Chip-seq analysis

From alignments to motifs

D. Puthier, C. Rioualen, J. van Helden

Galaxy Workshop — Cuernavaca, 2017

Slides prepared in collaboration with:
Morgane Thomas-Chollier, Carl Herrmann,
Matthieu Defrance, Stéphanie Le Gras

Protocol

1. Create a new history named **Peak_calling**
2. Go to **Saved histories**; switch to the history **ChIP-Seq_mapping**
   a. Go to **Copy datasets**
   b. In the **Source history**, select **MarkDuplicates_on_filtered_ESR1_chr1.bam (bam)**
   c. In the Destination history, select the newly created history **Peak_calling**
   d. Click on **Copy History Items**
3. Go to the history **INPUT** and transfer M**arkDuplicates_on_filtered_input_chr1. bam (bam)** to the **Peak_calling** history in the same manner

Protocol

The first question is **whether the numbers of aligned reads are roughly the same in the input and chip bam file**. Let's compute **alignments statistics** on the data we have just imported.

1. Use the tool **Flagstat** (search "flagstats" in the search field of the toolbox).
2. Select the **input bam file** in the dropdown list menu. **Execute**.
3. Select the newly created dataset from history. Click on **Run this job again** (↺)
4. Run the tool on **ESR1 bam file** dataset.

Rename the datasets

1. Click on **Edit Attributes** in a selected dataset (✎).
1. In the **Attribute tab**, enter a new name in the **Name** text field.
2. Click on **Save**

Rename the files as indicated:

- Flagstat on data ... to > **Flagstat on ESR1_chr1.bam**
- Flagstat on data ... to > **Flagstat on input_chr1.bam**

Protocol

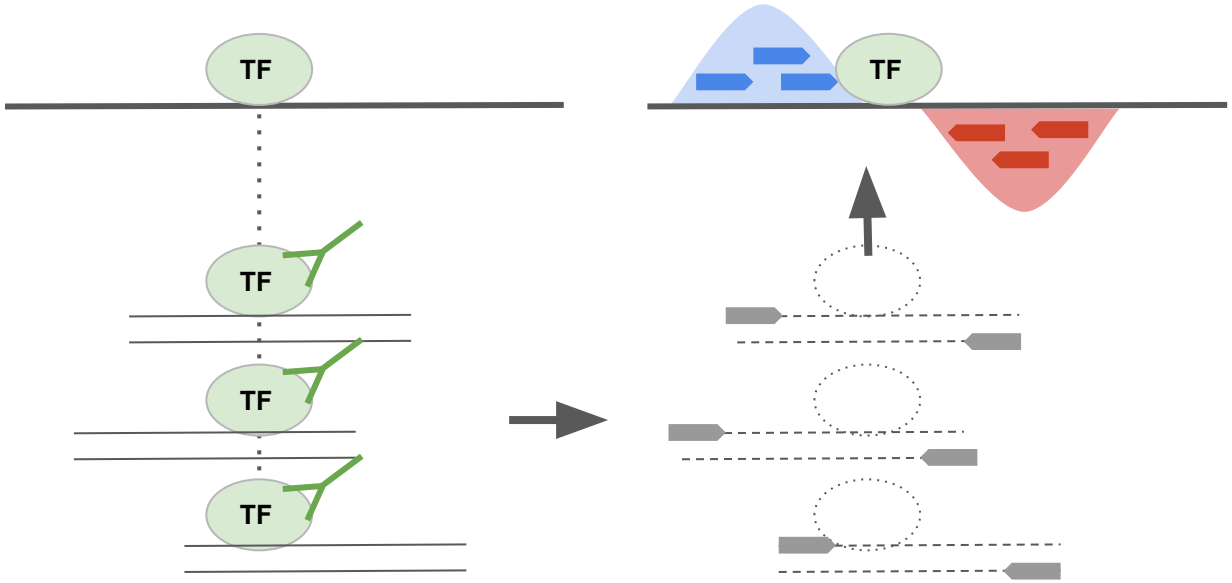
Hint: Rename the datasets all along the training session.

Q1: Is the number of **aligned reads** balanced ?

Hint: when running twice the same tool with the same parameters on several datasets, you can use  . Then, select all the datasets you want to analyse (select multiple datasets by pressing the **Ctrl key** when clicking on a dataset name in the drop down list.

Expected signal

- For a transcription factor, the signal is expected to be **sharp**.

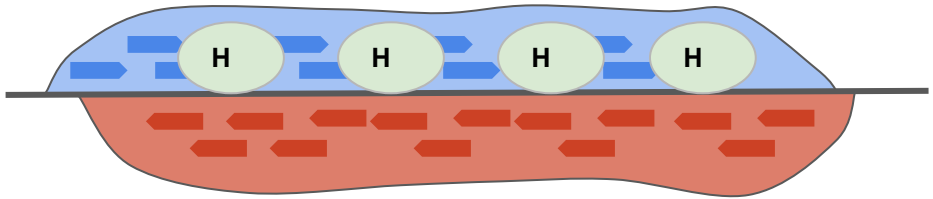


The binding site itself is generally not sequenced !

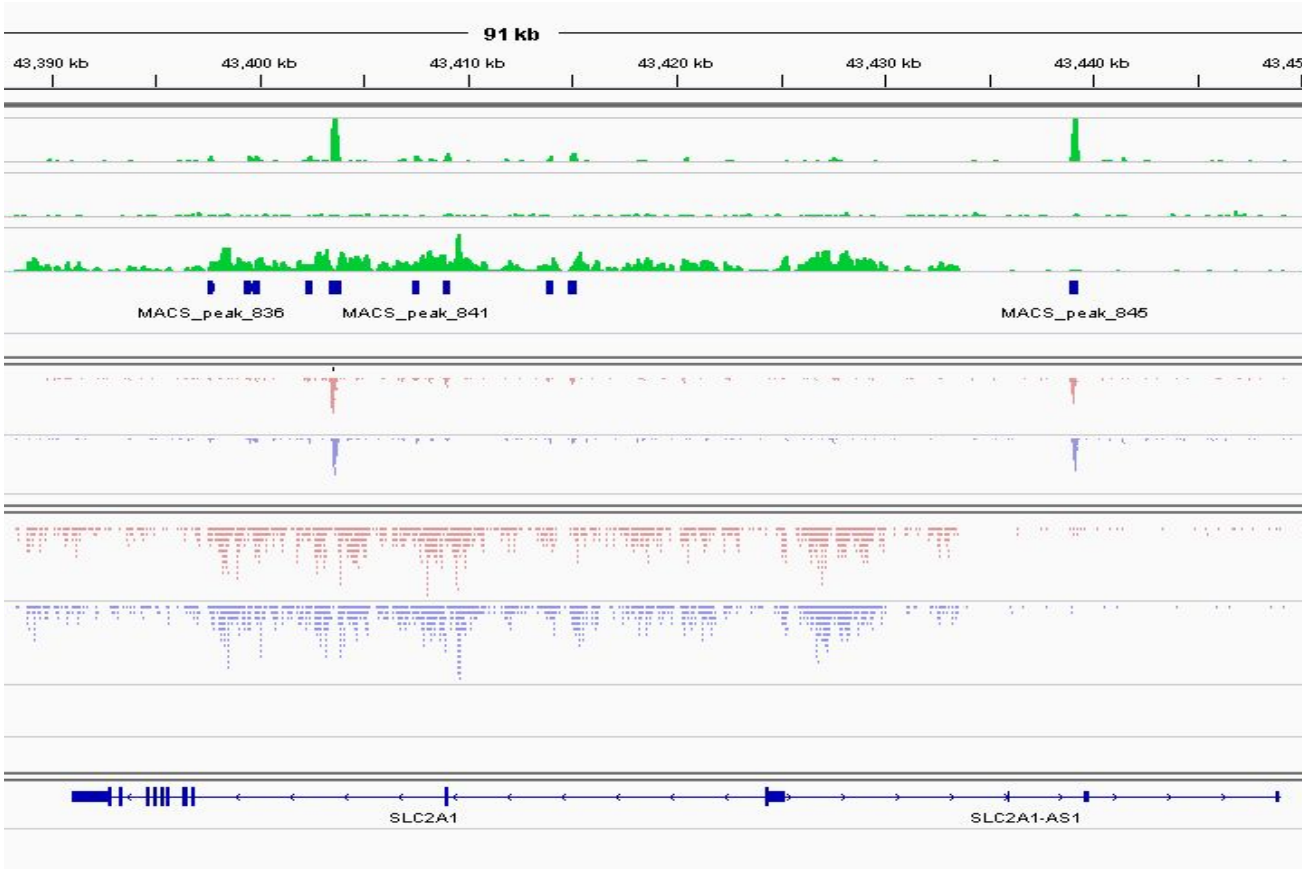
Forward alignments
Rev/comp alignments

Expected signal

- For most **histone marks** the signal is expected to be **broad**.
- Asymmetry is less/not pronounced.
- Peak calling algorithms need to adapt to these various signals.



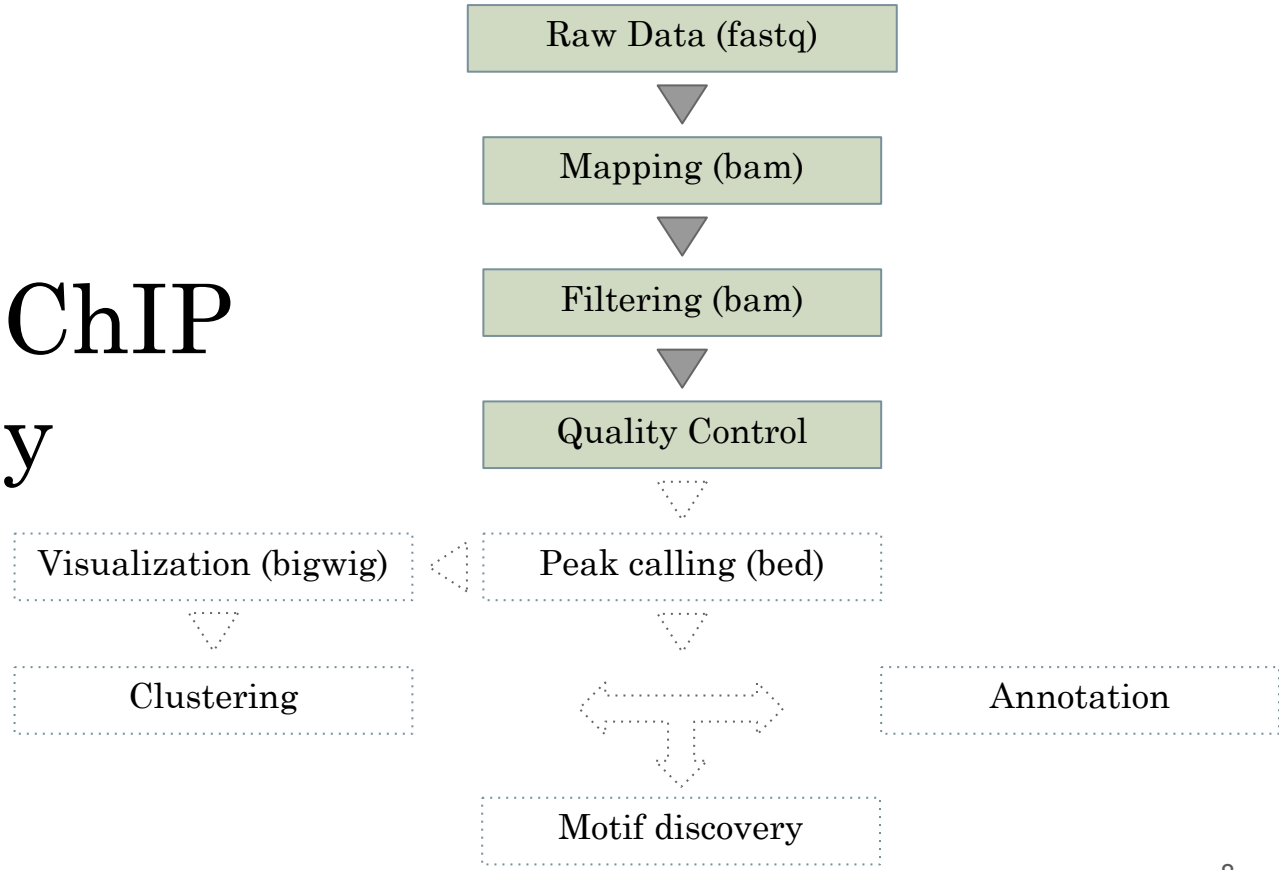
Observed signal



Trans. Factor (ESR1)

Histone mark (H3K4me1)

Checking ChIP quality

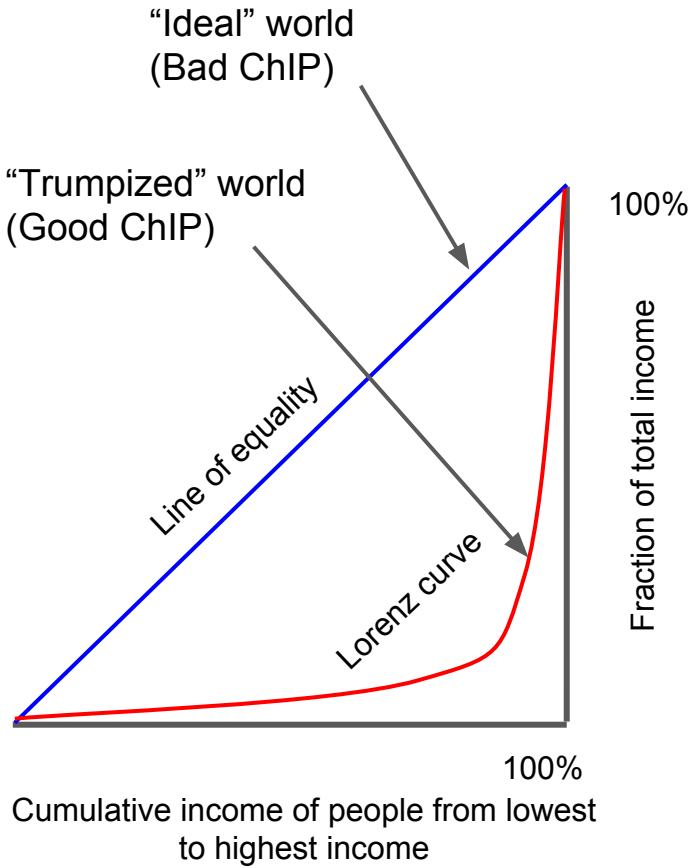


Assessing ChIP quality

- Various metrics
  - Check **duplicate** rate (see Filtering section)
  - Use a **Lorenz Curve** (implemented in **Deeptools fingerprint**)
  - Look at **strand cross-correlation** (implemented in **SPP** BioC package)
  - Fraction of reads in peaks (**FRiP**, as proposed by ENCODE)
    - Requires to find peaks.

Lorenz curve

- Analyze distribution of income among workers by computing cumulative sum.
  - If uniform income distribution (:
    - Straight line
  - If they were Trumpized
    - Lorenz curve
- Here the workers are the genome windows and incomes are reads

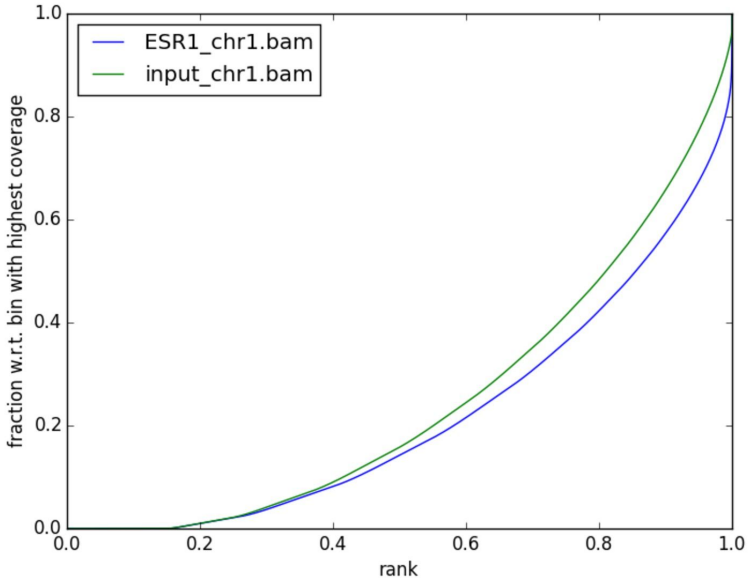


Protocol

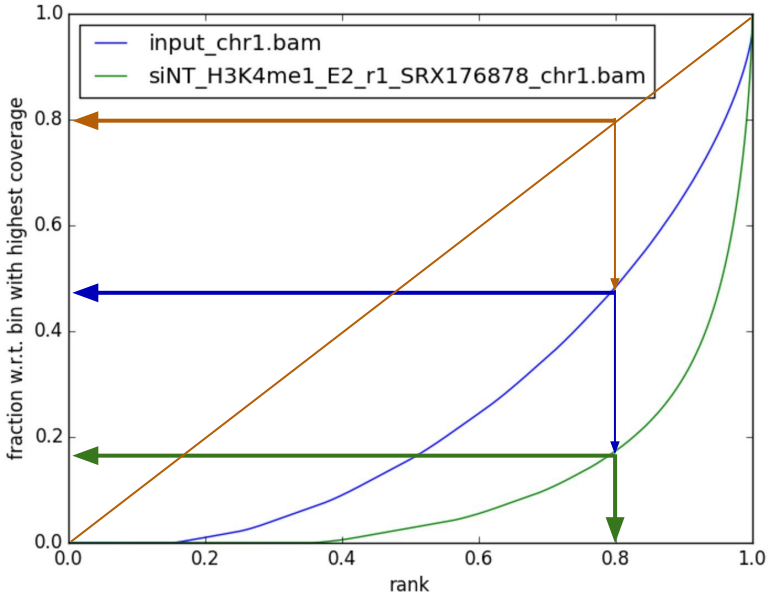
1. Select **plotFingerprint** from the toolbox.
   a. **Bam file** :
      i. **MarkDuplicates on filtered input_chr1.bam (bam)**
      ii. **MarkDuplicates on filtered ESR1_chr1.bam (bam)**
   b. To improve speed, set **Region of the genome to limit the operation** to **chr1**.
   c. Set **Show advanced options** to **yes**.
   d. Set **Bin size in bases** to 250.
   e. Set **Extend reads to the given average fragment size** to **A custom length**.
   f. Set the custom length to **150** (we will see later if these value is a good approximation).
   g. Set **Ignore duplicates** to **yes**.
   h. **Execute**.

Results of Lorenz curve analysis

Transcription factor
(we should expect narrow peaks)



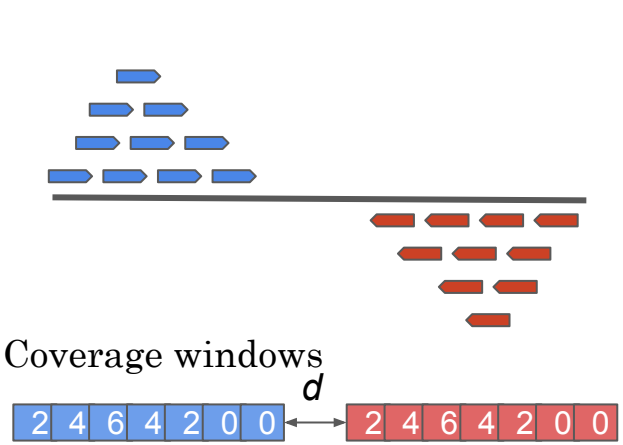
Methylation mark
(we should expect broad peaks)



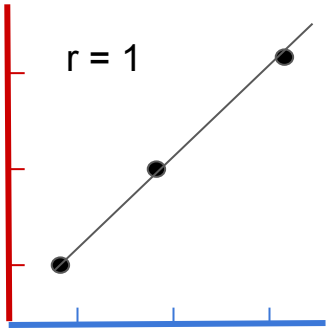
- **siNT (chipped-:** 80% of the genomic regions are covered by ~18% of he reads, whereas the 20% richest regions capture 92% of the reads.
- **Genomic input**: 80% genomic regions covered by 45% reads
- **Random expectation**: 80% genomic regions would be covered by 80% of the reads/

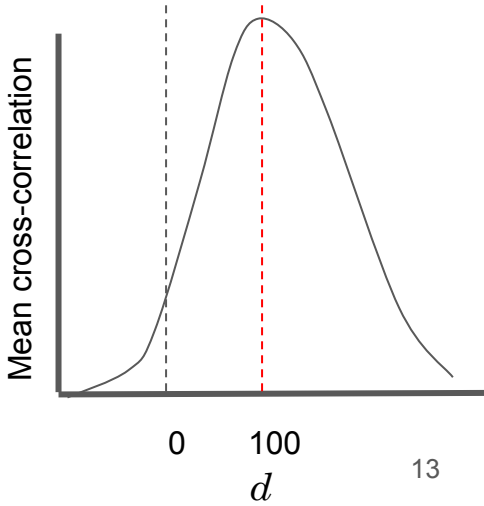
SPP (strand cross-correlation)

- Compute strand cross-correlation for each window w across the genome.
- Use various distances (***d***) and compute the mean cross-correlation observed.



Strand cross-correlation for each window and various d values

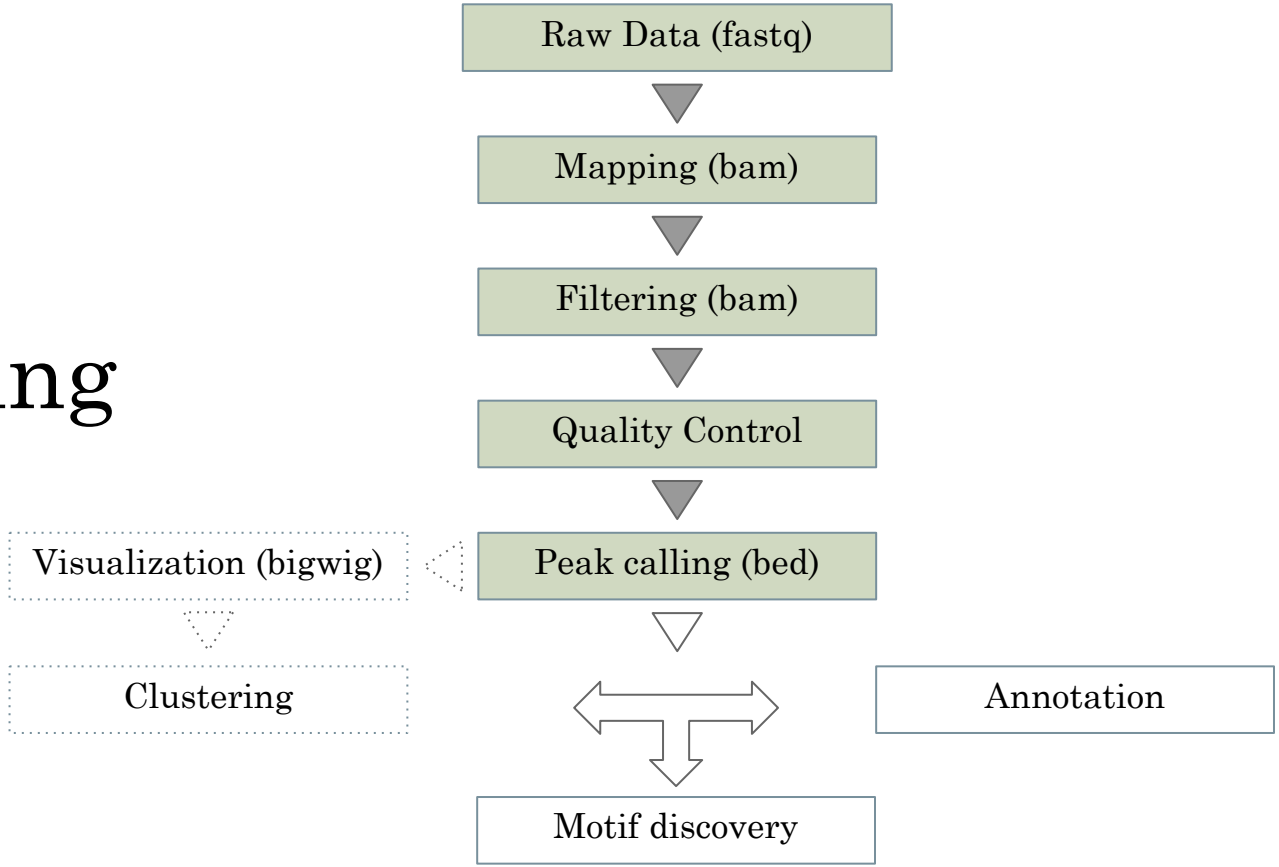




Protocol

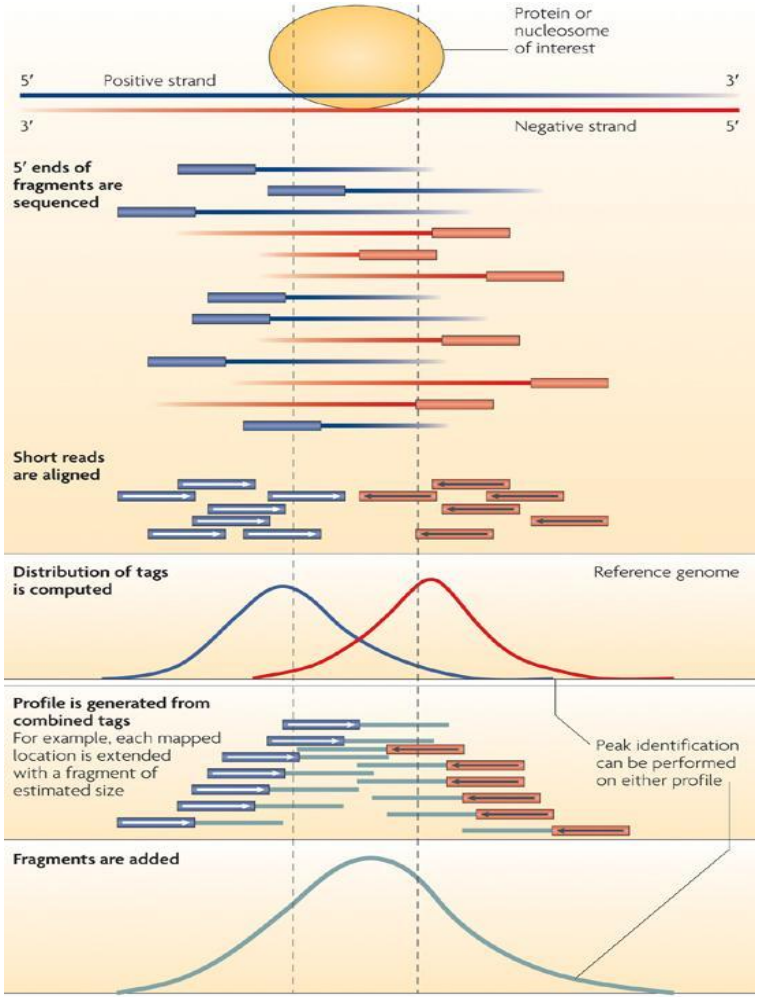
1. Select **SPP cross-correlation analysis package** tool. /!\ not working
2. Select **chip-quality-check** as Experiment name.
3. Select **Determine strand cross-correlation** as **Select action to be performed** (default).
4. **ChIP-Seq Tag File**: **MarkDuplicates on filtered ESR1_chr1.bam (bam)**
5. **ChIP-Seq Control File**: **MarkDuplicates on filtered input_chr1.bam (bam)**
6. Click execute.
7. What does the resulting plot indicate ?

Peak calling



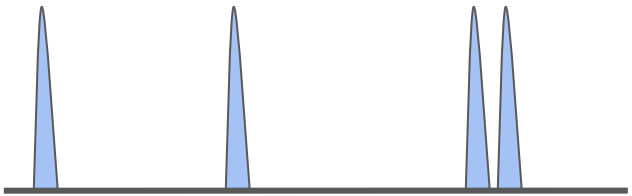
From reads to peaks

- Get the signal at the right position
  - Read shift
  - Extension
- Estimate the fragment size
- Do paired-end

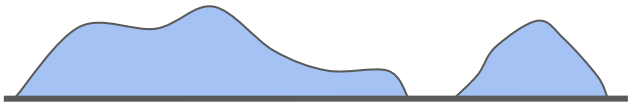


Peak callers

- The **peak caller** should be **chosen** based on
  - Experimental design
    - SE or PE (E.g MACS1.4 vs MACS2)
  - Expected signal
    - Sharp peaks (e.g. Transcription Factors).
      - E.g. MACS
    - Broad peaks (e.g. epigenetic marks).
      - E.g MACS, SICER,...



Sharp peaks



Broad peaks

Peak callers

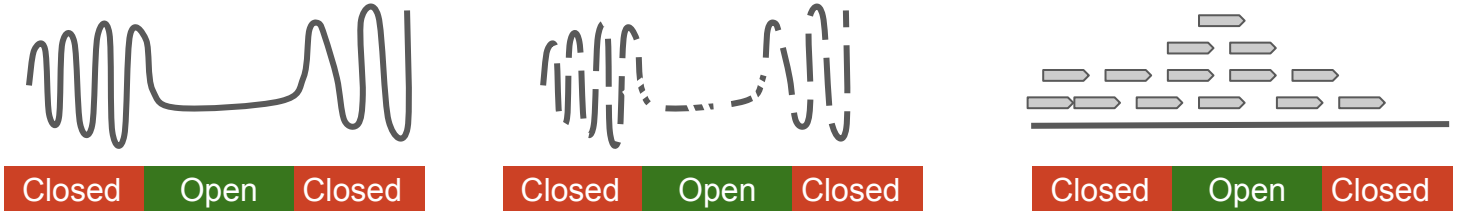
	Profile	Peak criteria[a]	Tag shift	Control data[b]	Rank by	FDR[c]	User input parameters[d]	Artifact filtering: strand-based/ duplicate[e]	Refs.
CisGenome v1.1	Strand-specific window scan	1: Number of reads in window 2: Number of ChIP reads minus control reads in window	Average for highest ranking peak pairs	Conditional binomial used to estimate FDR	Number of reads under peak	1: Negative binomial 2: conditional binomial	Target FDR, optional window width, window interval	Yes / Yes	10
ERANGE v3.1	Tag aggregation	1: Height cutoff High quality peak estimate, per-region estimate, or input	High quality peak estimate, per-region estimate, or input	Used to calculate fold enrichment and optionally *P* values	*P* value	1: None 2: $\frac{\#\ control}{\#\ ChIP}$	Optional peak height, ratio to background	Yes / No	4,18
FindPeaks v3.1.9.2	Aggregation of overlapped tags	Height threshold	Input or estimated	NA	Number of reads under peak	1: Monte Carlo simulation 2: NA	Minimum peak height, subpeak valley depth	Yes / Yes	19
F-Seq v1.82	Kernel density estimation (KDE)	*s* s.d. above KDE for 1: random background, 2: control	Input or estimated	KDE for local background	Peak height	1: None 2: None	Threshold s.d. value, KDE bandwidth	No / No	14
GLITR	Aggregation of overlapped tags	Classification by height and relative enrichment	User input tag extension	Multiply sampled to estimate background class values	Peak height and fold enrichment	2: $\frac{\#\ control}{\#\ ChIP}$	Target FDR, number nearest neighbors for clustering	No / No	17
MACS v1.3.5	Tags shifted then window scan	Local region Poisson *P* value	Estimate from high quality peak pairs	Used for Poisson fit when available	*P* value	1: None 2: $\frac{\#\ control}{\#\ ChIP}$	*P*-value threshold, tag length, mfold for shift estimate	No / Yes	13
PeakSeq	Extended tag aggregation	Local region binomial *P* value	Input tag extension length	Used for significance of sample enrichment with binomial distribution	*q* value	1: Poisson background assumption 2: From binomial for sample plus control	Target FDR	No / No	5
QuEST v2.3	Kernel density estimation	2: Height threshold, background ratio	Mode of local shifts that maximize strand cross-correlation	KDE for enrichment and empirical FDR estimation	*q* value	1: NA 2: $\frac{\#\ control}{\#\ ChIP}$ as a function of profile threshold	KDE bandwidth, peak height, subpeak valley depth, ratio to background	Yes / Yes	9
SICER v1.02	Window scan with gaps allowed	*P* value from random background model, enrichment relative to control	Input	Linearly rescaled for candidate peak rejection and *P* values	*q* value	1: None 2: From Poisson *P* values	Window length, gap size, FDR (with control) or *E*-value (no control)	No / Yes	15
SiSSRs v1.4	Window scan	$N_+ - N_-$ sign change, $N_+ + N_-$ threshold in	Average nearest paired tag distance	Used to compute fold-enrichment distribution	*P* value	1: Poisson 2: control	1: FDR 1,2: $N_+ + N_-$ threshold	Yes / Yes	11

- In 2009 there were already a dozen of peak callers.
- New tools and versions emerged since then.
- Which peak-caller should be used?
- How to tune its parameters?

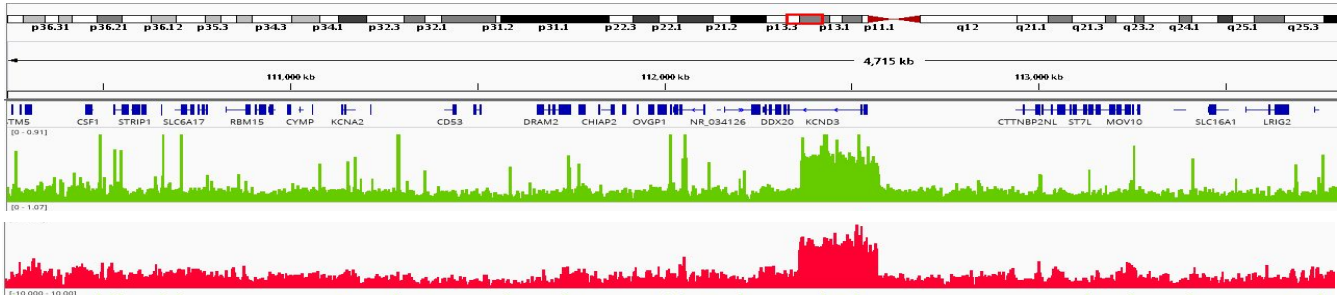
Pepke,S., Wold,B. and Mortazavi,A. (2009) Computation for ChIP-seq and RNA-seq studies. *Nature Methods*, **6**, S22–32.

Why is input mandatory ?

- The input is used to model **local noise level.**
  - **Accessible regions are expected to produce more reads.**



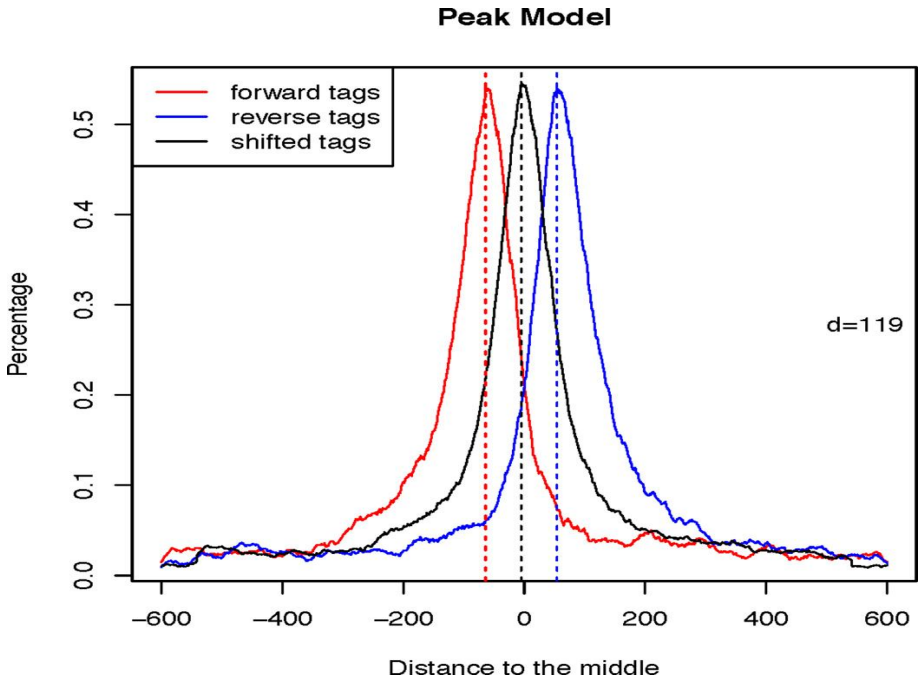
  - **Amplified regions (CNV) are expected to produce more reads.**



MACS [Zhang et al, 2008]

1. Modeling the shift size of ChIP-Seq tags

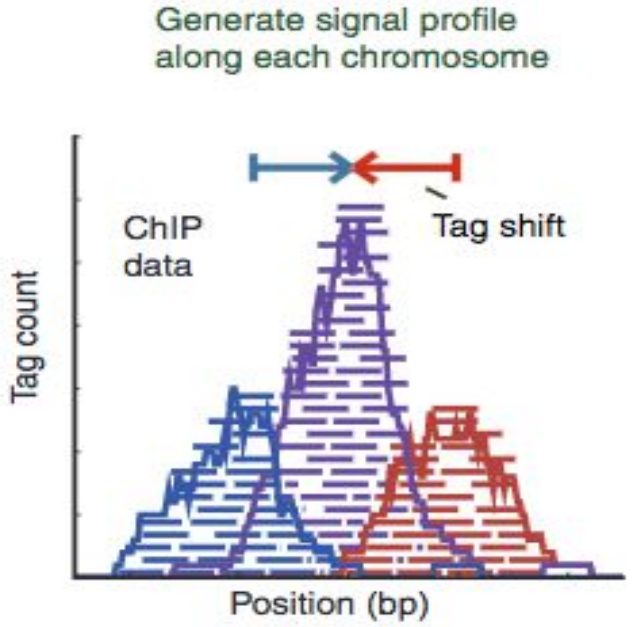
- **slides *2*bandwidth* windows** across the genome to find regions with tags **more than *mfold* enriched** relative to a random tag genome distribution
- randomly **samples 1,000** of these highly enriched regions
- **separates their Watson and Crick tags**, and **aligns them** by the midpoint between their Watson and Crick tag centers
- **define *d* as the distance** in bp between the summit of the two distribution



MACS [Zhang et al, 2008]

2. Peak detection

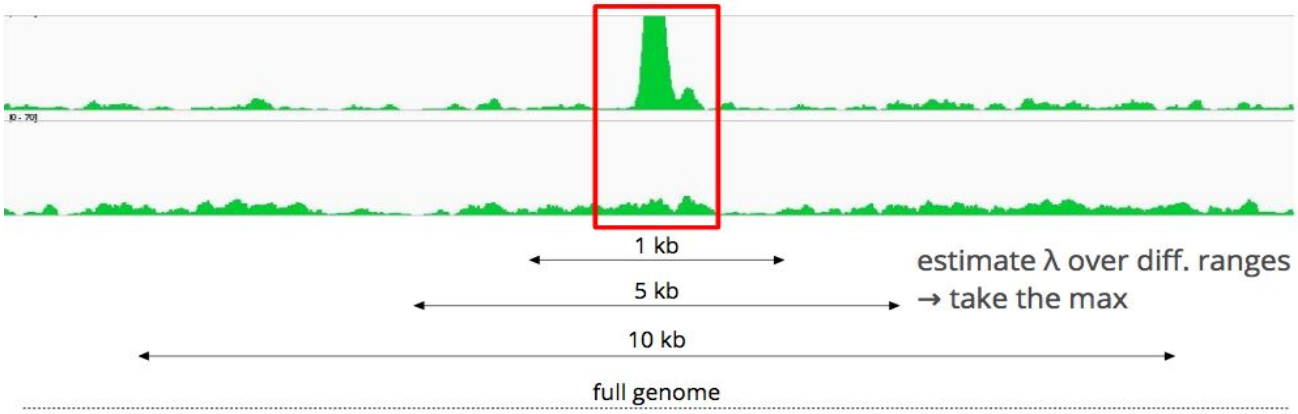
- Scales the total Input read count to be the same as the total ChIP read count
- Duplicate read removal
- Tags are shifted by *d*/2

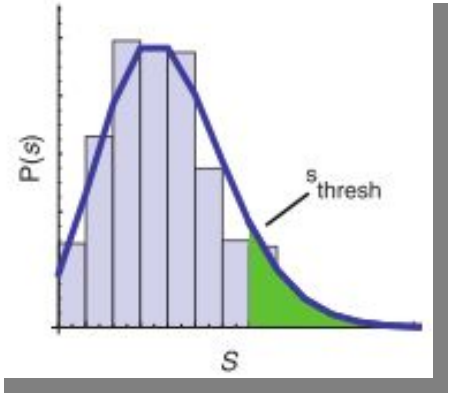


Pepke et al, 2009

MACS [Zhang et al, 2008]

- Slides 2d windows across the genome to find candidate peaks with a significant tag enrichment (Poisson distribution *p*-value based on λ_{BG}, default 10^{-5})
- Estimate parameter λ_{local} of Poisson distribution



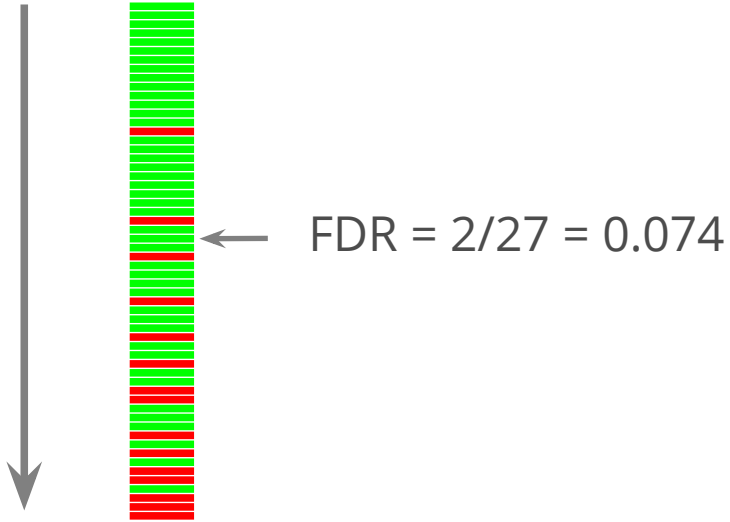


MACS [Zhang et al, 2008]

3. **Multiple testing correction with the False Discovery Rate (FDR)**

- Swap treatment and input and call negative peaks
- Take all the peaks (neg + pos) and sort them by increasing p-values (decreasing significance)
- Whilst going down in the list, you progressively increase
  - The number of true positives (green)
  - The number of false negatives (red)
- At each level, estimate the FDR

$$FDR(p) = \frac{\#\ \textbf{Negative}\ \text{peaks with p-value} < p}{\#\ \textbf{Negative}\ \text{peaks with p-value} < p \ + \ \#\ \textbf{Positive}\ \text{peaks with p-value} < p}$$



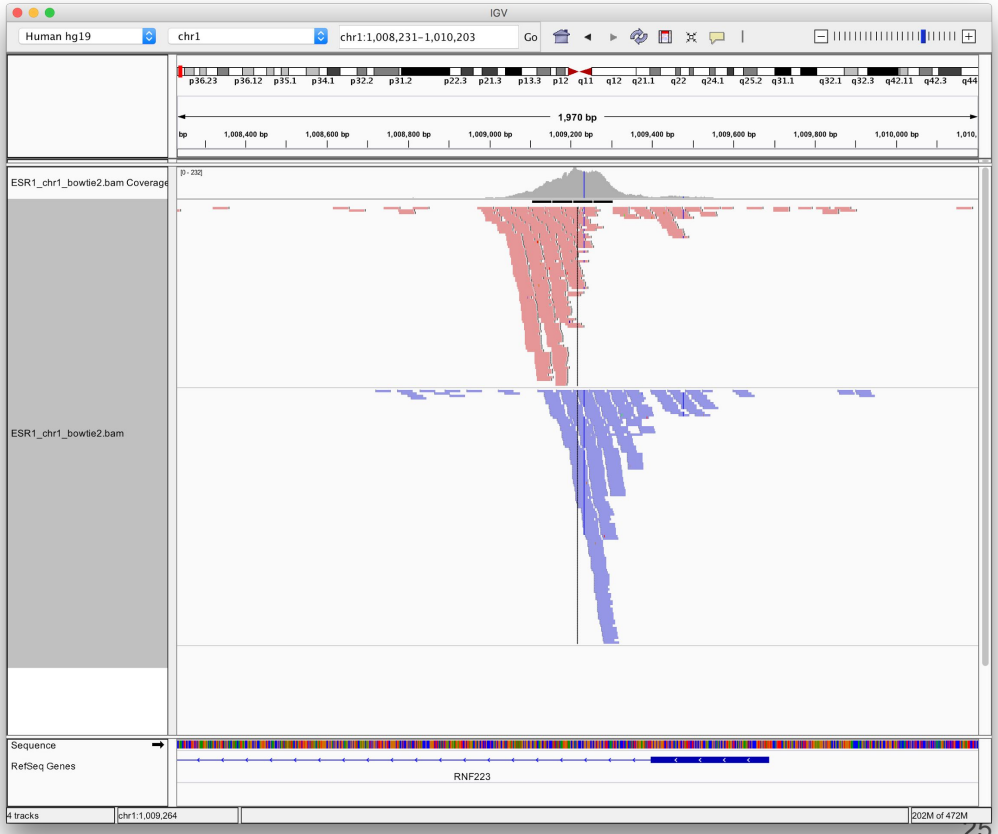
Protocol

Peak calling with MACS 1.4

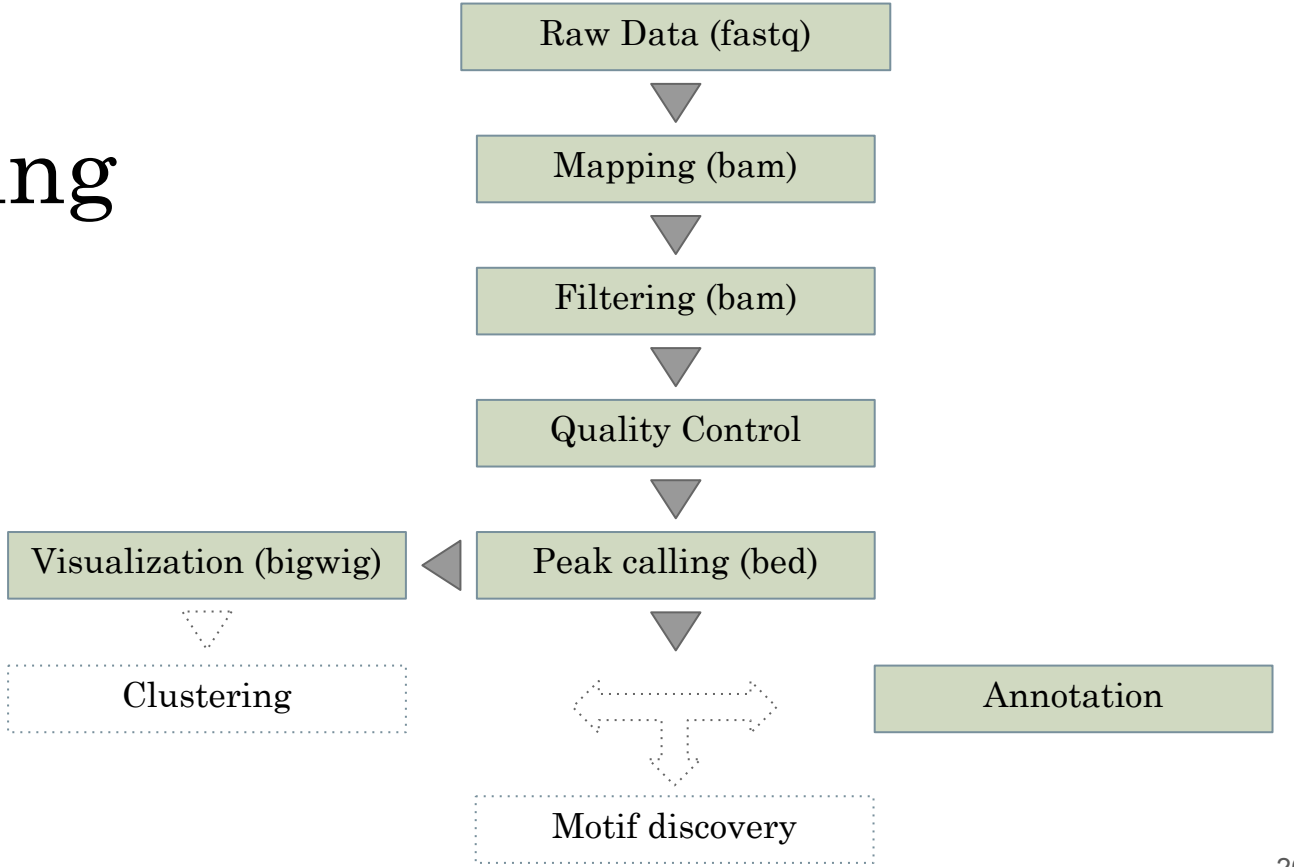
1. Select **MACS14** from the toolbox.
2. Set **Experiment name** to **ESR1_vs_input_macs14.**
3. MACS can handle single or paired-end data. Set **Paired end sequencing** to **Single End**.
4. Set **ChIP-seq tag file**: **MarkDuplicates on filtered ESR1_chr1.bam (bam).**
5. Set **ChIP-seq control file**: **MarkDuplicates on filtered input_chr1.bam (bam).**
6. **Effective Genome size:** 199400000 (80% of chr1 because we restricted the read mapping to this region).
7. **Tag size** : these are Illumina datasets of read size 36.
8. All other options should be set to default.
9. **Execute**.

A nice peak

- An example of well-shaped peak.
- The peak falls in the first intron of the gene RNF223.
- Note the characteristic shift between reads mapped on the forward and reverse strands.



Annotating peaks



Import coverage profiles (bigWig) from the shared library

Protocol

Perform the following steps to Import the BAM files that will be subsequently analyzed

1. Go to **Shared data > Data libraries > Theodorou > BigWig files**. Select the following **bigWig files** (full dataset):
   a. SRR540190_GSM986061_siNT_ER_E2_r2.UNIQALIGN.sorted_vs_Input_SES_log2.bw
   b. SRR540212_GSM986083_siNT_H3K4me1_Veh_r1.UNIQALIGN.sorted_vs_Input_SES_log2.bw
2. Click on **to History**. Enter the name **ESR1_peak_Annotation** to create a **new history**. Click on **Import to History.**
3. Rename the files as follows:
   a. SRR540190_GSM986061_siNT_ER_E2_r2.UNIQALIGN.sorted_vs_Input_SES_log2.bw **to ESR1_vs_input.bw**
   b. SRR540212_GSM986083_siNT_H3K4me1_Veh_r1.UNIQALIGN.sorted_vs_Input_SES_log2.bw **to H3K4me1.bw**

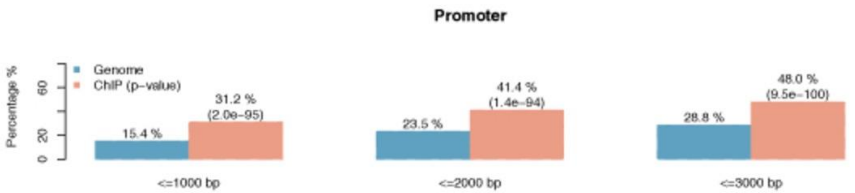
Import the peaks (bed format) from the shared library

Protocol

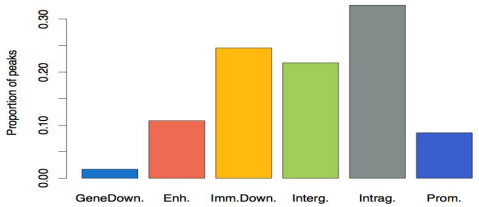
1. Go to **Shared data > data libraries > Theodorou > BED.** Select the following files and import them to ESR1_peak_Annotation history:
   a. macs.SRR540190_GSM986061_siNT_ER_E2_r2.UNIQALIGN.sorted_peaks.bed
   b. macs.SRR540212_GSM986083_siNT_H3K4me1_Veh_r1.UNIQALIGN.sorted_peaks.bed
2. Click on **to History**. Select **ESR1_peak_Annotation** as an history. Click on **Import**
3. Rename the files:
   a. **ESR1_peaks.bed**
   b. **H3K4me1_peaks.bed**

Are peaks biased toward any chromosome or genomic features?

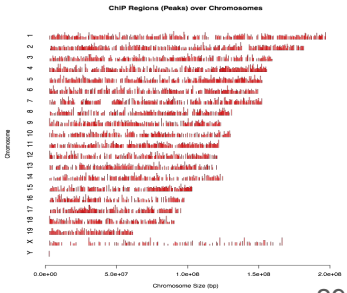
- What is the chromosomal distribution of the peak?
- What are the genomic regions associated with peaks?
  - Exonic, intronic, intergenic, promoter, bidirectional promoter, known epigenetic marks, any user defined regions (...)?
- What are the genomic regions enriched in peaks?
- Do some peaks overlap with other ones?



CEAS

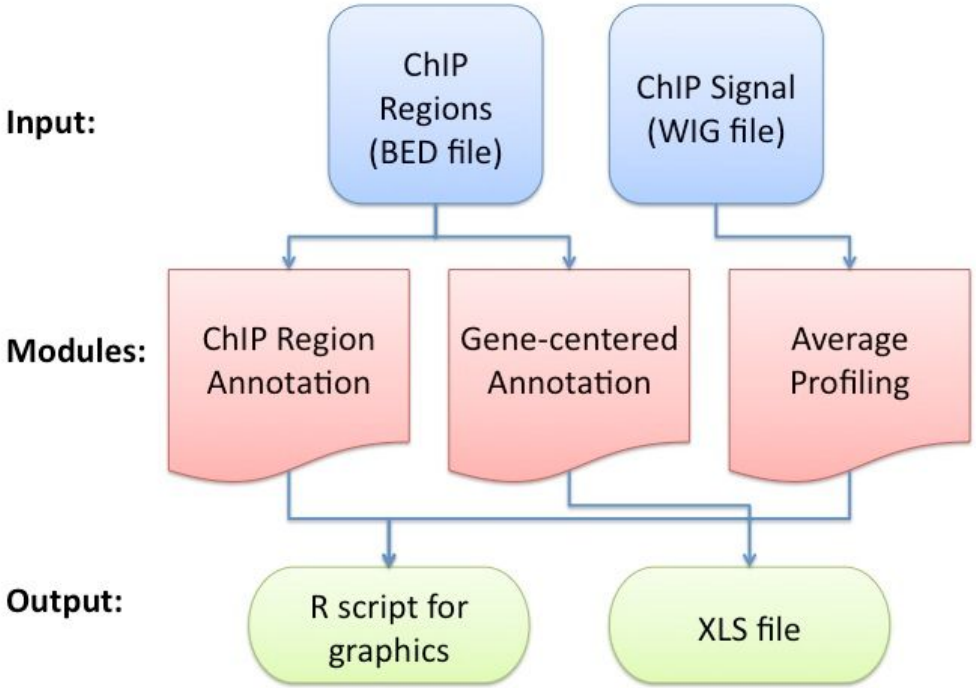


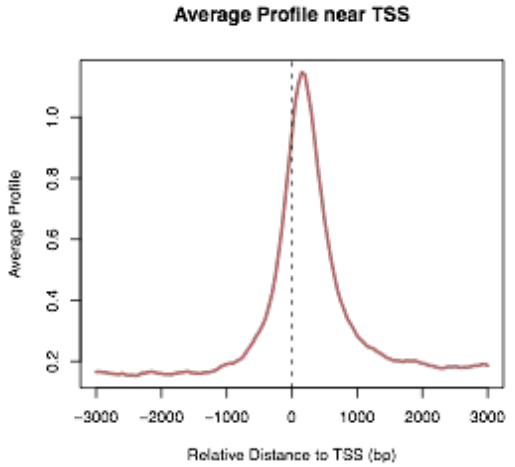
Galaxy::AnnotatePeaks



CEAS

CEAS (Cis-regulatory Element Annotation System)





http://liulab.dfci.harvard.edu/CEAS/

Using CEAS (Cis-regulatory Element Annotation System)

Protocol

CEAS can be used to annotate your peak and perform a first general analysis looking for chromosomes and genomic features enriched in peak of interest. One very interesting feature is that any kind of genomic feature can be provided for analysis. Here we will also try to test whether ESR1 peaks are biased toward regions marked for H3K4me1.

1. Search for **CEAS in the galaxy toolbox**
2. Set **Bed file of ChIP regions** to **ESR1_peaks.bed**
3. Set **Bed file of extra region of interest** to **H3K4me1_peaks.bed**
4. Set **Gene annotation table to hg19**
5. Click on **Execute**

NB: CEAS may also accept a wig file (e.g ESR1 signal) to draw coverage profile around features.

Interpretation of CEAS results

Exercise

Open the "graph results from CEAS" file:

Q1: Are there genomic location biases at the chromosome level?

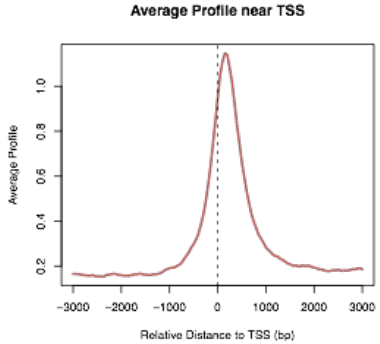
Q2: Are there genomic location biases at the feature level? Promoter, exons, introns (...)?

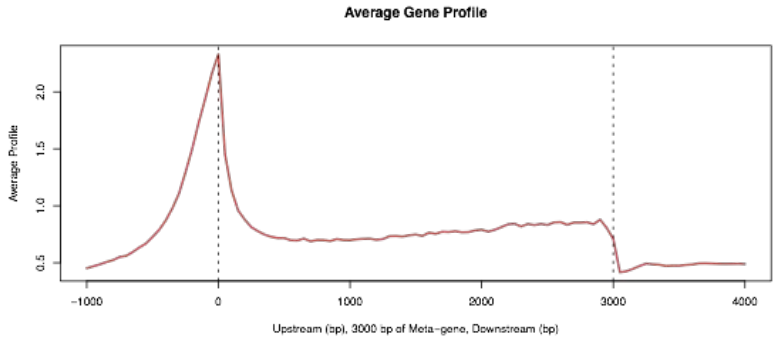
Open "Gene-centered annotation from CEAS" file.

Q3: What does this file contain?

How do the signal distribute relative to known features?

- How do the peaks distribute relative to
  - Transcriptional start site (TSS)
  - Transcriptional termination site (TTS)
  - Gene body, exon, intron,...





Computing H3K4me1 coverage profile around the TSS (step 1)

Protocol

1/ Getting transcript coordinates

1. Select **UCSC Main table browser** from the toolbox
2. Select **Clade**: Mammal ; **Genome**: Human ; **Assembly**: "Feb. 2009 (hg19, GRCh37)" ; **Group**: Genes and Gene Prediction tracks ; **Track**: UCSC gene ; **Table**: knownGene ; **Region**: genome ; **Output format:** bed ; **Send output to**: galaxy
3. Set **position** to chr1
4. Click on **Get the output**.
5. This opens a new page with additional options. Leave these options unchanged, and click **Send query to Galaxy**. The UCSC table generated by your query is automatically transferred to the current history of your
6. Rename the filtered bed file to **hg19_chr1.bed.**

Computing H3K4me1 coverage profile around the TSS (step 2)

Protocol

2/ Computing a coverage matrix

1. Select **ComputeMatrix** from the toolbox.
2. Set **Select Regions** to **hg19_chr1.bed.**
3. Set **Score file** to **H3K4me1.bw**
4. Set **ComputeMatrix has two main output options** to **reference-point**.
5. Set **Reference point for plotting** to **TSS**.
6. Set **upstream and downstream distance** to **5000**.
7. Set **Show advanced options** to **Yes**
8. **Set length of bins to 50.**
9. **Set Convert missing values to zero** to **YES**.
10. Execute.
11. Rename the result **H3K4me1_around_TSS_Matrix**.

Compute H3K4me1 coverage profile around the TSS

Protocol

3/ Computing profile

1. Select **plotProfile** from the toolbox.
2. For the Matrix File option, select the result of the previous step (**H3K4me1_around_TSS_Matrix**)
3. Set **The input matrix was computed in scale-regions mode** to No.
4. Set Press **Execute**.

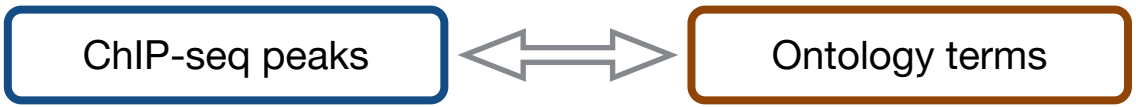
4/ Computing heatmap

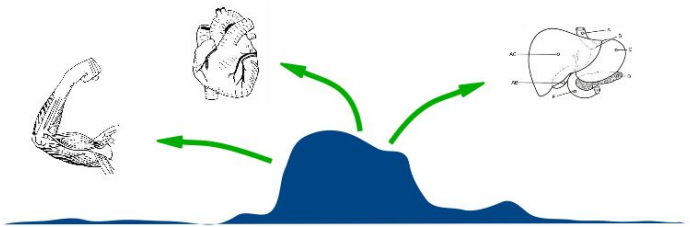
1. Select **plotHeatmap** from the toolbox.
2. For the Matrix File option, select the result of the previous step (**H3K4me1_around_TSS_Matrix**)
3. Press **Execute**.

Q1: What would you say from the average coverage profile around the TSS. What is the depression observed at the TSS position?

What is hidden functional meaning?

- What are the genes associated to the peaks?
  - Find the closest genes (see Annotate peaks).
- More globally
  - Are some functional categories overrepresented?



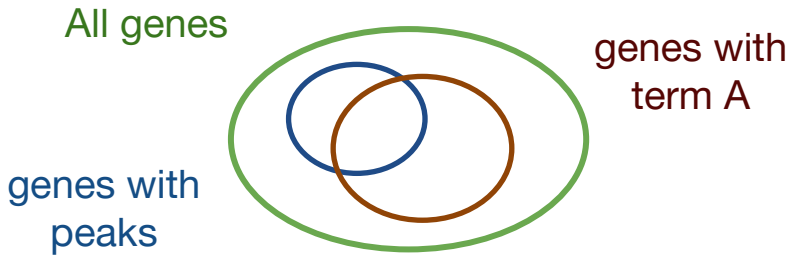


GO Molecular Function
GO Biological Process
Disease Ontology
Pathways
…

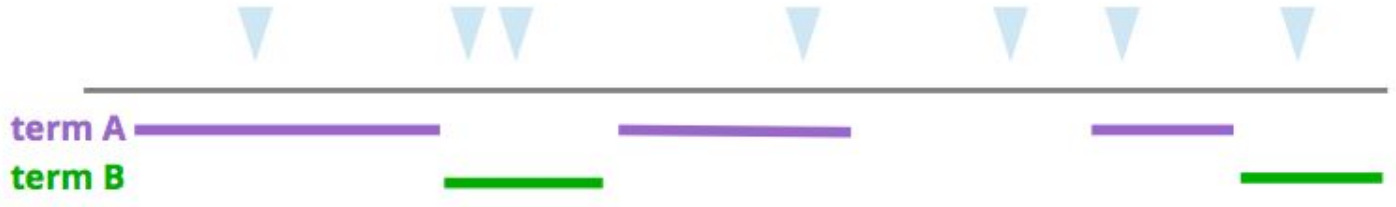
GREAT

- Performs functional enrichment test
  - Hypergeometric
    - Using closest genes
    - Tends to be biased
  - Binomial test
    - More robust

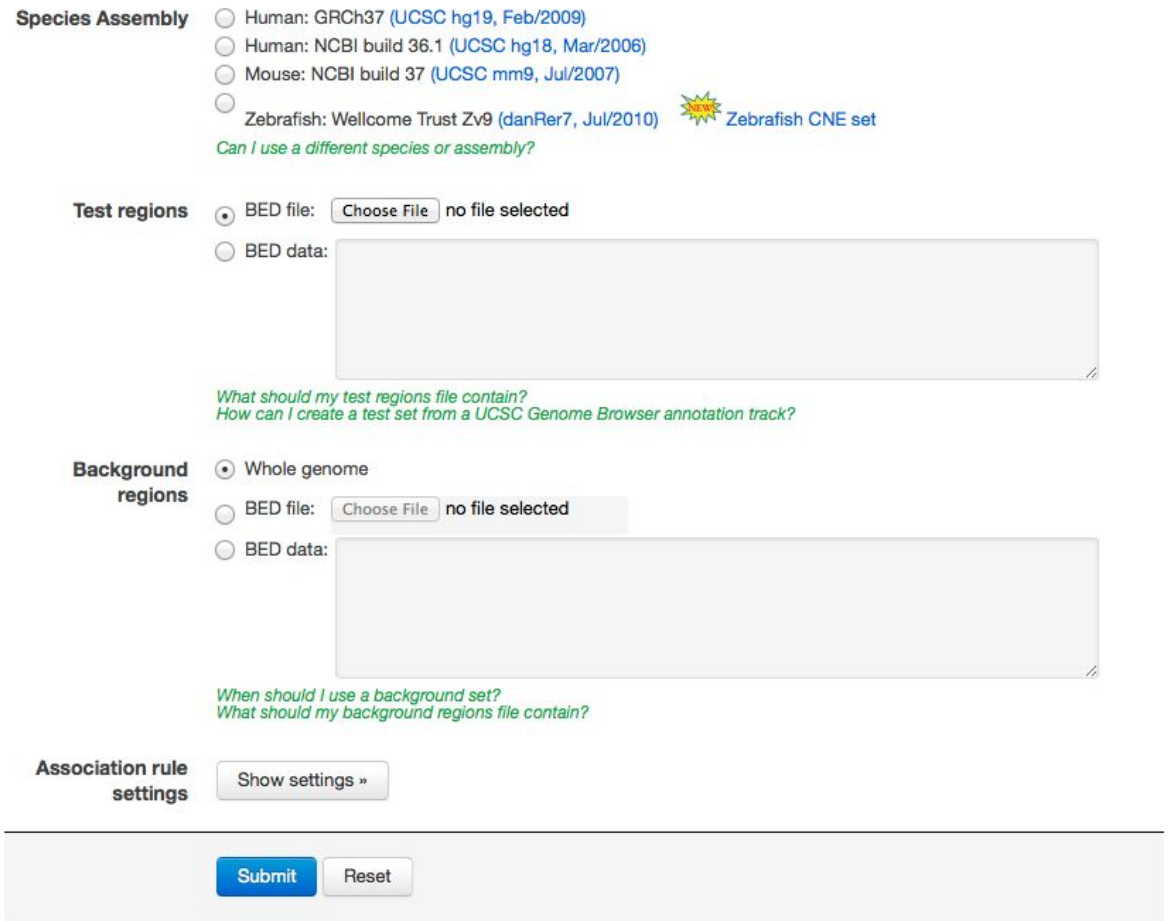
Hypergeometric test over gene list



Binomial test over regions



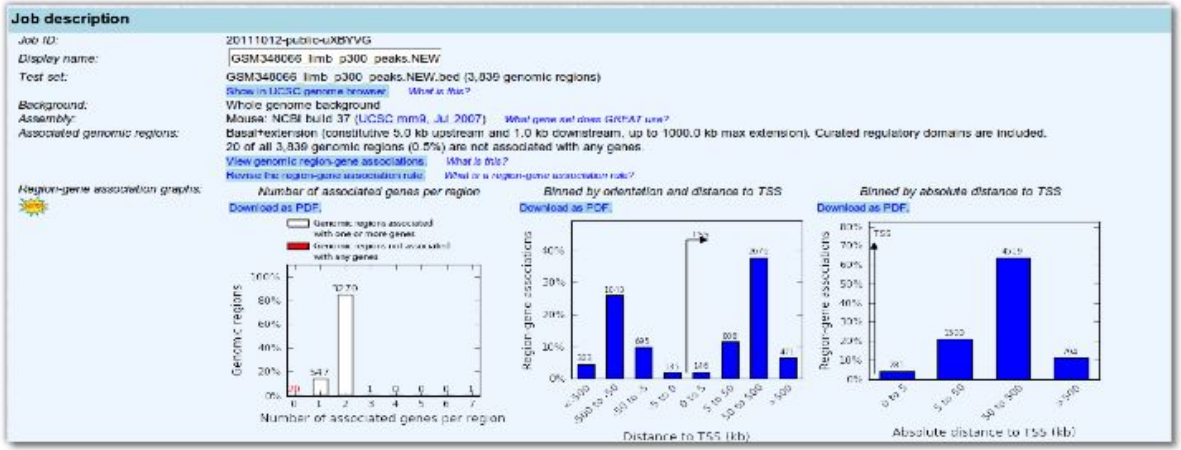
GREAT

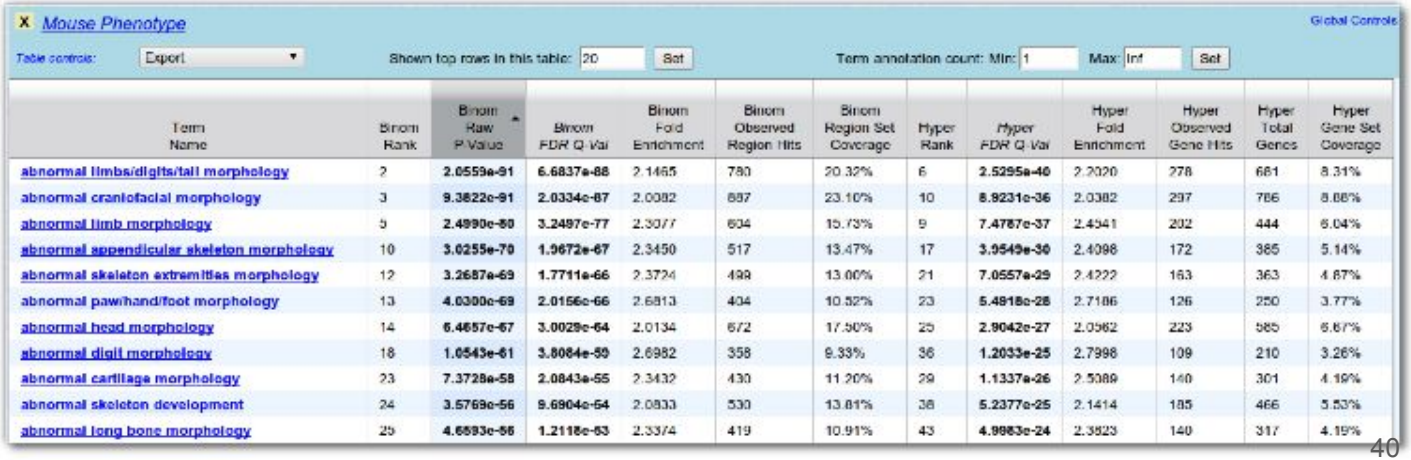


Note: Only human (hg19 and hg18), mouse (mm9) and zebrafish (danRer7) genomes are supported

GREAT

- Input
  - bed file with peaks
- Output
  - Enriched GO terms and functions





Term Name	Binom Rank	Binom Raw P Value	Binom FDR Q Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
abnormal limbs/digits/tail morphology	2	2.0559e-91	6.6837e-88	2.1465	780	20.32%	6	2.5295e-40	2.2020	278	681	8.31%
abnormal craniofacial morphology	3	9.3822e-91	2.0334e-87	2.0082	887	23.10%	10	8.9231e-36	2.0382	297	786	8.88%
abnormal limb morphology	5	2.4890e-80	3.2497e-77	2.3077	604	15.73%	9	7.4787e-37	2.4541	202	444	6.04%
abnormal appendicular skeleton morphology	10	3.0255e-70	1.9672e-67	2.3450	517	13.47%	17	3.9549e-30	2.4098	172	385	5.14%
abnormal skeleton extremities morphology	12	3.2687e-69	1.7711e-66	2.3724	499	13.00%	21	7.0557e-29	2.4222	163	363	4.87%
abnormal paw/hand/foot morphology	13	4.0300e-69	2.0156e-66	2.6813	404	10.52%	23	5.4818e-28	2.7186	126	250	3.77%
abnormal head morphology	14	6.4657e-67	3.0029e-64	2.0134	672	17.50%	25	2.9042e-27	2.0562	223	585	6.67%
abnormal digit morphology	18	1.0543e-61	3.8084e-59	2.6982	358	9.33%	36	1.2033e-25	2.7998	109	210	3.26%
abnormal cartilage morphology	23	7.3728e-58	2.0843e-55	2.3432	430	11.20%	29	1.1337e-26	2.5089	140	301	4.19%
abnormal skeleton development	24	3.5769e-56	9.6904e-54	2.0833	530	13.81%	38	5.2377e-25	2.1414	185	466	5.53%
abnormal long bone morphology	25	4.6593e-56	1.2118e-53	2.3374	419	10.91%	43	4.9963e-24	2.3823	140	317	4.19%

Selecting top peaks based on MACS2 score

TO BE WRITTEN

Protocole

TO BE WRITTEN

Relating peaks to Gene Ontology (GO) terms

For that specific step we will use the **GREAT** annotation tools to perform a GO annotation for the ESR1 peaks. Alternatively GREAT can be launched directly from UCSC web server (using Table browser Custom track and by selecting send to GREAT).

Before submitting the peaks, we will need to convert the bed file in a 3-column bed file, which is the required input format for GREAT.

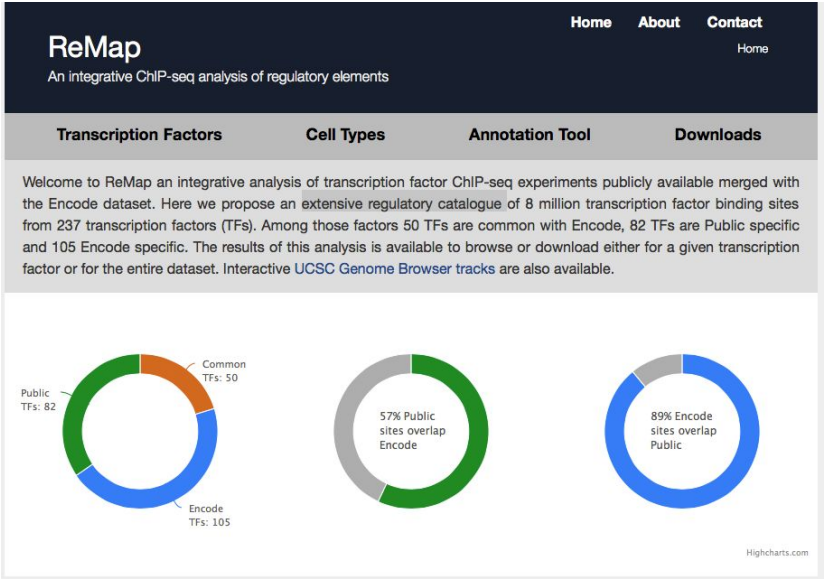
Protocole

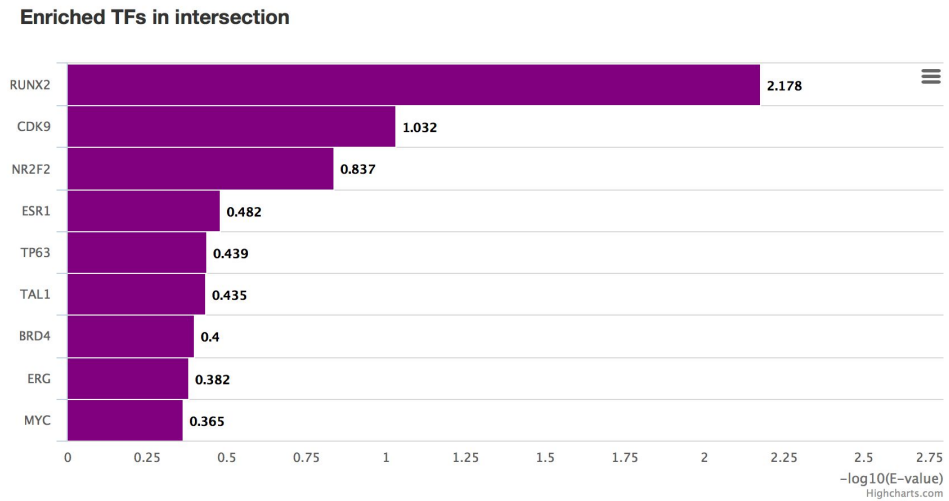
1. In Galaxy, select the tool **Cut** (in the section Text Manipulation).
2. Set “cut columns” to c1,c2,c3
3. **… TO BE COMPLETED …**
4. Connect the GREAT web server http://great.stanford.edu.
5. Select the genome assembly version (hg19).
6. Upload or paste the peaks obtained previously in BED format.
7. Use the whole genome as background and run the software

Q: Examine the enriched functional categories.

ReMap: a regulatory catalog of transcription factor binding sites

- ReMAP (http://tagc.univ-mrs.fr/remap/)
  - Is my peak dataset enriched for known TF peaks?





Integrative ChIP-seq analysis of regulatory elements using ReMap

In this part, we will use the ReMap software to compare the peaks obtained in the peak-calling tutorial to an extensive regulatory catalog of 8 million transcription factor binding regions defined by collecting all the peaks from ChIP-seq experiments from the ENCODE project + other public datasets from the GEO database. Note that on the ReMap Web site, the term “site” is used to denote a ChIP-seq peak, rather than the precise binding location of a transcription factor.

Protocole

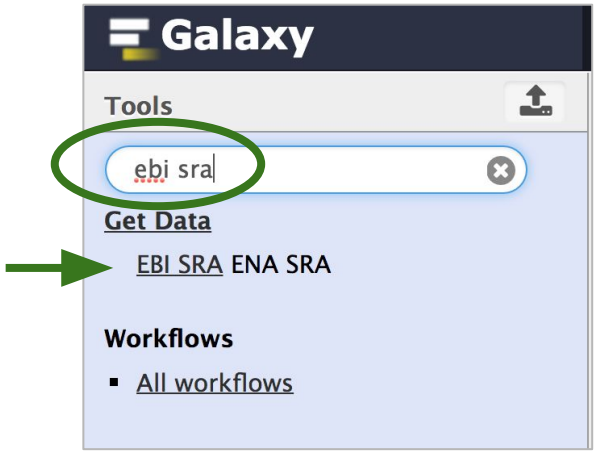
1. Connect the ReMap web server (http://tagc.univ-mrs.fr/remap/)
2. Go to the Annotation Tool
3. upload or paste the peaks in BED format (select BED format in the data format selector)
4. Add your email and run the software with default parameters

Q: What are the TFs associated to your peaks?

Going further: importing full ChIP-seq datasets from SRA/ENA

Importing short reads from RSA via ENA

- Short read sequences are available
  - in the NCBI SRA database
  - in the ENA database, which mirrors the SRA database.
- Galaxy includes a tool called “EBI SRA” that enables to import short read sequences in fastq format.

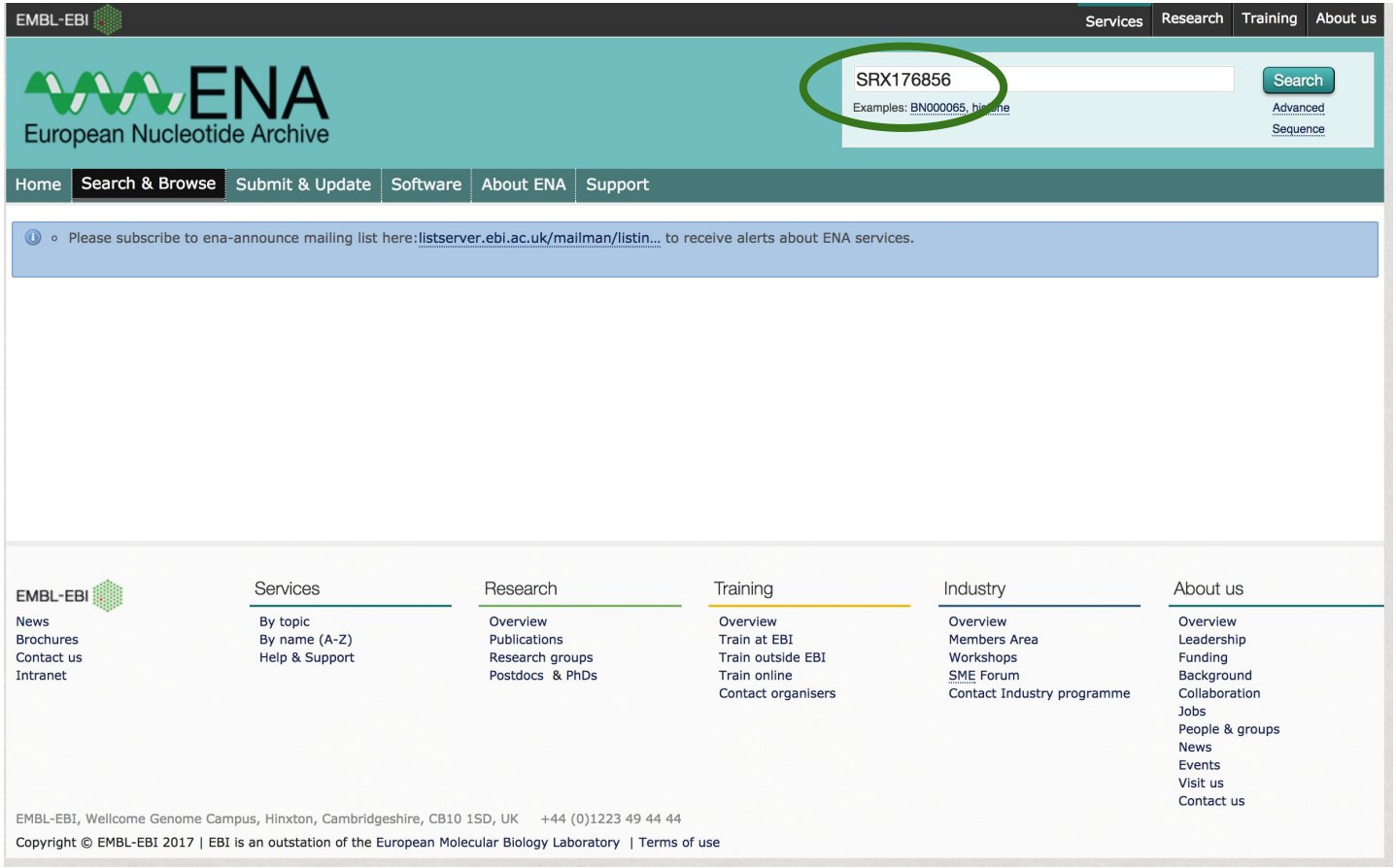


Importing short red sequences from EBI SRA

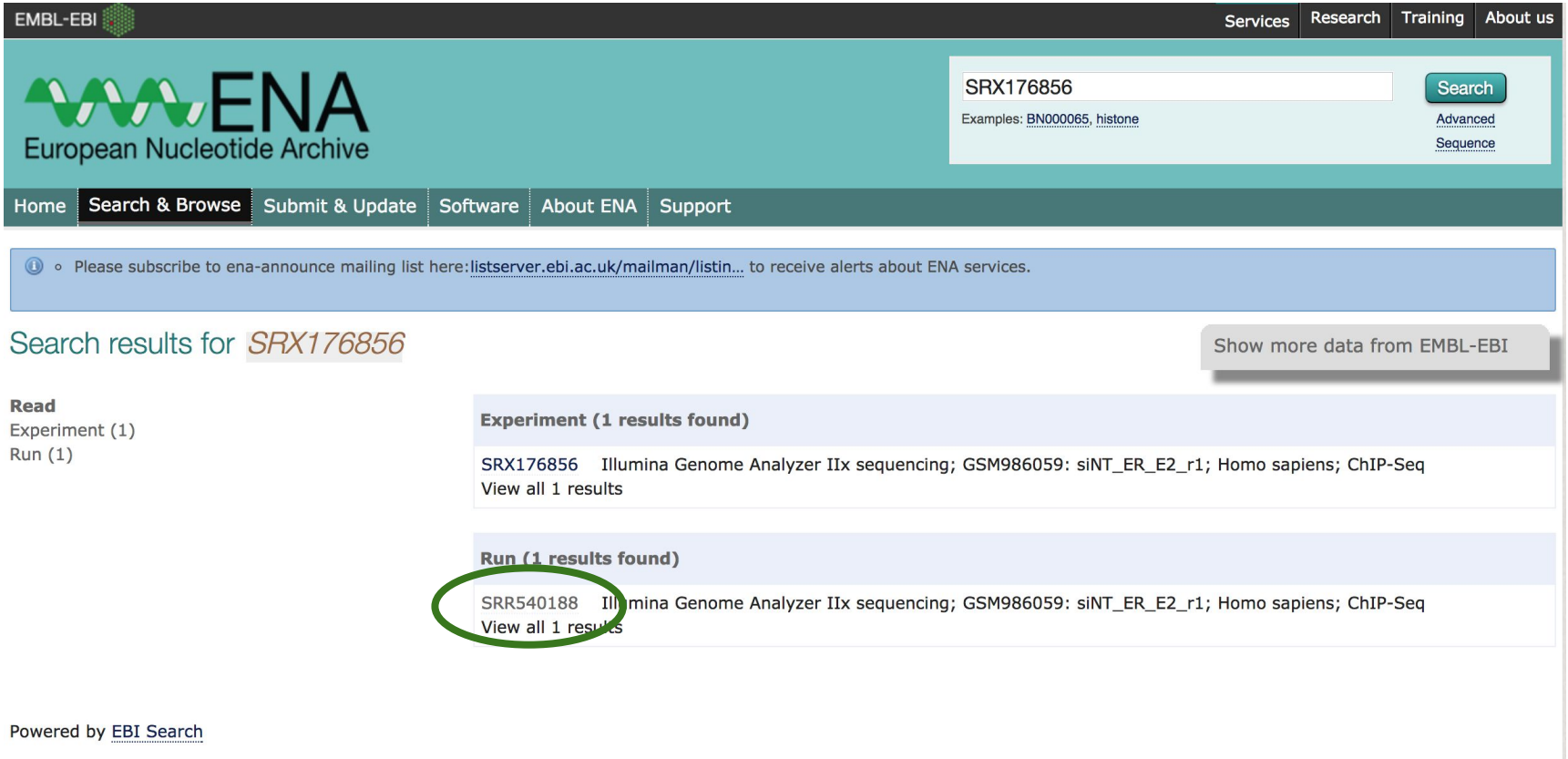
Protocol

1. Create a new history and name it **Theodorou_GATA3_ER**.
2. In the Galaxy toolbox, choose **EBI SRA**. This opens a query form at ENA.
3. Type the identifier of interest (for example the GEO series identifier **GSE40129**) and click **Search**.
4. Above the result table, click **Select columns**, check the option **Experiment title**, the click **Hide selected columns**.
5. Identify the run of interest (e.g. **SRR540188**) and click **File 1** in the column **Fastq files (galaxy)**. This will automatically import the short read file in your Galaxy history.

Find the experiment of interest (SRX176856)



Select the desired run (SRR540188)



Home | Search & Browse | Submit & Update | Software | About ENA | Support

Please subscribe to ena-announce mailing list here:listserver.ebi.ac.uk/mailman/listin... to receive alerts about ENA services.

Contact Helpdesk

Run: SRR540188

Illumina Genome Analyzer IIx sequencing; GSM986059: siNT_ER_E2_r1; Homo sapiens; ChIP-Seq

View: XML **Download:** XML

Submitting Centre	Platform	Model	Read Count	Base Count
GEO	ILLUMINA	Illumina Genome Analyzer IIx	29,991,295	1,079,686,620

Library Layout	Library Strategy	Library Source	Library Selection	Library Name
SINGLE	ChIP-Seq	GENOMIC	ChIP	GSM986059: siNT_ER_E2_r1

Broker Name
NCBI

Navigation | Read Files

This table contains the files for run SRR540188

Bulk Download Files (Using Firefox is recommended to launch the bulk downloader app.)

Download: 1 - 1 of 1 results in TEXT

Select columns

Showing results 1 - 1 of 1 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	Fastq files (ftp)	Fastq files (galaxy)	Submitted files (ftp)	Submitted files (galaxy)	NCBI SRA file (ftp)	NCBI SRA file (galaxy)	CRAM Index files (ftp)	CRAM Index files (galaxy)
PRJNA172877	SAMN01113304	SRS356212	SRX176856	SRR540188	9606	Homo sapiens	Illumina Genome Analyzer IIx	SINGLE	File 1	File 1			File 1	File 1		

Prev Next **Go to page:** 1 Go

For Aspera download, please download and install Aspera Connect