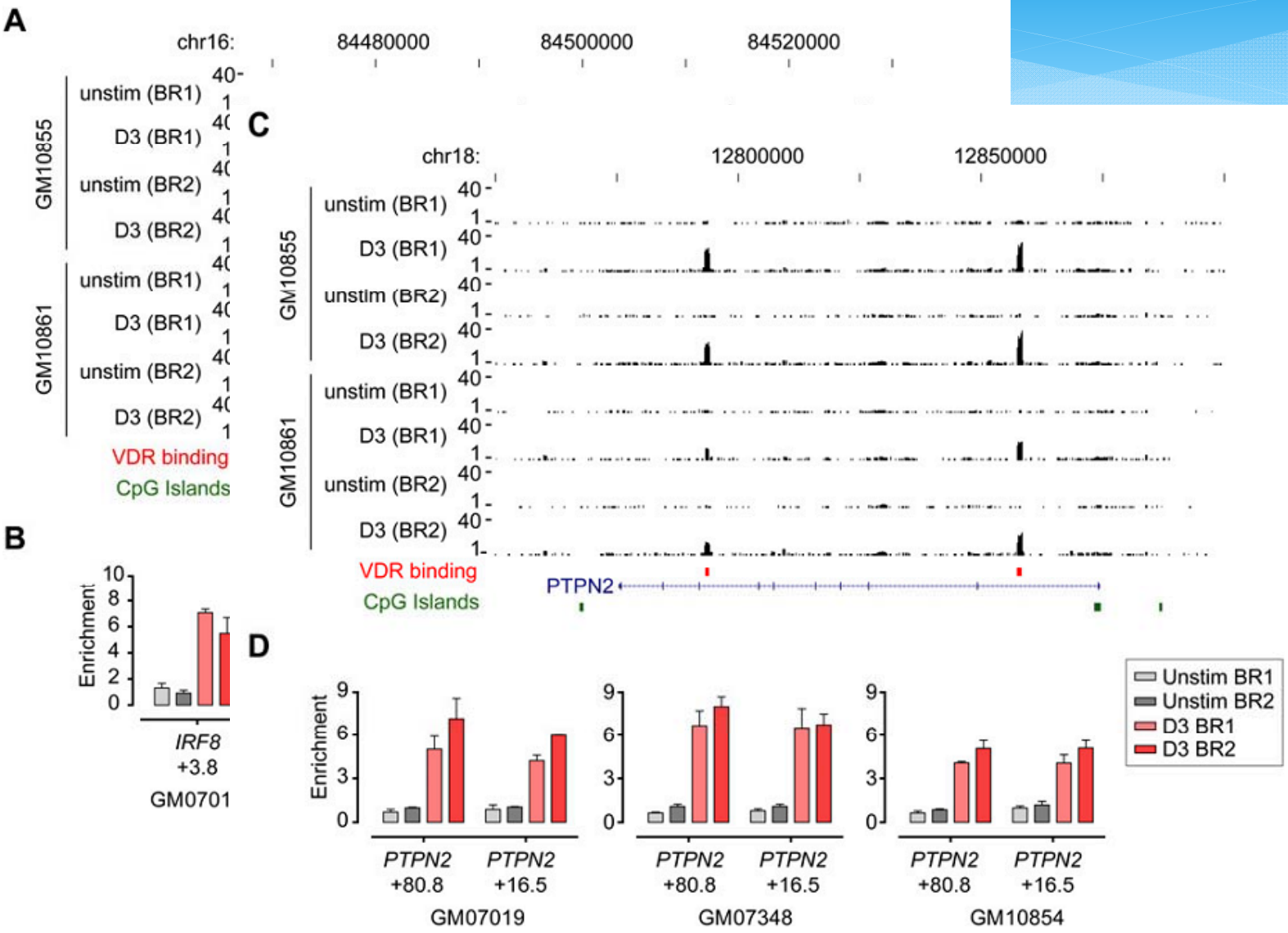
Calcitrol-stimulated cells

Table 1. Number of peaks called using different thresholds and average interval size

	No. of intervals	Average interval size (bp)
FDR 1%		
Calcitriol stimulated	2776	781
Unstimulated	623	1161
FDR 1%, 20-fold enriched		
Calcitriol stimulated	789	658
Unstimulated	57	1100






A
chr16:
84480000
84500000
84520000
GM10855
unstim (BR1)
D3 (BR1)
unstim (BR2)
D3 (BR2)
GM10861
unstim (BR1)
D3 (BR1)
unstim (BR2)
D3 (BR2)
VDR binding
CpG Islands
B
Enrichment
IRF8
+3.8
GM0701
C
chr18:
12800000
12850000
GM10855
unstim (BR1)
D3 (BR1)
unstim (BR2)
D3 (BR2)
GM10861
unstim (BR1)
D3 (BR1)
unstim (BR2)
D3 (BR2)
VDR binding
CpG Islands
PTPN2
D
Enrichment
PTPN2 +80.8
PTPN2 +16.5
GM07019
GM07348
GM10854
Unstim BR1
Unstim BR2
D3 BR1
D3 BR2

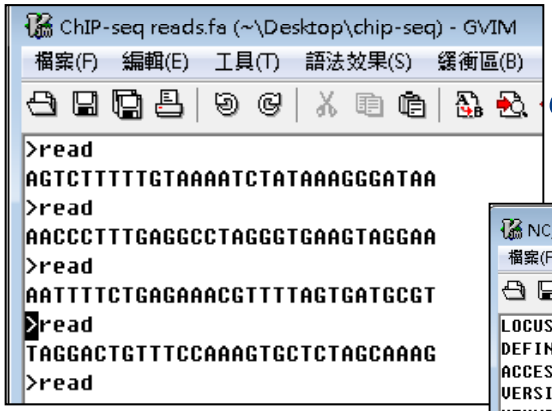
Part II

Chip-seq data analysis steps

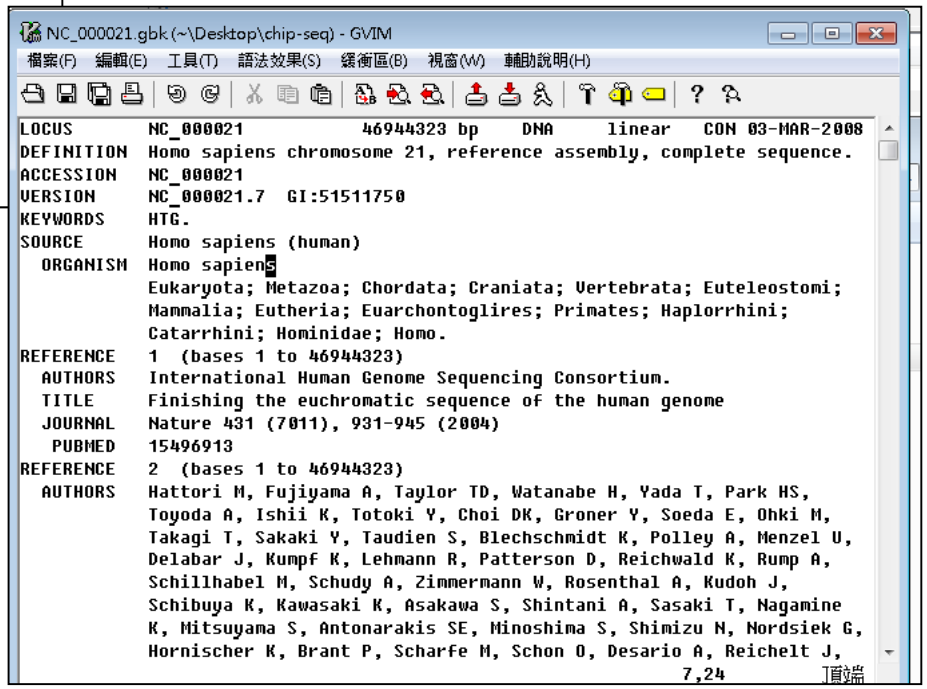
* import the data
* map the reads to a reference
* use the ChIP sequencing tool to detect significant peaks in the sample.

* wget http://192.168.75.28/class/chipseq/ChIP-seq%20reads%20-%20subset.fa
* wget http://192.168.75.28/class/chipseq/NC_000073.gbk
* wget http://192.168.75.28/class/chipseq/Mouse_Reads_subset.fa
* wget http://192.168.75.28/class/chipseq/NC_000021%20-%20subset.gbk

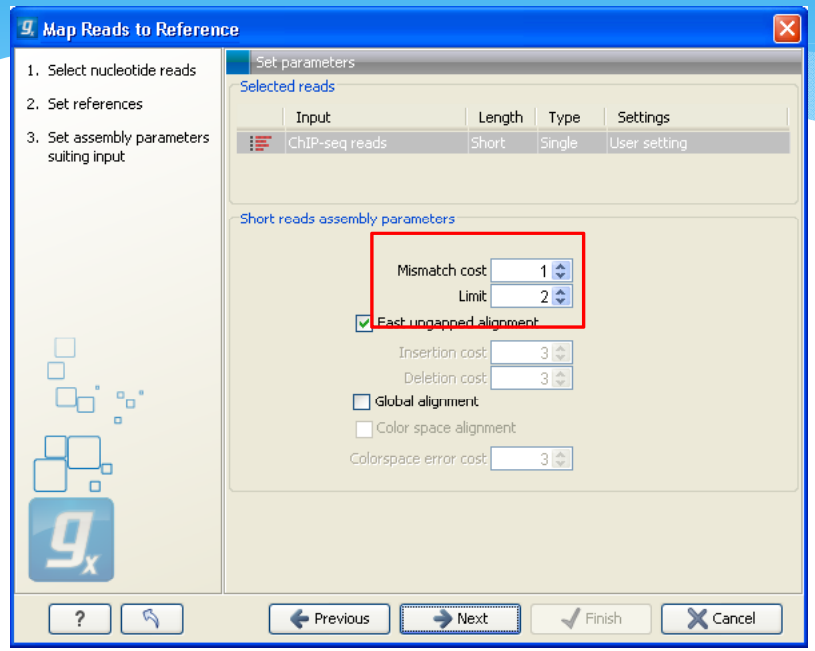
Input

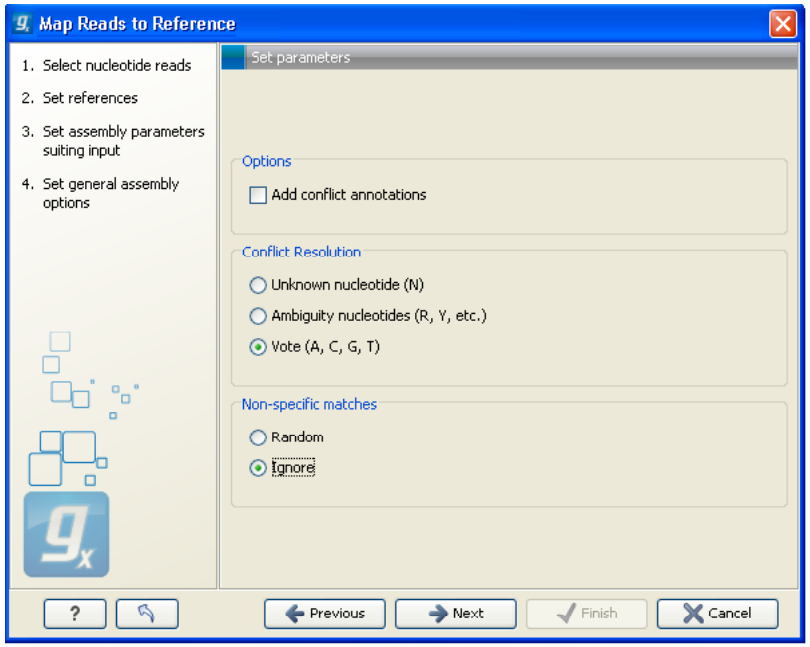


& reference from:

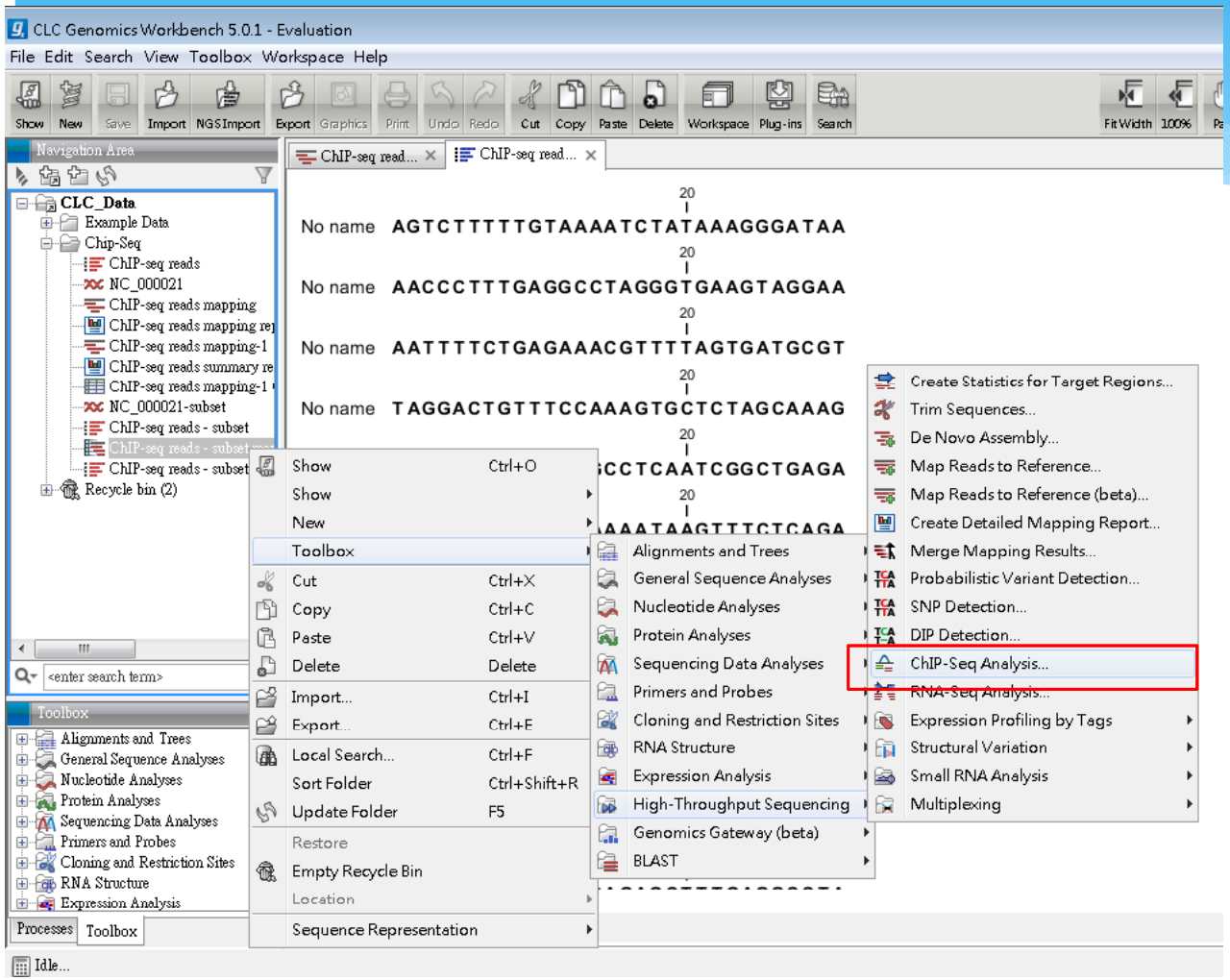


map the reads to a reference

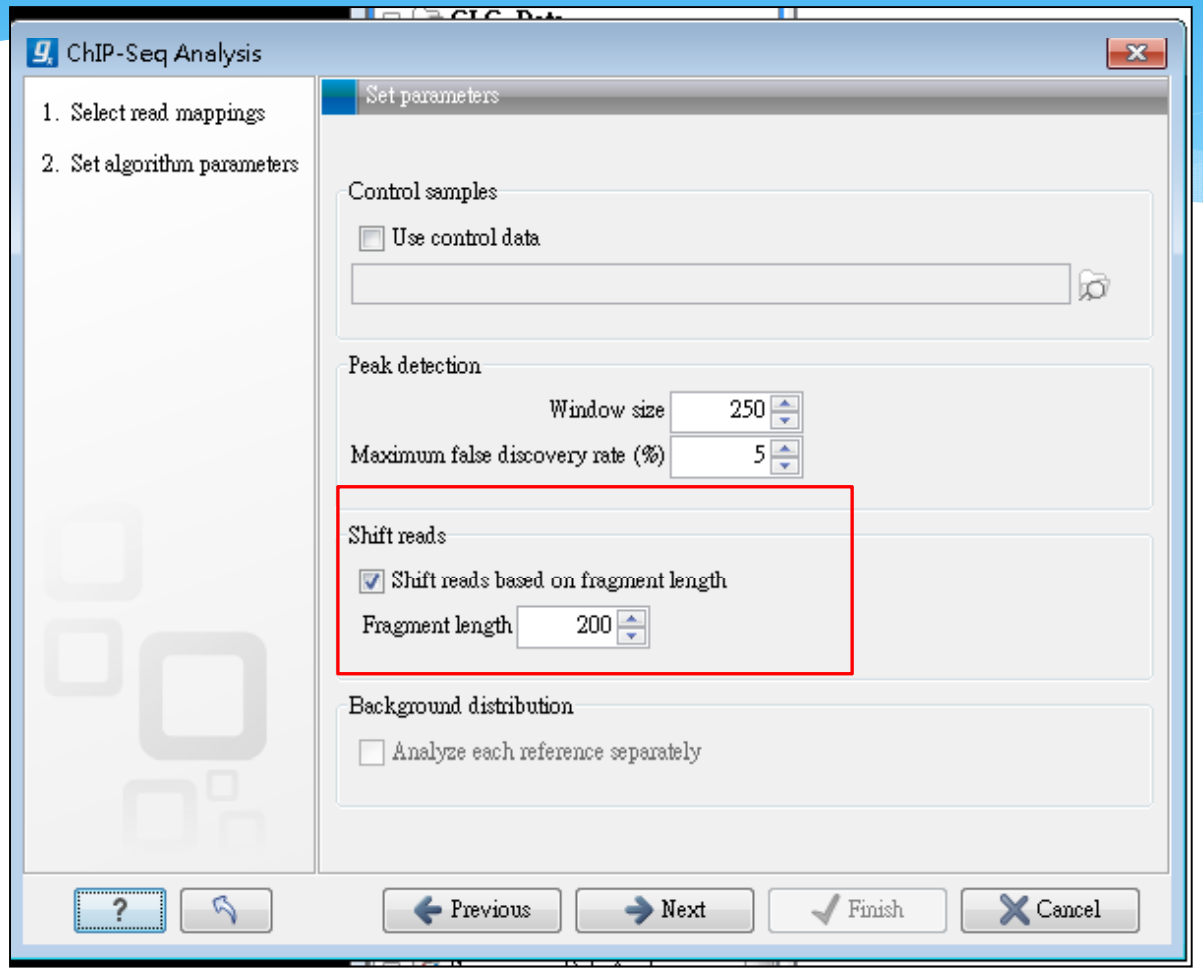




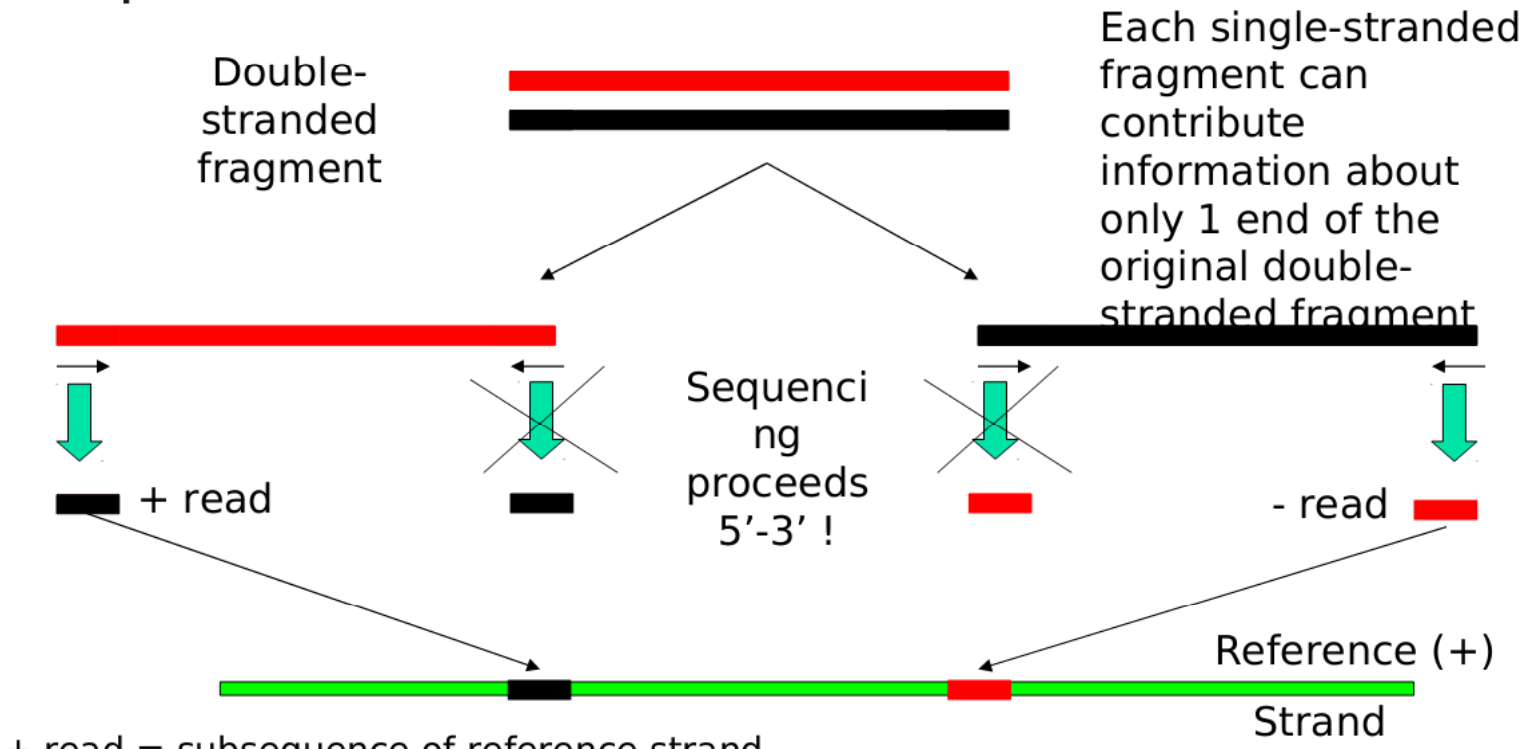
detect significant peaks



parameters

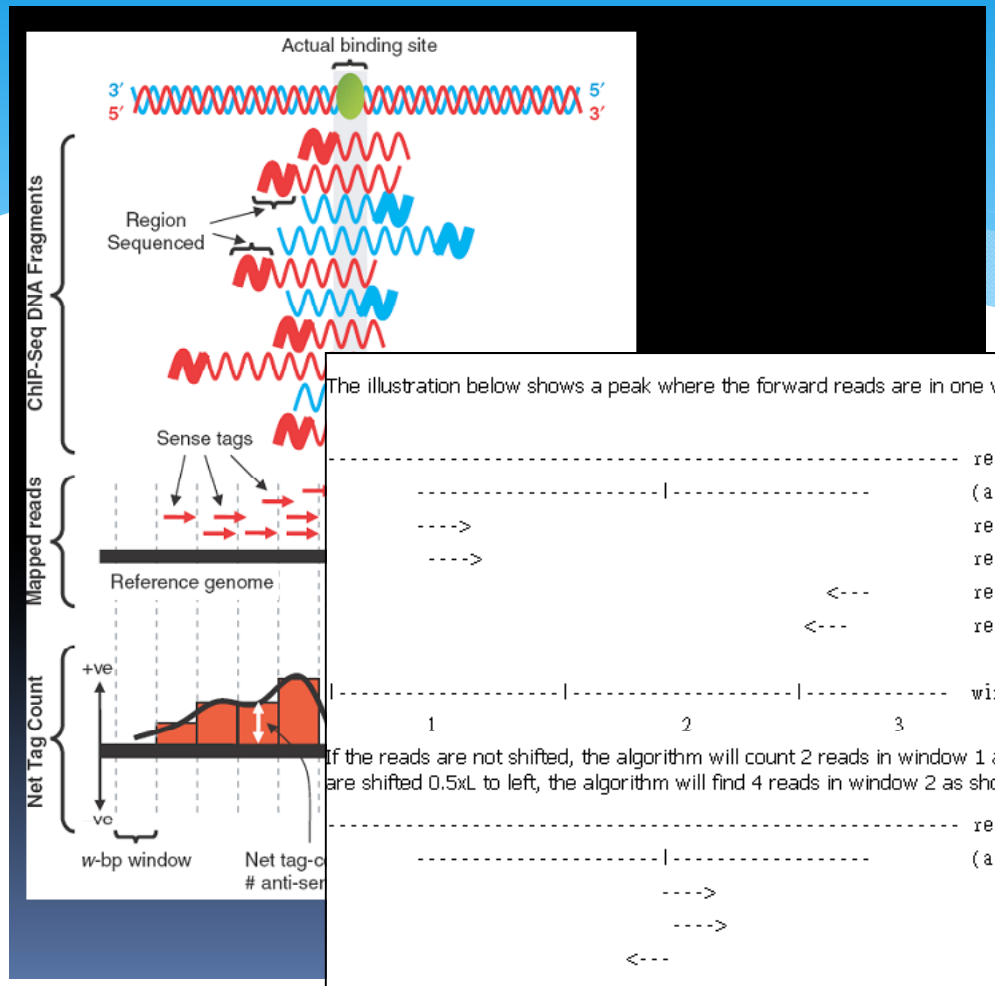

ChIP-Seq Analysis
1. Select read mappings
2. Set algorithm parameters
Set parameters
Control samples
Use control data
Peak detection
Window size
250
Maximum false discovery rate (%)
5
Shift reads
Shift reads based on fragment length
Fragment length
200
Background distribution
Analyze each reference separately
Previous
Next
Finish
Cancel

Sequencing: Single End



+ read = subsequence of reference strand

- read = reverse complement of subsequence of reference strand



The illustration below shows a peak where the forward reads are in one window and the reverse reads fall in another window (window 1 and 3).

```
---------------------------------------------------------- reference
        ----------------------|------------------          (actual sequenced fragment length = L bp)
        ---->                                               reads
         ---->                                              reads
                                              <---          reads
                                            <---            reads.

|--------------------|--------------------|-------------   window size W
         1                    2                   3
```

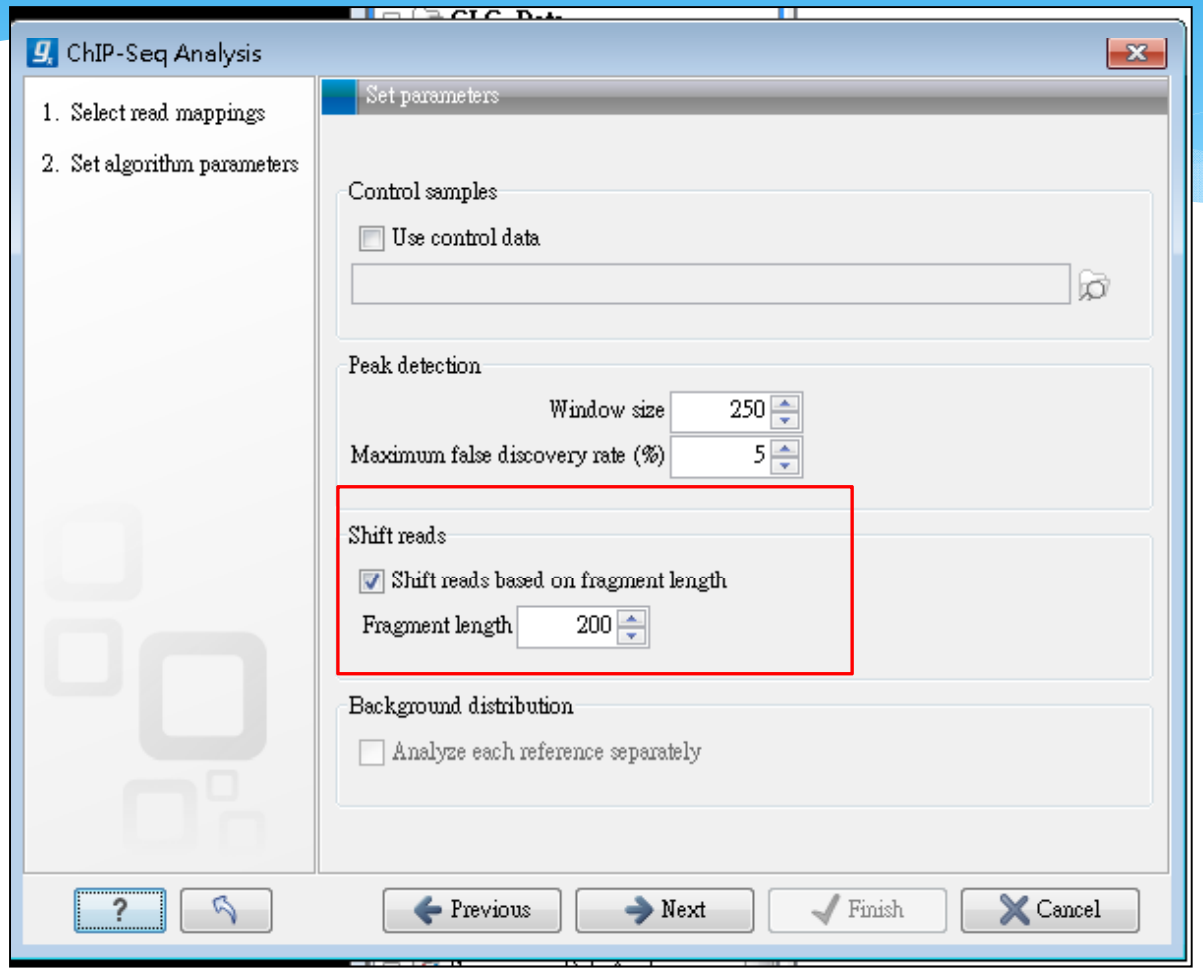
If the reads are not shifted, the algorithm will count 2 reads in window 1 and 3. But if the forward reads are shifted 0.5XL to the right and reverse reads are shifted 0.5xL to left, the algorithm will find 4 reads in window 2 as shown below:

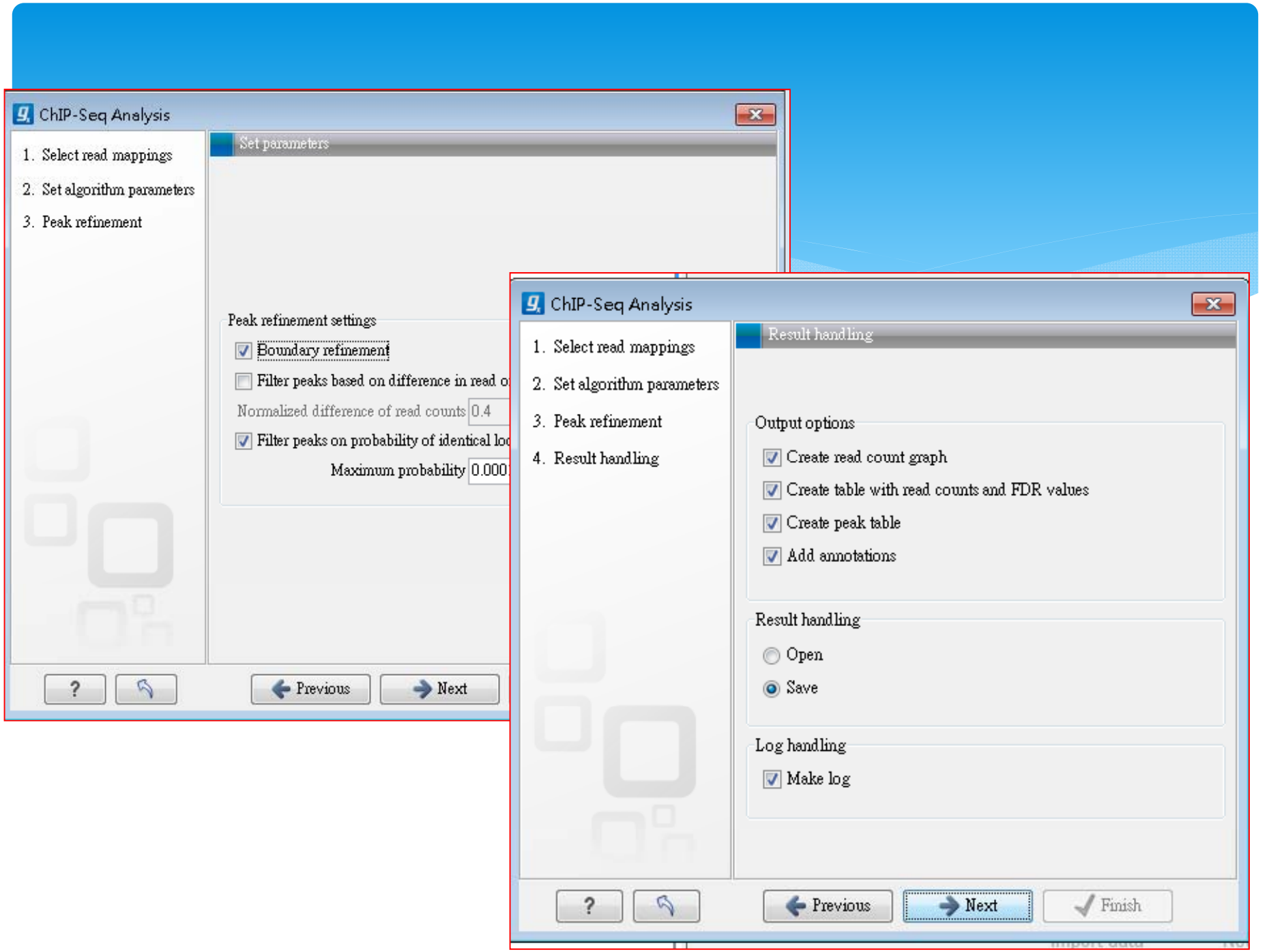
```
---------------------------------------------------------- reference
        ----------------------|------------------          (actual sequenced fragment length = L basepairs)
                                ---->                         reads
                                 ---->                        reads
                              <---                            reads
                             <---                             reads

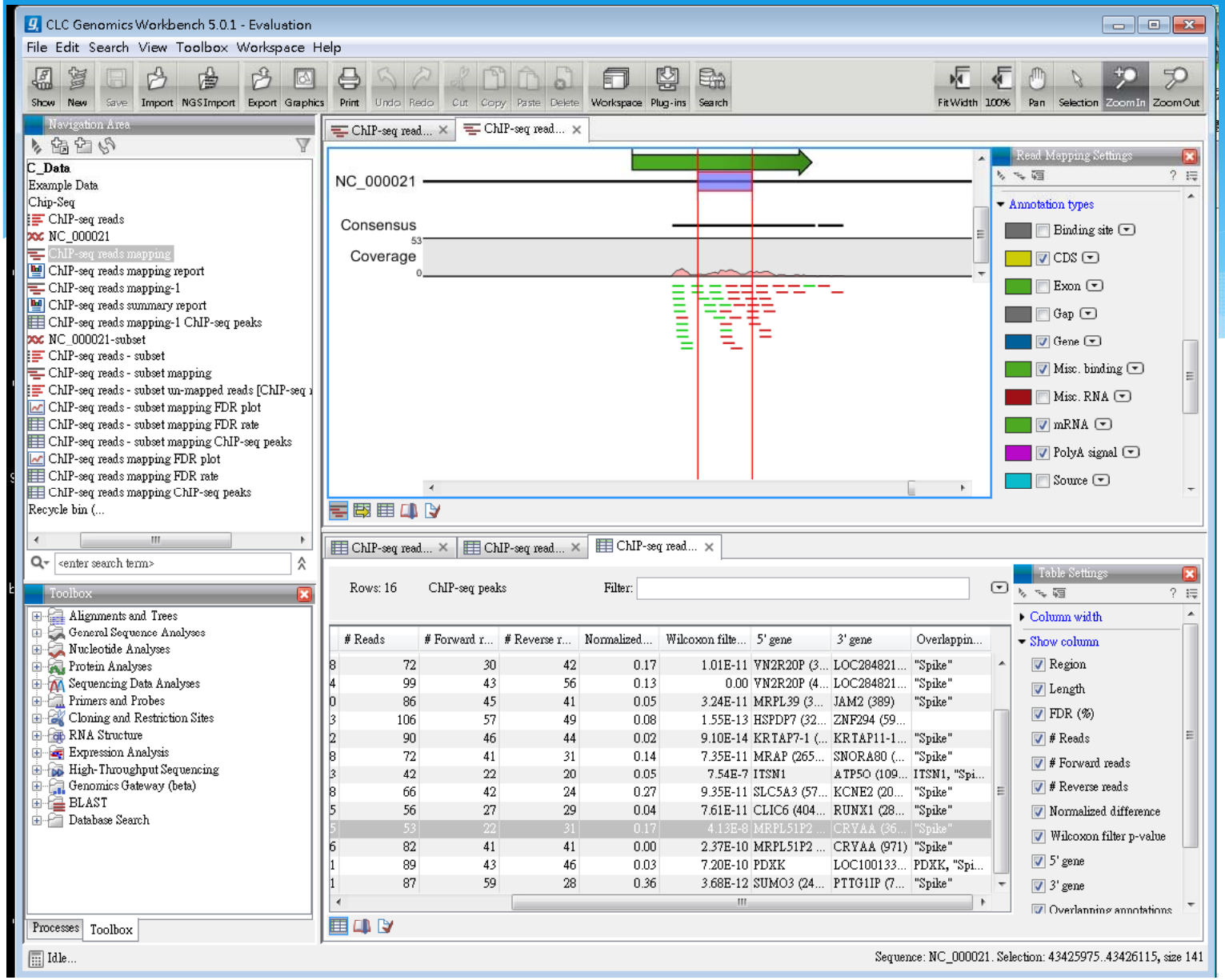
|--------------------|--------------------|-------------   window size W
         1                    2                   3
```

So shifting reads will increase the signal to noise ratio.

parameters


ChIP-Seq Analysis
1. Select read mappings
2. Set algorithm parameters
Set parameters
Control samples
Use control data
Peak detection
Window size
250
Maximum false discovery rate (%)
5
Shift reads
Shift reads based on fragment length
Fragment length
200
Background distribution
Analyze each reference separately
Previous
Next
Finish
Cancel


ChIP-Seq Analysis
1. Select read mappings
2. Set algorithm parameters
3. Peak refinement
Set parameters
Peak refinement settings
Boundary refinement
Filter peaks based on difference in read o
Normalized difference of read counts 0.4
Filter peaks on probability of identical loc
Maximum probability 0.000
Previous
Next
ChIP-Seq Analysis
1. Select read mappings
2. Set algorithm parameters
3. Peak refinement
4. Result handling
Result handling
Output options
Create read count graph
Create table with read counts and FDR values
Create peak table
Add annotations
Result handling
Open
Save
Log handling
Make log
Previous
Next
Finish



	# Reads	# Forward r...	# Reverse r...	Normalized...	Wilcoxon filte...	5' gene	3' gene	Overlappin...
8	72	30	42	0.17	1.01E-11	VN2R20P (3...	LOC284821...	"Spike"
4	99	43	56	0.13	0.00	VN2R20P (4...	LOC284821...	"Spike"
0	86	45	41	0.05	3.24E-11	MRPL39 (3...	JAM2 (389)	"Spike"
3	106	57	49	0.08	1.55E-13	HSPDP7 (32...	ZNF294 (59...	
2	90	46	44	0.02	9.10E-14	KRTAP7-1 (...	KRTAP11-1...	"Spike"
8	72	41	31	0.14	7.35E-11	MRAP (265...	SNORA80 (...	"Spike"
3	42	22	20	0.05	7.54E-7	ITSN1	ATP5O (109...	ITSN1, "Spi...
8	66	42	24	0.27	9.35E-11	SLC5A3 (57...	KCNE2 (20...	"Spike"
5	56	27	29	0.04	7.61E-11	CLIC6 (404...	RUNX1 (28...	"Spike"
5	53	22	31	0.17	4.13E-8	MRPL51P2 ...	CRYAA (36...	"Spike"
6	82	41	41	0.00	2.37E-10	MRPL51P2 ...	CRYAA (971)	"Spike"
1	89	43	46	0.03	7.20E-10	PDXK	LOC100133...	PDXK, "Spi...
1	87	59	28	0.36	3.68E-12	SUMO3 (24...	PTTG1IP (7...	"Spike"

practices

Data resource

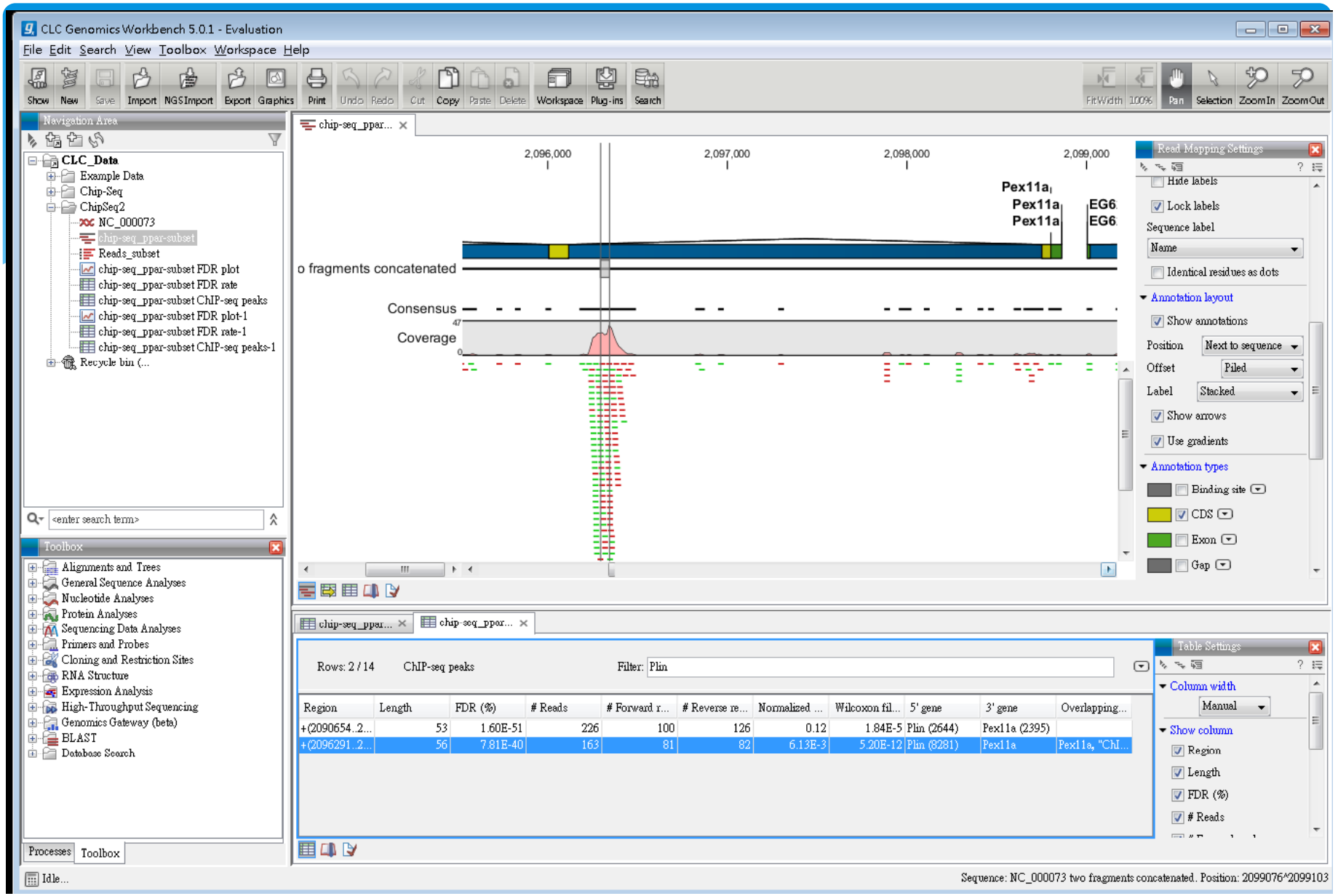


Genome-wide profiling of PPARγ:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis

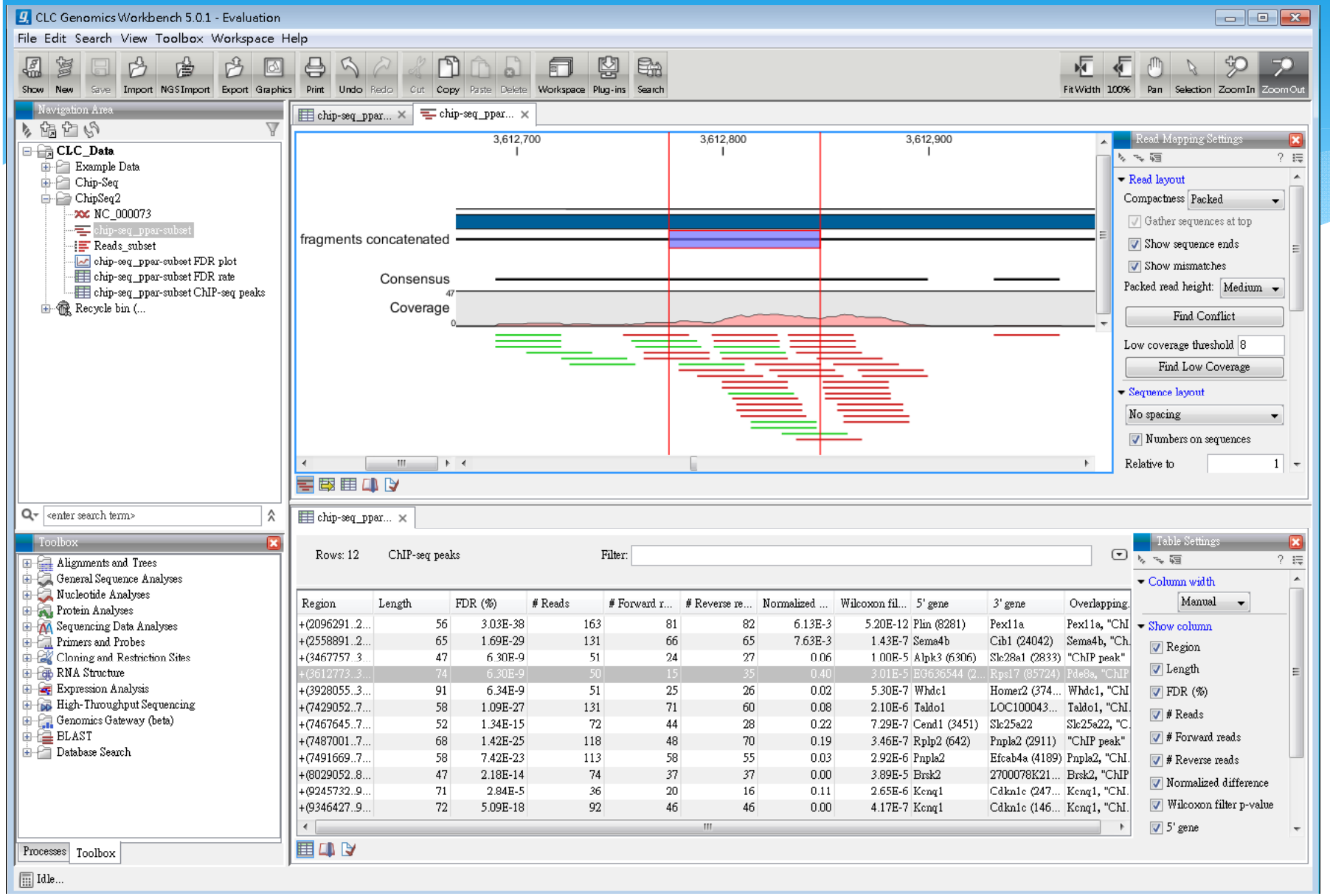
Ronni Nielsen, Thomas Åskov Pedersen, Dik Hagenbeek, et al.

Genes Dev. 2008 22: 2953-2967
Access the most recent version at doi:10.1101/gad.501108

Supplemental Material	http://genesdev.cshlp.org/content/suppl/2008/11/06/22.21.2953.DC1.html
References	This article cites 52 articles, 32 of which can be accessed free at: http://genesdev.cshlp.org/content/22/21/2953.full.html#ref-list-1
	Article cited in: http://genesdev.cshlp.org/content/22/21/2953.full.html#related-urls
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here



The paper comments on the gene *perilipin (Plin)*, so we will now take a look at the binding sites surrounding that gene.


CLC Genomics Workbench 5.0.1 - Evaluation
File Edit Search View Toolbox Workspace Help
Show New Save Import NGS Import Export Graphics Print Undo Redo Cut Copy Paste Delete Workspace Plug-ins Search
Fit Width 100% Pan Selection ZoomIn ZoomOut
Navigation Area
CLC_Data
Example Data
Chip-Seq
ChipSeq2
NC_000073
chip-seq_ppar-subset
Reads_subset
chip-seq_ppar-subset FDR plot
chip-seq_ppar-subset FDR rate
chip-seq_ppar-subset ChIP-seq peaks
Recycle bin (...
<enter search term>
Toolbox
Alignments and Trees
General Sequence Analyses
Nucleotide Analyses
Protein Analyses
Sequencing Data Analyses
Primers and Probes
Cloning and Restriction Sites
RNA Structure
Expression Analysis
High-Throughput Sequencing
Genomics Gateway (beta)
BLAST
Database Search
Processes
Toolbox
Idle...
chip-seq_ppar...
chip-seq_ppar...
3,612,700
3,612,800
3,612,900
fragments concatenated
Consensus
Coverage
47
0
Read Mapping Settings
Read layout
Compactness Packed
Gather sequences at top
Show sequence ends
Show mismatches
Packed read height: Medium
Find Conflict
Low coverage threshold 8
Find Low Coverage
Sequence layout
No spacing
Numbers on sequences
Relative to 1
chip-seq_ppar...
Rows: 12 ChIP-seq peaks
Filter:
Region Length FDR (%) # Reads # Forward r... # Reverse re... Normalized ... Wilcoxon fil... 5' gene 3' gene Overlapping...
+(2096291..2... 56 3.03E-38 163 81 82 6.13E-3 5.20E-12 Plin (8281) Pex11a Pex11a, "ChI
+(2558891..2... 65 1.69E-29 131 66 65 7.63E-3 1.43E-7 Sema4b Cib1 (24042) Sema4b, "Ch.
+(3467757..3... 47 6.30E-9 51 24 27 0.06 1.00E-5 Alpk3 (6306) Slc28a1 (2833) "ChIP peak"
+(3612773..3... 74 6.30E-9 50 15 35 0.40 3.01E-5 EG636544 (2... Rps17 (85724) Pde8a, "ChIP
+(3928055..3... 91 6.34E-9 51 25 26 0.02 5.30E-7 Whdc1 Homer2 (374... Whdc1, "ChI
+(7429052..7... 58 1.09E-27 131 71 60 0.08 2.10E-6 Taldo1 LOC100043... Taldo1, "ChI.
+(7467645..7... 52 1.34E-15 72 44 28 0.22 7.29E-7 Cend1 (3451) Slc25a22 Slc25a22, "C.
+(7487001..7... 68 1.42E-25 118 48 70 0.19 3.46E-7 Rplp2 (642) Pnpla2 (2911) "ChIP peak"
+(7491669..7... 58 7.42E-23 113 58 55 0.03 2.92E-6 Pnpla2 Efcab4a (4189) Pnpla2, "ChI.
+(8029052..8... 47 2.18E-14 74 37 37 0.00 3.89E-5 Brsk2 2700078K21... Brsk2, "ChIP
+(9245732..9... 71 2.84E-5 36 20 16 0.11 2.65E-6 Kcnq1 Cdkn1c (247... Kcnq1, "ChI.
+(9346427..9... 72 5.09E-18 92 46 46 0.00 4.17E-7 Kcnq1 Cdkn1c (146... Kcnq1, "ChI.
Table Settings
Column width
Manual
Show column
Region
Length
FDR (%)
Reads
Forward reads
Reverse reads
Normalized difference
Wilcoxon filter p-value
5' gene