

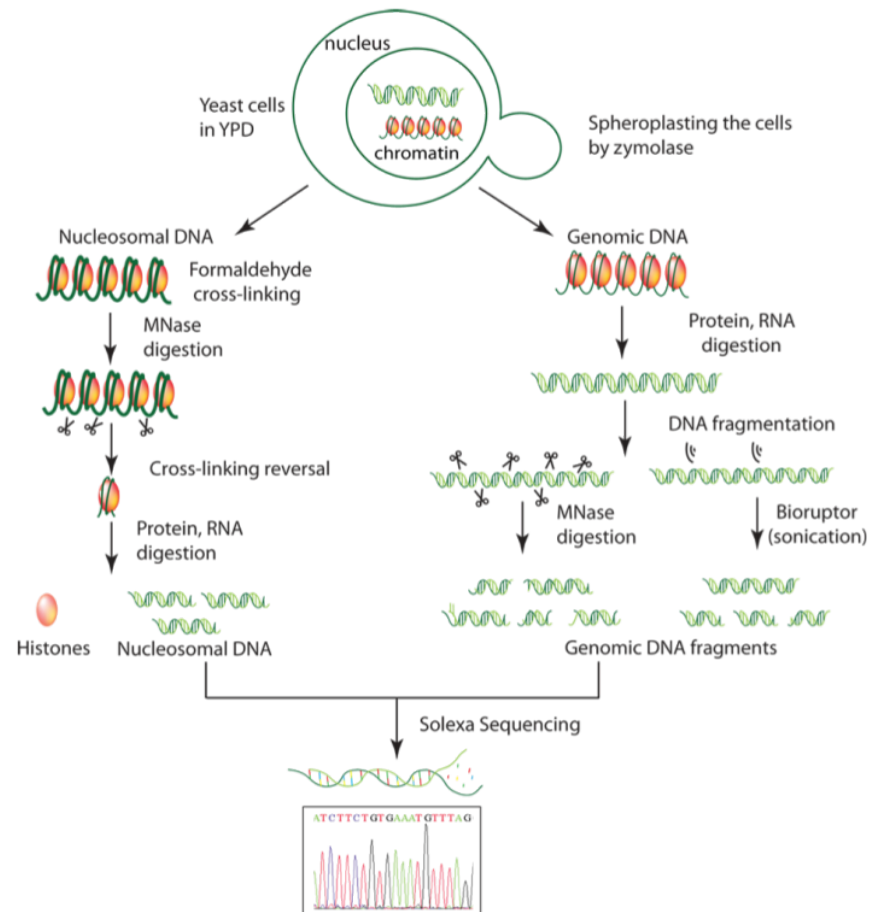
NucleosomeDynamics portal

Ricard Illa Pujagut
Molecular Modeling and Bioinformatics Group



A portal to easily analyze and visualize
MNase-seq data

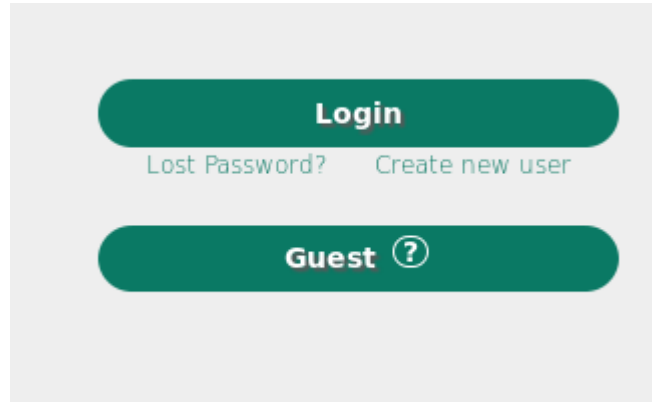
MNase-seq



NucleosomeDynamics portal

url: <http://mmb.irbbarcelona.org/NucleosomeDynamics/>

User creation

A light gray rectangular box containing two dark green rounded buttons. The top button is labeled 'Login' in white text, with the links 'Lost Password?' and 'Create new user' in a smaller, lighter green font below it. The bottom button is labeled 'Guest' in white text, followed by a white question mark icon in a circle.

Login
Lost Password? Create new user

Guest ?

User Registration:

A registration form with a light gray border. It contains five labels on the left and corresponding input fields on the right. The labels are 'Email:', 'Name', 'Surname', 'Institution', and 'Password (twice):'. The 'Password (twice):' label is positioned to the left of two adjacent input fields.

Email:

Name

Surname

Institution

Password (twice):

Register

Your guest user

ND58002de0e3ba9

<http://mmb.irbbarcelona.org/NucleosomeDynamics/datamanager/workspace.php?userId=ND58002de0e3ba9>

User workspace

- Users can
 - Upload inputs
 - Run calculations
 - Visualize those calculations
 - See some statistics on the calculations run

The screenshot displays a web-based user workspace interface. At the top, there are buttons for 'Upload file', 'Run', 'Visualize', and 'Statistics'. A search bar is located on the right. Below these, a table lists files under the 'File' tab. The table has columns for 'Format', 'Project', 'Date', 'Size', 'Expires', and 'Actions'. The files listed are 'sample_cellcycle_M.bam' and 'sample_cellcycle_G1.bam', both of size 3.52 M and 4.26 M respectively, with a date of 2016/07/22 12:10. The 'Actions' column contains icons for edit, download, and delete. At the bottom, there is a 'Disk use' section showing 7.78 MB of 24 GB (0.03%) used.

File	Format	Project	Date	Size	Expires	Actions
File	All	All				
uploads		uploads	2016/07/22 12:10	0.00 B		
sample_cellcycle_M.bam	BAM	uploads	2016/07/22 12:10	3.52 M	-40	
sample_cellcycle_G1.bam	BAM	uploads	2016/07/22 12:10	4.26 M	-40	

Showing 1 to 3 of 3 entries

Previous 1 Next

Upload file Run Visualize

Disk use 7.78 MB of 24 GB (0.03%)

Upload file

- BAM: main input, MNase-seq reads mapped to a reference genome
- GFF/GTF or GFF3: to add additional annotations
- Bedgraph, Wig, Bed or Bigwig: to visualize continuous data as well
- Any of those formats compressed as .gz

The interface is divided into two main sections for file selection:

- Select from your computer:** Contains a "Browse..." button and the text "No files selected.".
- Download from web by URL:** Contains a URL input field with the placeholder text "http://public/path/to/file" and a small document icon with "URL" written on it.

A central "Upload File" button is located below these two sections.

Below the upload options, there are three input fields:

- File Format:** A dropdown menu with the text "Select the file format".
- Reference genome:** A dropdown menu with the text "Select the reference genome".
- Description:** A text area with the placeholder text "Write a short description here" and "..." below it.

At the bottom of the form, there are two buttons: "Validate" and "Cancel".

Run calculations

- NucleR for peak calling and nucleosome positioning
- Nucleosome Dynamics to look for changes between MNase-seq experiments
- Detection Nucleosome-Free Regions
- Analysis of nucleosome phasing in gene bodies
- Transcription start site classification
- Stiffness constant estimation

Inputs

1	2016/07/22 12:10	3.52 M	sample_cellcycle_M.bam
2	2016/07/22 12:10	4.26 M	sample_cellcycle_G1.bam

[Edit inputs](#) Selected analyses will run for each input file

Analyzes

Project name: run002

Short description:

☐ NucleR

☐ Nucleosome Dynamics

☐ Nucleosome Free Regions

☐ Nucleosome phasing

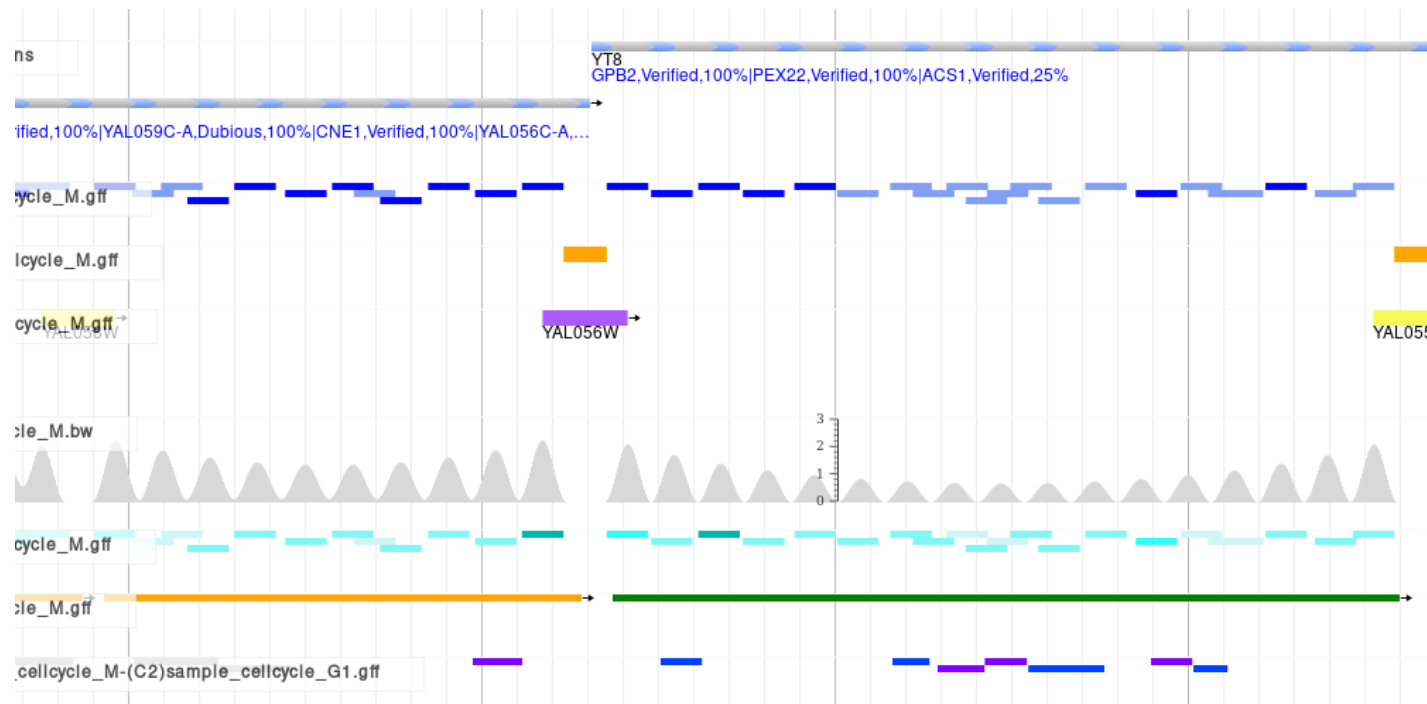
☐ TSS classification

☐ Stiffness

Compute

Visualize results

- Information can be visualized using the Jbrowse genome browser
 - Calculation results
 - Annotations uploaded by the user
 - Pre-uploaded annotations from reference papers



View statistics

- Statistics and a summary table for all genes are generated from the calculation results

General statistics

Statistics by gene

NR_sample_cellcycle_M.gff

run001

NucleR

More

NFR_sample_cellcycle_M.gff

run001

Nucleosome Free Regions

More

TSS_sample_cellcycle_M.gff

run000

TSS Classification

More

STF_sample_cellcycle_M.gff

run000

Stiffness

More

P_sample_cellcycle_M.gff

run000

Periodicity

More

ND_(C1)sample_cellcycle_M-(C2)sample_cellcycle_G1.gff

run000

Nucleosome Dynamics

More

General statistics

Statistics by gene

Download CSV

Show 10 entries

Search:

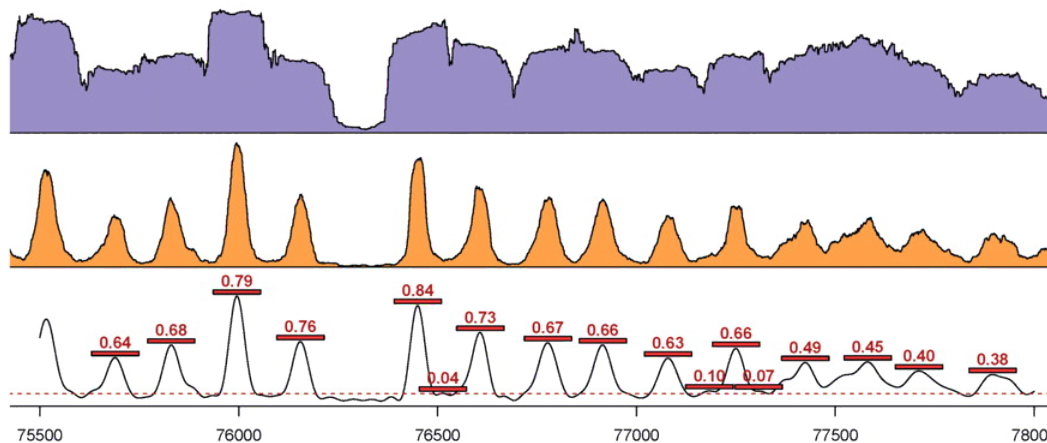
Gene	Total Fuzzy	Total Uncertain	TSS_sample_cellcycle_M.gff	STF_sample_cellcycle_M.gff	P_sample_cellcycle_M.gff	ND_(C1)sample_cellcycle_M-(C2)sample_cellcycle_G1.gff			
			TSS class	Distance from -1 to +1	Mean_STF	StdDev_STF	Score phase	Score autocorrelation	Included
1	0		F-open-W	280	0.219273238424362	0.181298551724703	10	0.642151396199483	0
7	0		-1_missing	NA	0.16447883509854	0.13996739927808	6.5	0.807172006756489	0
28	0		-1_missing	NA	0.116543860395167	0.0710435954568777	44	0.761777007429757	0
3	0		W-open-W	278	0.265061360250617	0.243429862284676	53	0.513681894065867	0
4	0		W-open-W	278	0.214558135258915	0.173991819106295	11	0.665636881288794	0
6	0		F-open-W	257	0.165556062465743	0.0873340717406752	54	0.684717472788911	0
1	0		W-open-W	249	0.150833639752188	0.0545407302316928	19	0.55548189359363	0
6	0		W-open-W	249	0.169317493401558	0.0828388201538131	67	0.622626339292175	2
2	0		F-open-W	223	0.26872356899037	0.169662999021871	66	0.516215826226627	0
0	0		F-open-W	242	0.293063235968596	0.184996958939678	11	0.542746577915214	0

Showing 31 to 40 of 5,680 entries

Previous 1 2 3 4 5 ... 568 Next

NucleR

- NucleR is used for peak calling and nucleosome positioning



 NucleR 

NucleR finds nucleosome positions from MNase experiments using Fourier transform filtering and classifies nucleosomes according to their fuzziness ([more information](#))

Width 	Minimum Overlap 	Dyad Length 
<input type="text" value="147"/>	<input type="text" value="80"/>	<input type="text" value="50"/>
Background level 	Height Threshold 	Width Threshold 
<input type="radio"/> Absolute value 	<input type="text" value="0.4"/>	<input type="text" value="0.6"/>
<input checked="" type="radio"/> Percentage 		
<input type="text" value="35%"/>		

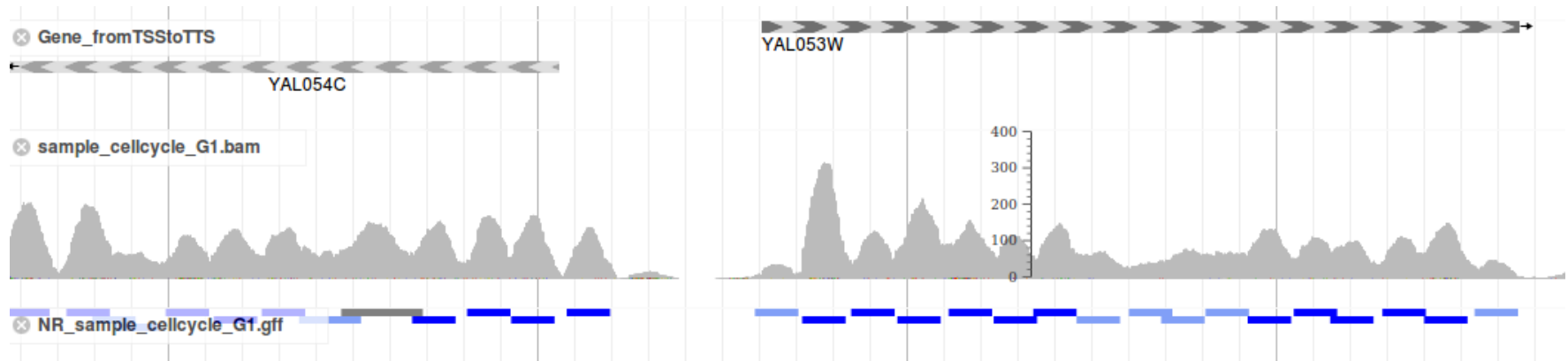
 Advanced Settings

Coverage Smoothing 

NucleR's parameters

- Width: Size of each nucleosome, in bp, to be considered by NucleR. Default = 147.
- Background level: Minimum number of reads (Coverage) to call a nucleosome. Can be given as a percentage or as an absolute coverage value. Default = 35%.
- Dyad Length: Length of the reads that should be used for nucleosome calling to define the dyad of the nucleosomes. Default = 50.
- Minimum Overlap: Minimum overlap between two nucleosomes for merging them. Default:80
- Width Threshold: Derived from the dispersion of the reads around the dyad. Used for W/F classification. Default = 0.6.
- Height Threshold: Derived from the number of reads at the dyad. Used for W/F classification. Default = 0.4.
- Coverage smoothing: Number of signal components used in the fourier transformation filtering. Default = 0.02.

Visualization of NucleR's output



Nucleosome

Primary Data

Type	Nucleosome
Score ?	0.963747880020362
Position	chrI:45897..46043
Length	147 bp

Attributes

class ?	W
score_height ?	0.9999999999998
score_width ?	0.927495760040923
source	nucleR

Region sequence

[FASTA](#)

```
>chrI chrI:45897..46043 class=Nucleosome length=147
CCATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACTGTTCGTACTGTGCAGCGGTA
CAGCACGTTCTCTGACACAAACGACACTACTCCGGCGTCTGCAAAGCATTTCAGACCACTT
CTTTATTGACGTGTATGGACA
```

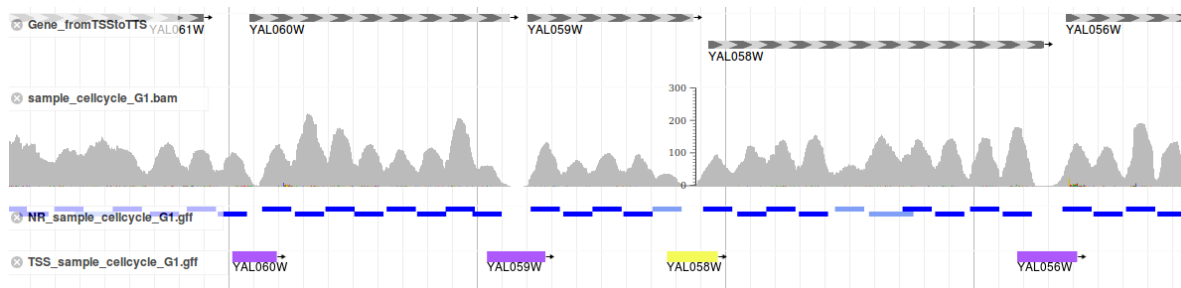
OK

NucleR's output parameters

- Score: Positioning score. Calculated from score_height and score_width
- Class: Can be W for well-positioned nucleosomes, F for fuzzy nucleosomes and uncertain for extremely fuzzy nucleosomes.
- score_height: Indicates how big or small the peak is.
- score_width: Indicates for sharp or wide the peak is

Classification of Transcription Start Sites

- Nucleosome -1 can be W, F or missing
- Nucleosome +1 can be W, F or missing
- Distance between both can be open, closed or overlapped



About track: TSS_sample_cellcycle_G1.gff

Name	TSS_sample_cellcycle_G1.gff
Category	Your Data / Projects / run000
Type	TSS nucleosome classification
Description	Classification of The Transcription Start Sites (TSS) according to the nucleosome architecture. They are classified based on the width of the NFRs (closed (c) or open (o)) and on the fuzziness of the -1 and +1 nucleosomes (missed (M), fuzzy (F) or well-positioned (W)).
Details	<ul style="list-style-type: none">-1_missingF-close-FF-close-WF-open-FF-open-WF-overlap-FF-overlap-WW-close-FW-close-WW-open-FW-open-W
Track type	JBrowse/View/Track/CanvasFeatures

OK

NucleR's statistics

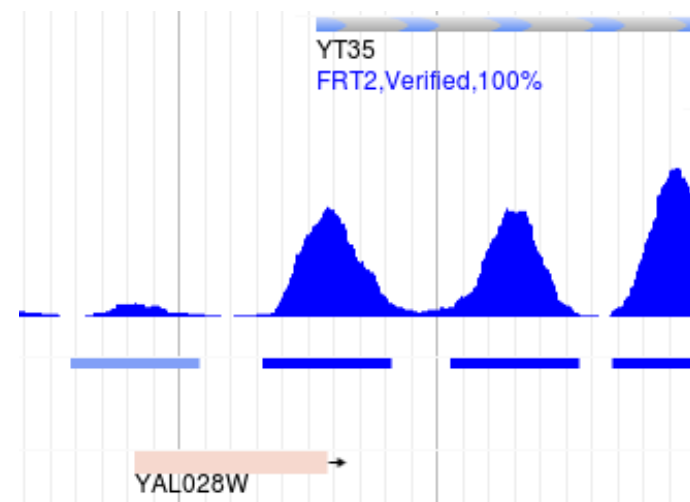
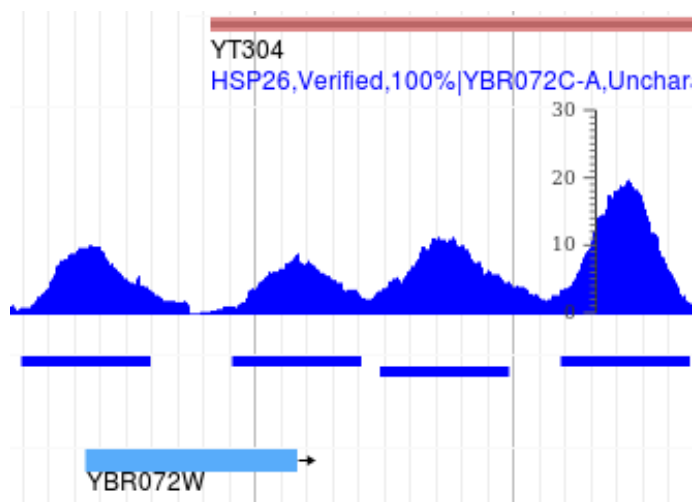
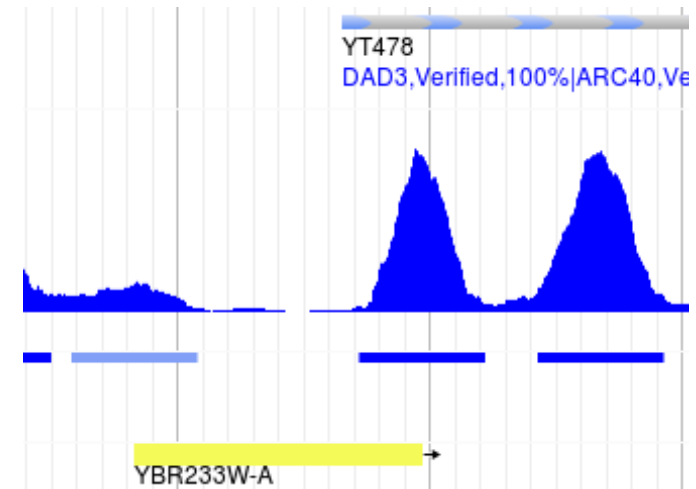
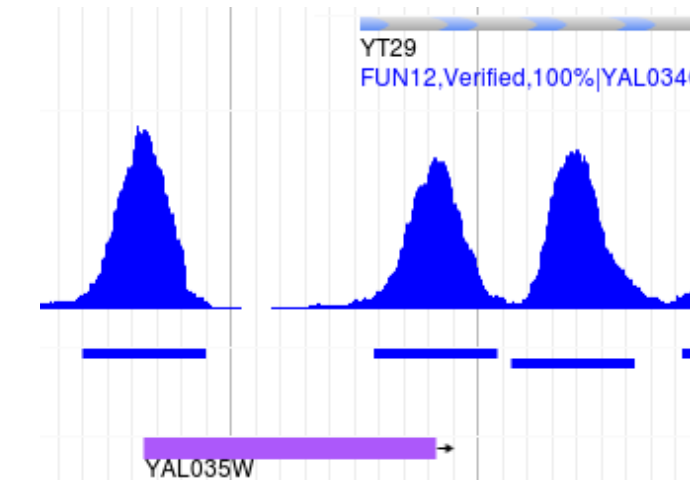
Class	Frequency
Fuzzy	336
Uncertain	1
Well-positioned	267
Total	604

Show entries

Search:

NR_sample_cellcycle_M.gff				
Name	Total Nucleosomes	Total Well-Positioned	Total Fuzzy	Total Uncertain
YAL027W ▾	6	5	1	0
YAL028W ▾	11	4	7	0
YAL029C ▾	36	8	28	0
YAL030W ▾	5	2	3	0
YAL031C ▾	17	13	4	0
YAL032C ▾	10	4	6	0
YAL033W ▾	4	3	1	0
YAL034C ▾	12	6	6	0
YAL034C-B ▾	8	6	2	0
YAL034W-A ▾	6	6	0	0

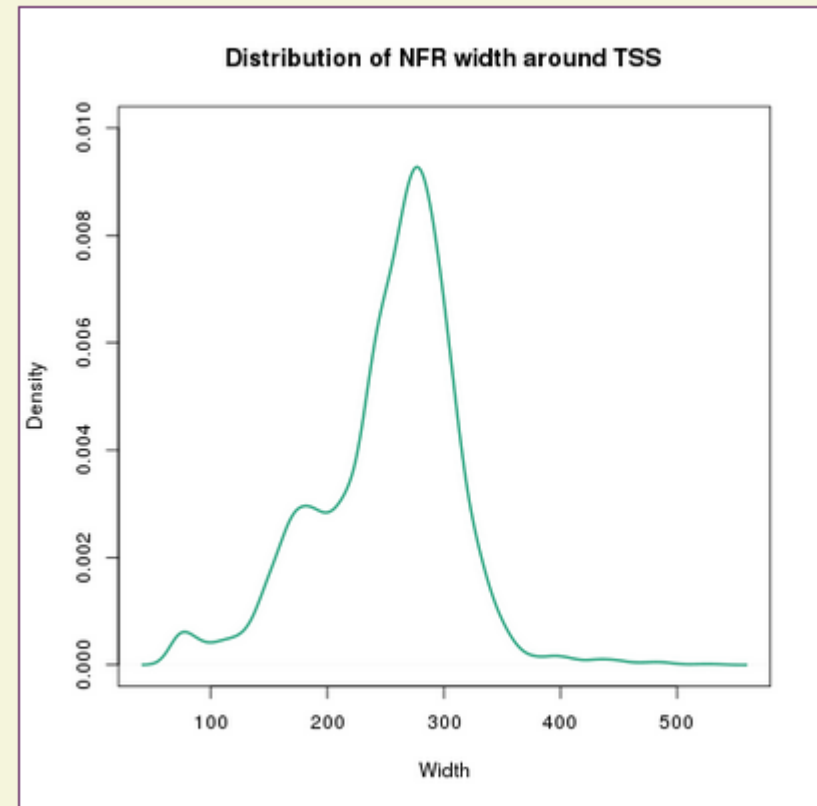
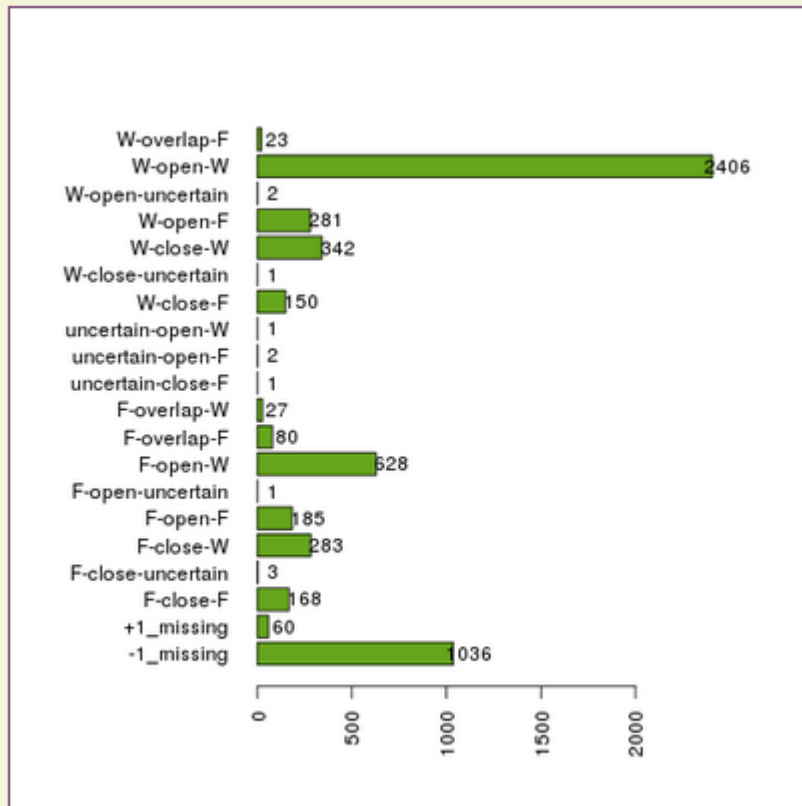
Classification of Transcription Start Sites



Transcription Start Sites output parameters

- classification: TSS classification (ex: W-close-W, F-open-W, etc)
- distance: Distance between the nucleosomes -1 and +1
- nucleosome_minus1: Position of the nucleosome -1
- nucleosome_plus1: Position of the nucleosome +1
- TSS_position: Position of the Transcription Start Site

Transcription Start Sites' Statistics



Nucleosome Free Regions

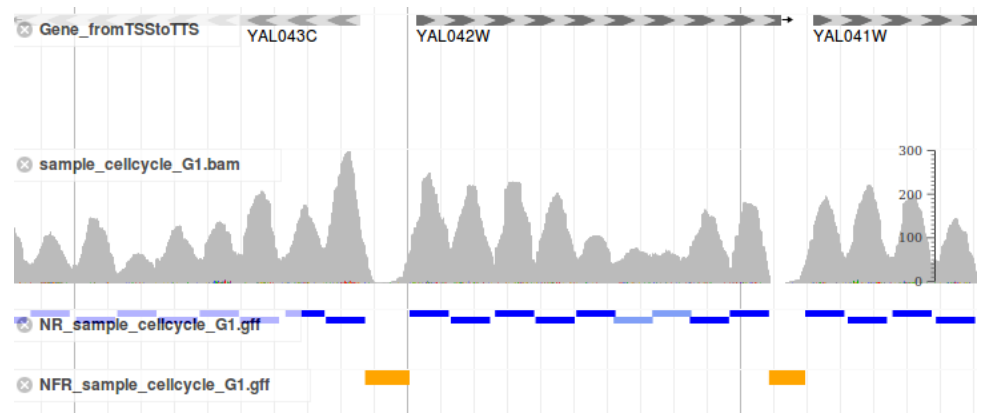
- Maximum Width: Maximum width for a linker fragment to be considered a nucleosome-free region. Default = 400.
- Minimum Width: Minimum width for a linker fragment to be considered a nucleosome-free region. Default = 110.

☒ Nucleosome Free Regions

Nucleosome-free regions (NFR) are regions depleted of nucleosomes and larger than an average linker fragment. ([more information](#))

Minimum Width

Maximum width



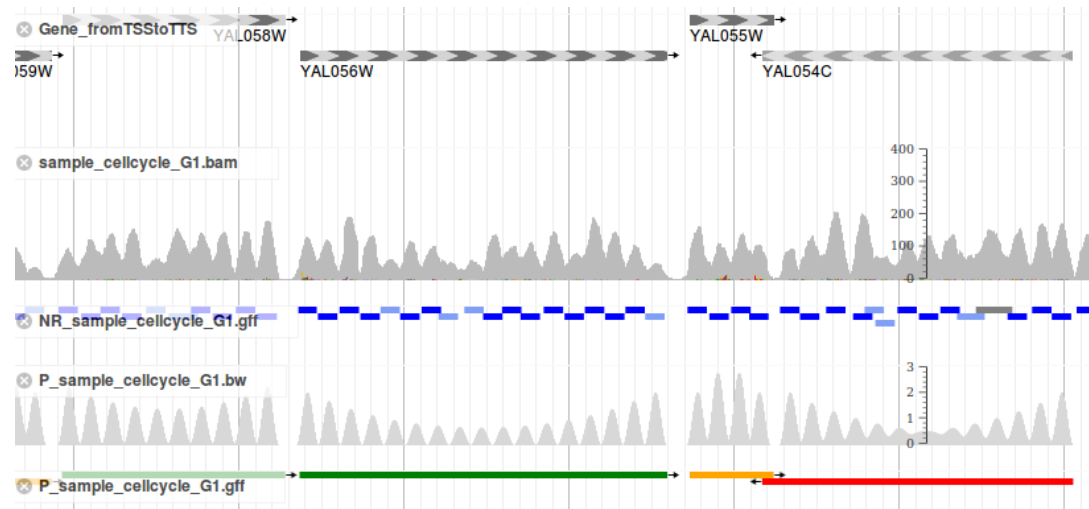
Nucleosome Phasing

- Analyses on the periodicity of nucleosomes in each gene
- Parameter: period
- $\text{phase}_{\text{score}} = \text{nuc}_{\text{dist}} - T [\text{nuc}_{\text{dist}}/T]$
- $\text{autocor}_{\text{score}} = R(T) / T(0)$

Nucleosome Phasing output parameters

- nucleosome_first: position of the first nucleosome of the gene
- nucleosome_last: position of the last nucleosome of the gene
- score_autocorrelation: autocorrelation of the nucleosome coverage in the gene body
- score_phase: phase score of the gene
- a bw file showing a theoretical periodic coverage

Nucleosome Phasing



About track: P_sample_cellcycle_G1.gff

Name	P_sample_cellcycle_G1.gff
Category	Your Data / Projects / run000
Type	Phasing
Description	Nucleosome phasing along a given gene between the first and last nucleosome
Details	<p>Score phase: remainder of the division of the distance between first and last nucleosomes by the period specified by the user. Values between 0 and 82. Number of nucleotides left after dividing the distance between the first and the last nucleosome by the period. A score phase of 0 means the nucleosomes are completely phased and a score of 82 corresponds to totally anti-phased nucleosomes. Depending on the score phase, the tracks have the following colors:</p> <ul style="list-style-type: none">Score [0-25]Score [26-55]Score [56-82] <p>Score autocorrelation: autocorrelation function computed from the experimental coverage of a given gene using the period given by the user. Values are between 0 (nonperiodic) and 1 (periodic).</p>
Track type	JBrowse/View/Track/CanvasFeatures


OK

Phasing Statistics

Phased genes	1430
Not-phased genes	1954
Other genes	2263

Show 10 ▾ entries

Search:

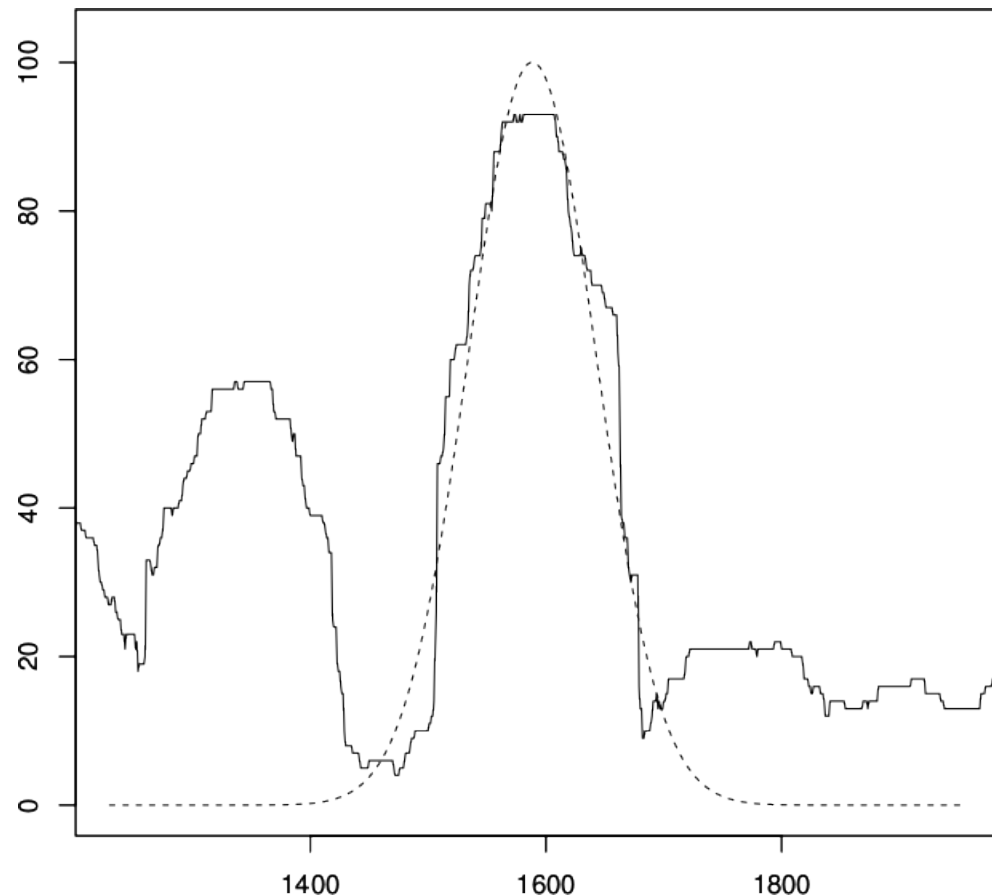
 <input type="text" value="Name"/>		P_rep2_30m_S.gff	
		Score phase	Score autocorrelation
YAL017W ▾		6	0.789500351093349
YAL018C ▾		42	0.721704056819486
YAL019W ▾		47	0.806241546302602
YAL020C ▾		54	0.700257662449389
YAL021C ▾		31	0.79071004916251
YAL022C ▾		40	0.69078945442945
YAL023C ▾		40	0.785182668058834
YAL024C ▾		22	0.770969469340187
YAL025C ▾		69.5	0.758468905213693
YAL026C ▾		64	0.739794730102729

Showing 21 to 30 of 5,680 entries

Previous 1 2 3 4 5 ... 568 Next

Stiffness constant

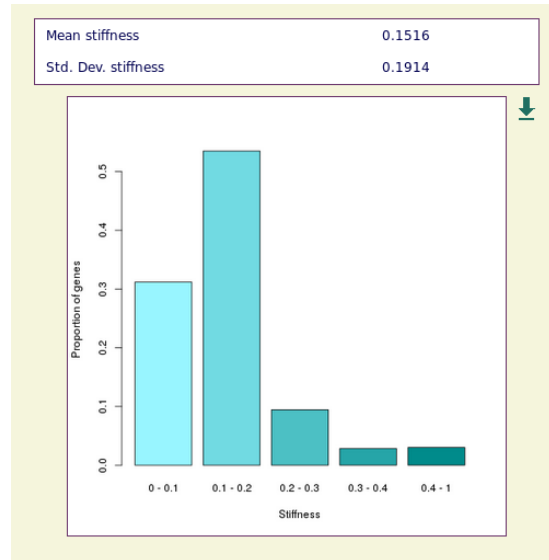
- A stiffness can be estimated from a gaussian fitting
- The results are given as kcal/mol/bp²
- $f(x) = ke^{-(x-m)^2/2sd^2}$
- $\text{stiffness} = Rt/sd$



Stiffness constant's output parameters

- Score: Estimated stiffness for that nucleosome, expressed in kcal/mol/bp²
- gauss_k: Height of the fitted gaussian curve
- gauss_m: Mean of the fitted gaussian curve
- gauss_sd: Standard deviation of the fitted gaussian curve

Stiffness's Statistics



Show 10 entries

Search:

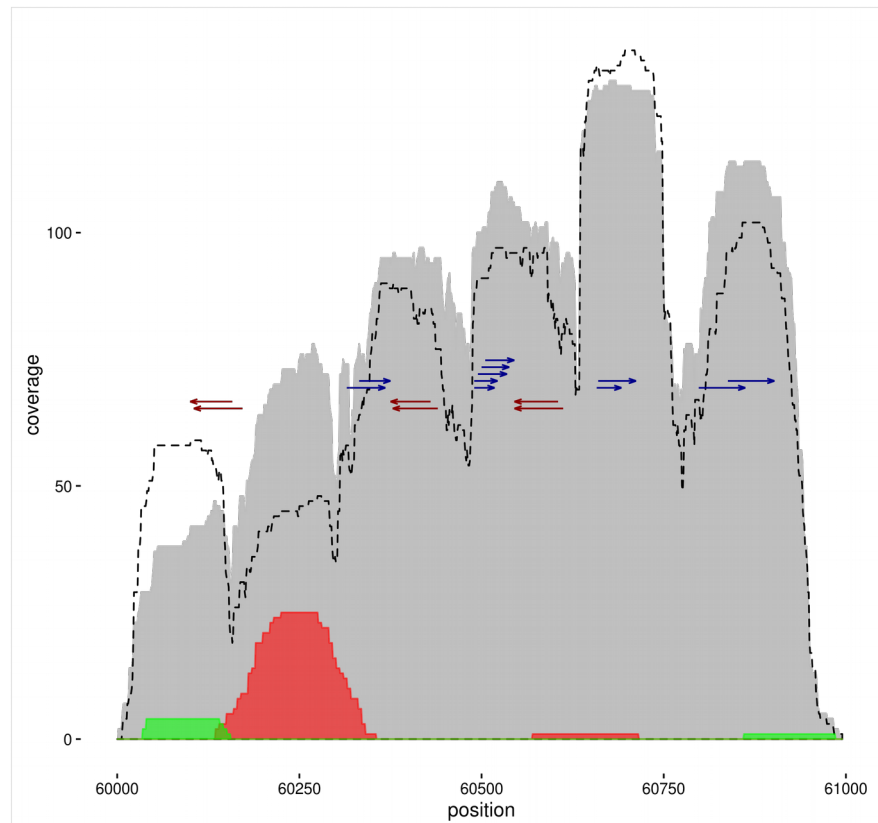
Name	STF_rep2_30m_S.gff		P_rep2_30m_S.gff	
	Mean_STF	StdDev_STF	Score phase	Score autocorrelation
YAL008W	0.131515550193258	0.0508612821810746	67	0.70816568748032
YAL009W	0.152205087325554	0.0552656834858515	6	0.6305777337569
YAL010C	0.170579424256106	0.102461531977557	52	0.569972873624679
YAL011W	0.157881134832941	0.0625068037477226	9	0.792309164688768
YAL012W	0.116490597214048	0.0465897802298949	13.5	0.796861566597235
YAL013W	0.321043118254027	0.373073905951499	15	0.755382351590362
YAL014C	0.133334742830977	0.0805096512118026	60	0.666040033966019
YAL015C	0.154861059632687	0.040228528768057	20	0.73744717606401
YAL016C-B	0.201962945574583	0.0318371374370273	17	NA
YAL016W	0.113670420321227	0.0491440156356997	13.5	0.660726466575716

Showing 11 to 20 of 5,680 entries

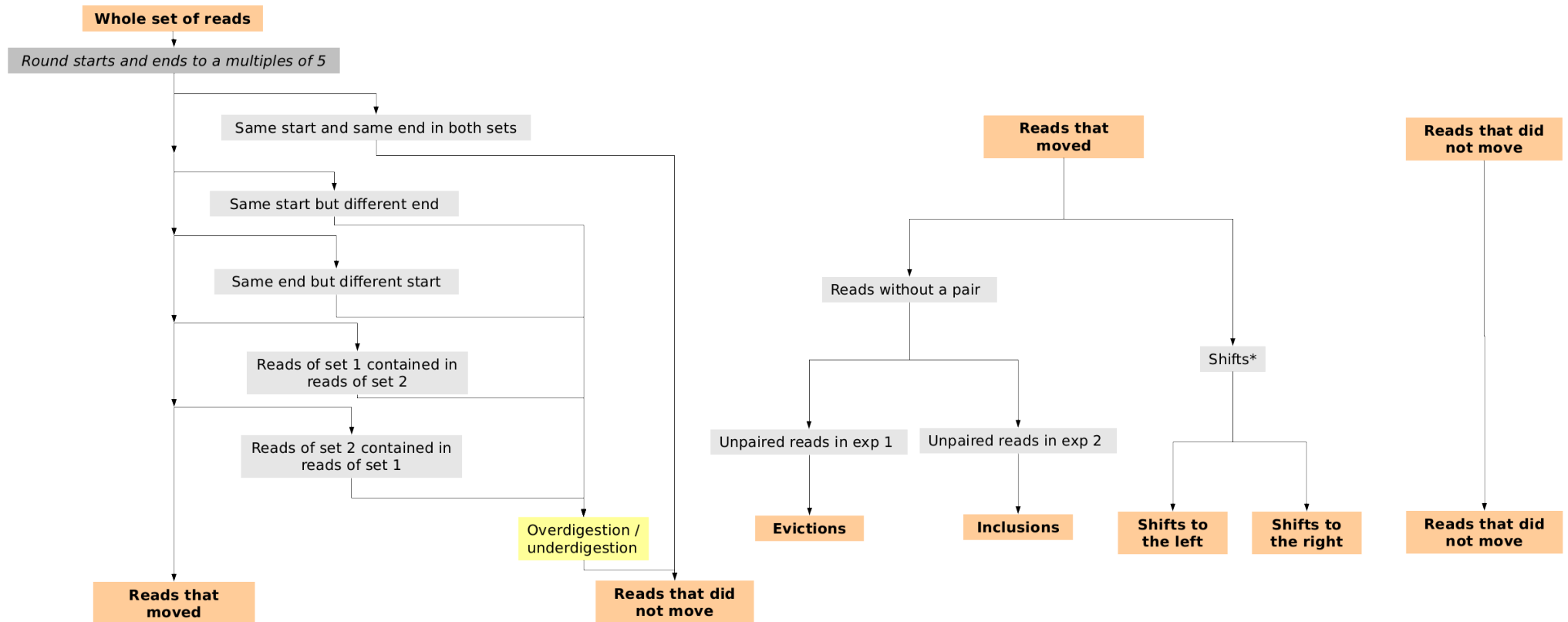
Previous 1 2 3 4 5 ... 568 Next

NucleosomeDynamics

- Allows to compare between two MNase-seq experiments
- Works at a read-level



NucleosomeDynamics



NucleosomeDynamics parameters

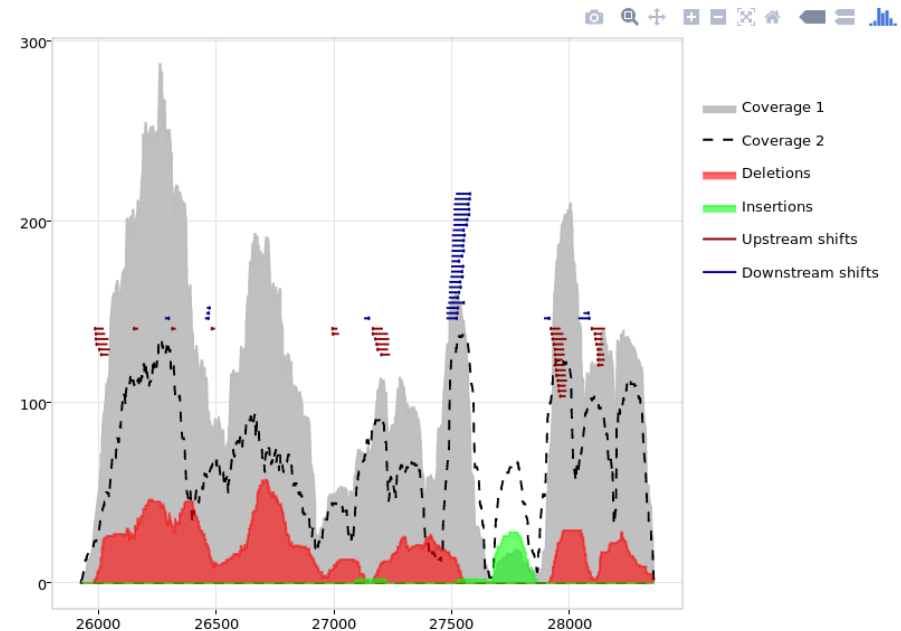
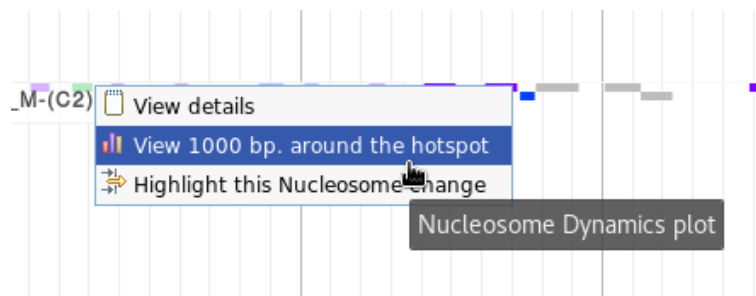
- Genomic Range: Portion of the genome to be analyzed (default: all of it)
- Maximum Diff: Maximum distance from beginning to end on shifts (default 70 bp)
- Maximum Length: Maximum lengths of reads to be considered (default 140 bp)
- Equal Size: If set to TRUE, all reads will be set to the same length (default: FALSE)
- Combined: If set to TRUE, nearby movement hotspots will be combined (default: TRUE)
- Round Power: Read lengths will be rounded to the closest power of this number (default: 5)
- Same Magnitude: When combining two hotspots this is the maximum ratio value between two hotspots for them to be considered the same magnitude
- Shift minimum num. reads: Minimum number of reads in a shift hotspot
- Shifts threshold: Minimum score in a shift hotspot
- Indels minimum num. reads: Minimum number of reads in an indel hotspot
- Indels threshold: Minimum score in an indel hotspot

NucleosomeDynamics' output parameters

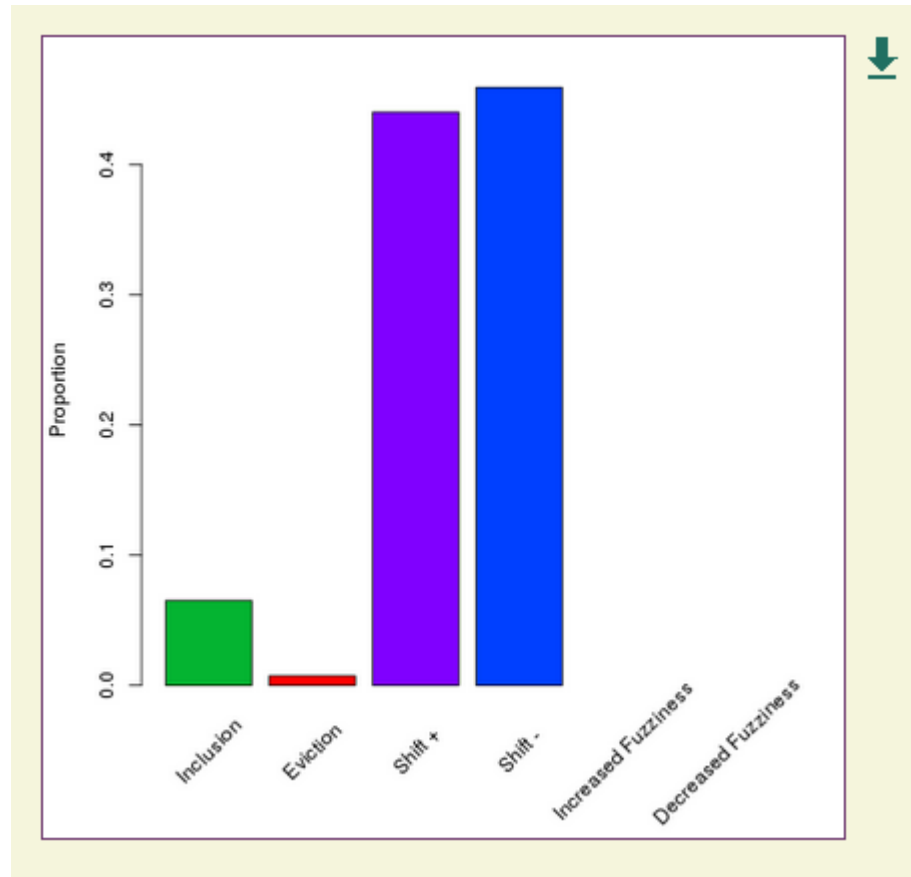
- Score: How relevant that movement hotspot is. It is calculated as the movement area relative to the coverage area in that zone.
- class: Type of movement. Can be
 - Shift -: Upstream shift
 - Shift +: Downstream shift
 - Inclusion: Increased coverage in condition 2 relative to condition 1
 - Eviction: Decreased coverage in condition 2 relative to condition 2
 - Increased fuzziness: Shift - followed by a shift + in the same nucleosome
 - Decreased fuzziness: Shift + followed by a shift - in the same nucleosome
- nreads: Number of reads at the hotspot peak

NucleosomeDynamics plot

- Additionally, an interactive plot showing the movements can be viewed using plotly



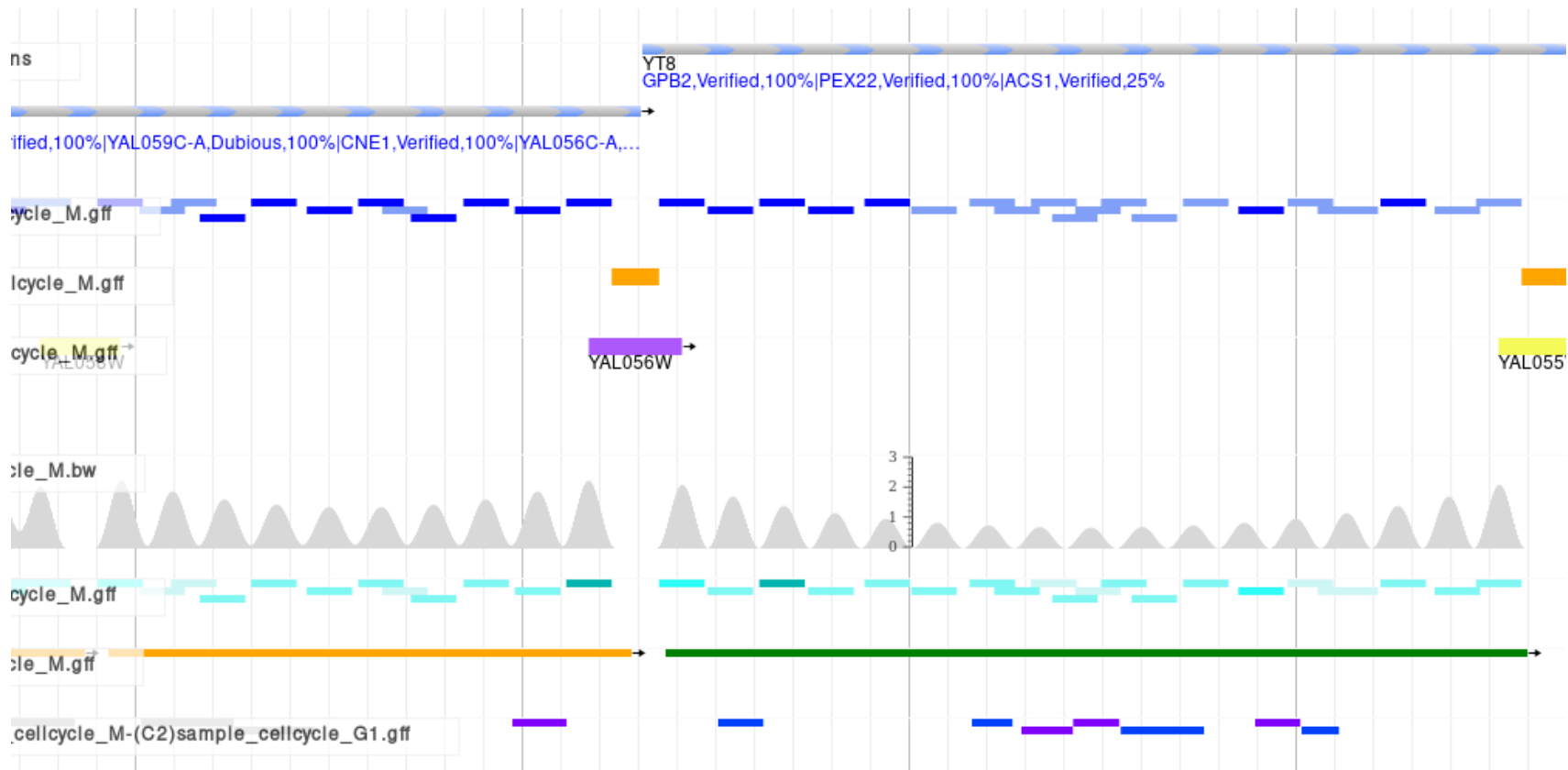
NucleosomeDynamics' Statistics



Additional reference annotations available

▼ Reference sequence	1
<input checked="" type="checkbox"/> Reference sequence	
▼ Genes and Gene Predictions	3
<input type="checkbox"/> SGD_other	
<input checked="" type="checkbox"/> SGD_ORF	
<input checked="" type="checkbox"/> Gene_fromTSStoTTS	
▼ Mapping and Sequencing	1
<input type="checkbox"/> GC content	
▼ Gene Structure	4
▶ UTRs	2
▶ Gene Models	2
▼ RNA Expression Profiling	2
▶ mRNA	1
▶ Unannotated	1
▼ Chromatin structure	50
▼ Histone modifications	37
▶ Kirmizis et al. (2007)	18
▶ Guillemette et al. (2011)	4
▶ Liu et al. (2005)	15
▼ Nucleosome organization	8
▶ Mavrich et al. (2008)	1
▶ Field et al. (2008)	3
▶ Schep et al. (2015)	4
▼ DNase I hypersensitive sites	1
▶ Hesselberth et al. (2009)	1
▼ Histone variants	4
▶ Albert et al. (2007)	1
▶ Guillemette et al. (2005)	3
▼ Replication and Recombination	6
▶ Origins of replication	6
▼ Transcription Regulation	24
▼ Transcription Factors	5
▶ Rhee et al. (2012)	1
▶ Rhee et al. 2011	4
▼ RNA Pol II ChIP	19
▶ Mayer et al. 2010	19

NucleosomeDynamics portal allows to visualize and integrate analysis derived from MNase-seq data



Akwnowledgements

Laia Codó

Diana Buitrago

Isabelle Heath

Romina Royo

Pau de Jorge

Josep Lluís Gelpí

Adam Hospital

Oscar Flores

Modesto Orozco

