

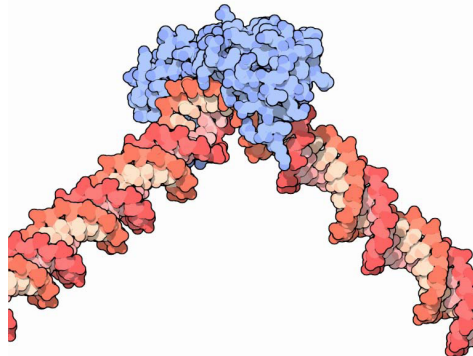
Análisis de datos de ChIP-seq: Identificación de Sitios de Unión de Factores de Transcripción y Marcas Epigenéticas

Francisco J. Romero Campero
<http://www.cs.us.es/~fran/>

**Dpt. de Ciencias de la Computación e
Inteligencia Artificial
Universidad de Sevilla**

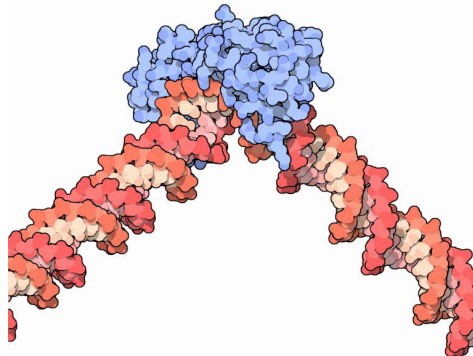
Los Factores de Transcripción se unen a regiones cis para regular la transcripción

- Los **Factores de Transcripción (FTs)** son proteínas que controlan la transcripción de genes mediante la unión física directa con ciertos patrones de DNA llamados **motivos**.
- Comúnmente, estos motivos están localizados aguas arriba de los genes regulados en **regiones cis**.



El cistroma o conjunto de regiones cis asociadas a un FT puede ser enormemente plástico

- El conjunto global de sitios de unión de un factor de transcripción se denomina **cistroma**.
- El cistroma de un FT puede ser **enormemente plástico** ya que condiciones externas e internas pueden cambiar sustancialmente su estado o el del correspondiente complejo proteico produciendo la unión a diferentes regiones cis.



ChIP-Seq determina el cistroma de un FT

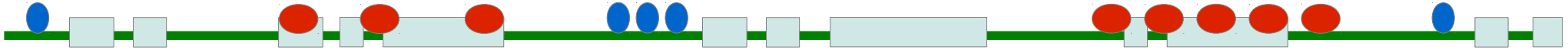
- La técnica ómica que permite determinar el cistroma de un factor de transcripción en unas condiciones específicas se denomina **ChIP-Seq** (Chromatin Immunoprecipitation coupled with Sequencing).
- Esta técnica combina dos metodologías ya establecidas:
 - **ChIP**: Chromatin Immuno-Precipitation
 - **Seq**: High throughput sequencing of DNA

Chromatin Immunoprecipitation coupled with sequencing

Chromatin Immunoprecipitation coupled with sequencing



Chromatin Immunoprecipitation coupled with sequencing



Chromatin Immunoprecipitation coupled with sequencing

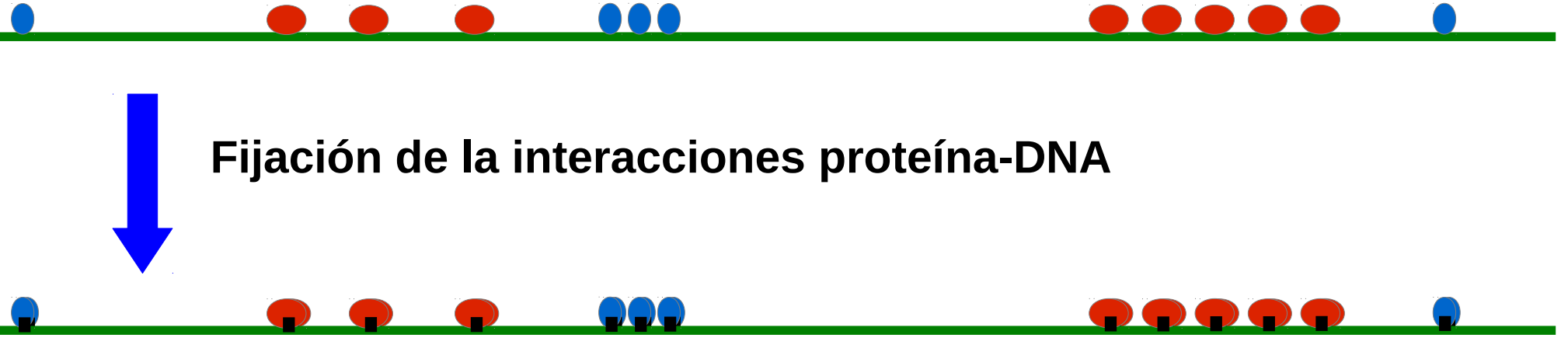


Chromatin Immunoprecipitation coupled with sequencing

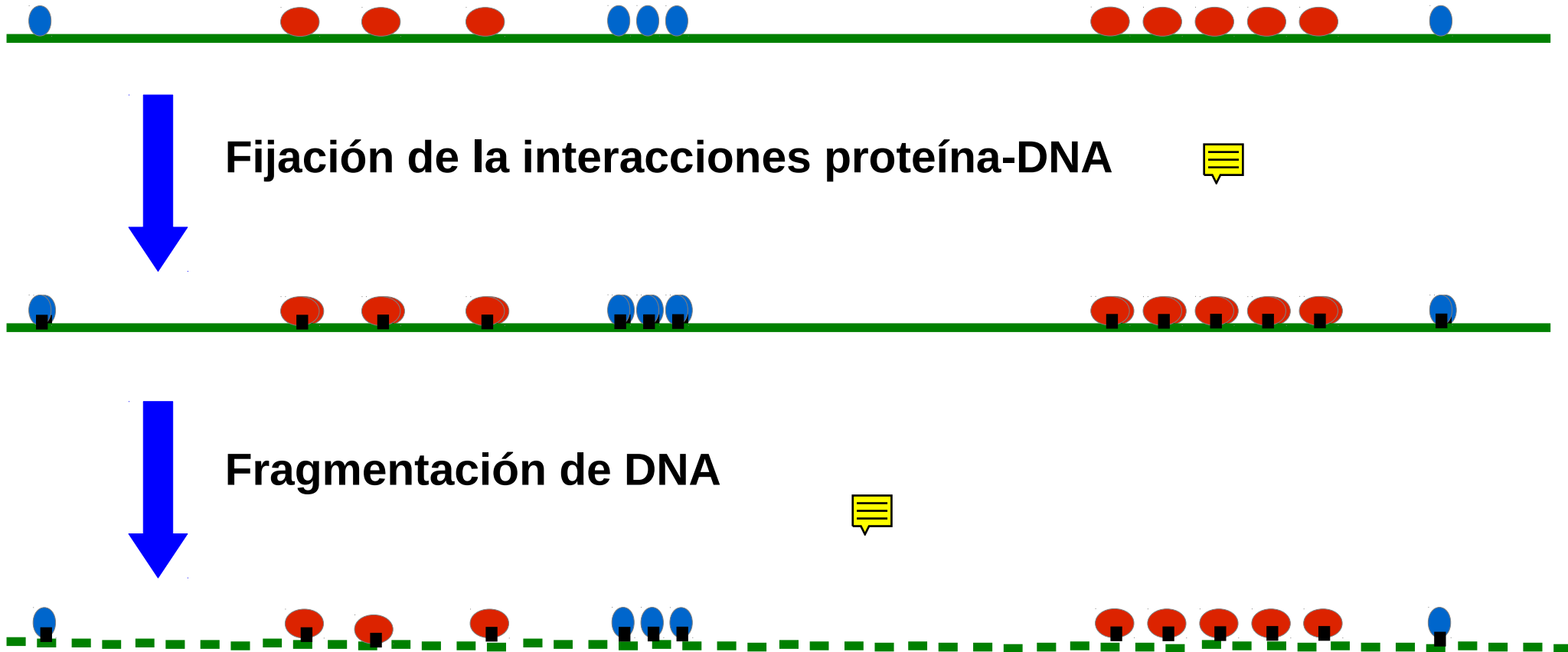


Chromatin Immunoprecipitation coupled with sequencing

Fijación de la interacciones proteína-DNA



Chromatin Immunoprecipitation coupled with sequencing



Chromatin Immunoprecipitation coupled with sequencing



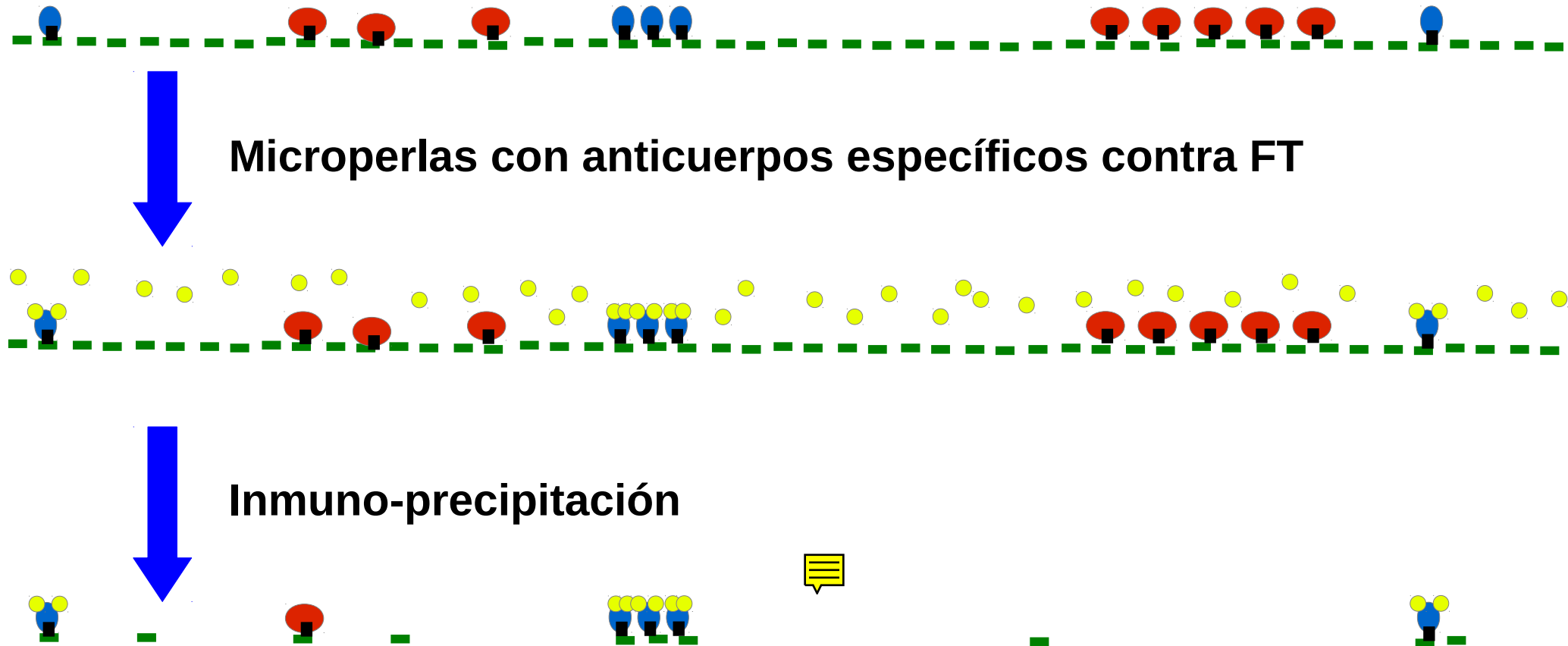
Chromatin Immunoprecipitation coupled with sequencing



Microperlas con anticuerpos específicos contra FT



Chromatin Immunoprecipitation coupled with sequencing



Chromatin Immunoprecipitation coupled with sequencing

Célula 1



Chromatin Immunoprecipitation coupled with sequencing

Célula 1



Célula 2



Chromatin Immunoprecipitation coupled with sequencing

Célula 1



Célula 2



Célula 3



Chromatin Immunoprecipitation coupled with sequencing

Célula 1



Célula 2



Célula 3



Célula 4



Chromatin Immunoprecipitation coupled with sequencing

Célula 1



Célula 2



Célula 3



Célula 4

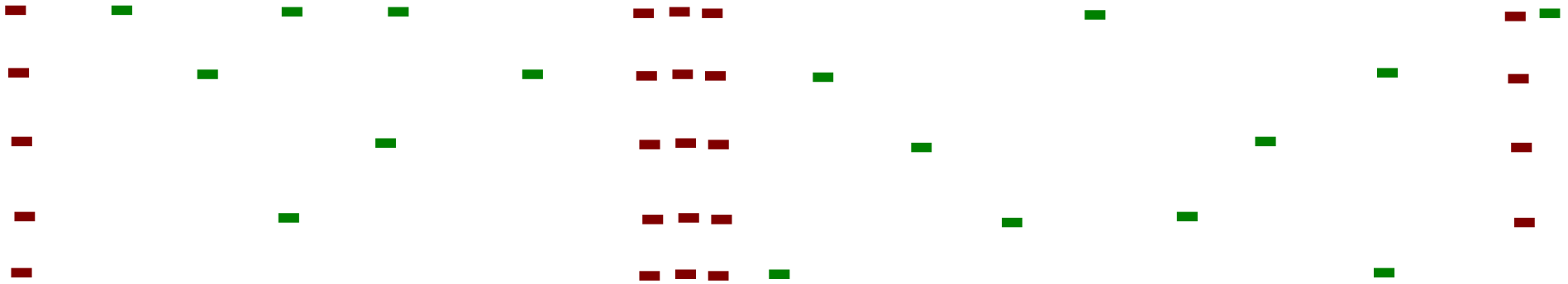


...

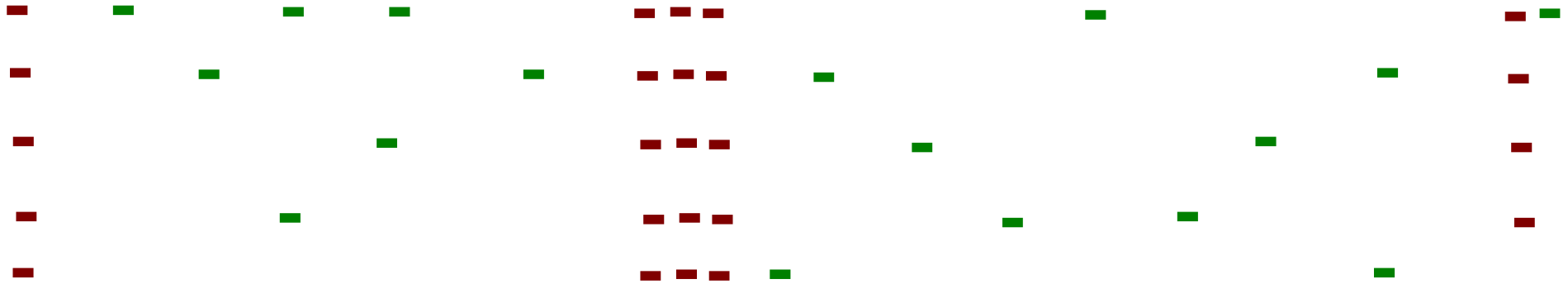
Célula N



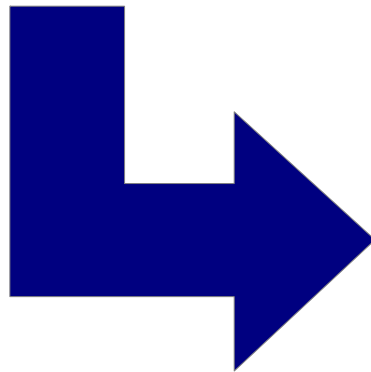
Chromatin Immunoprecipitation coupled with sequencing



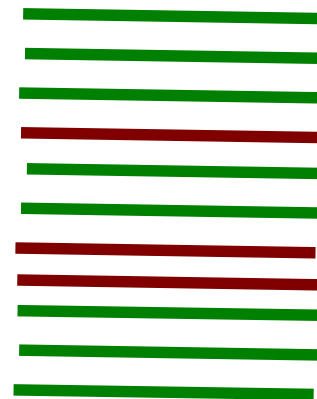
Chromatin Immunoprecipitation coupled with sequencing



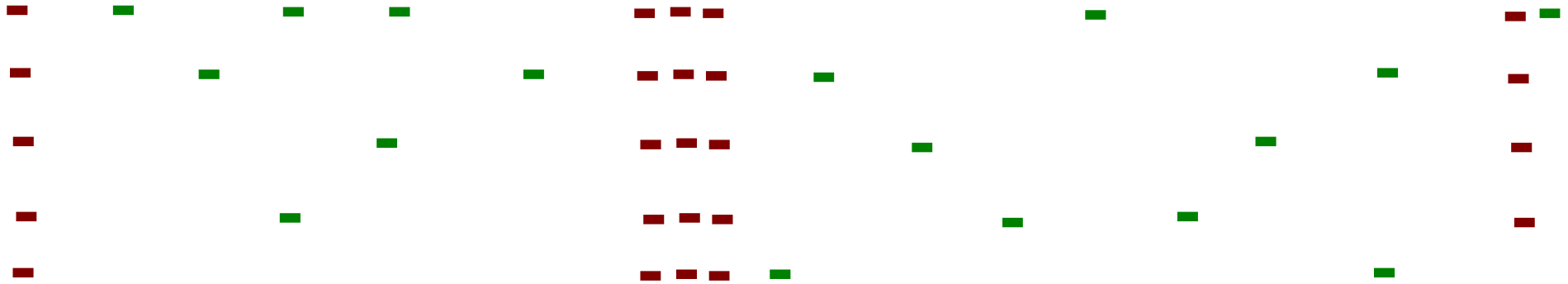
Secuenciación
de altas prestaciones



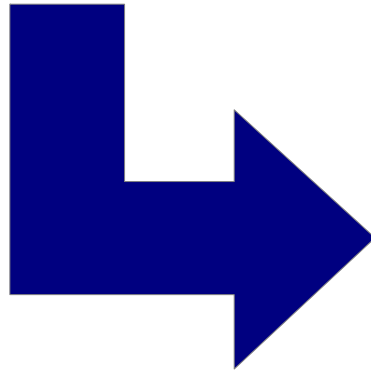
Fichero fastq



Chromatin Immunoprecipitation coupled with sequencing



Secuenciación
de altas prestaciones



Fichero fastq



Dependiendo de la proteína diana la técnica de ChIP-Seq puede usarse para determinar:

- Los sitios de unión y genes diana de un factor de **transcription factor**.
- Los genes activamente transcritos en una condición específica por la **polimerasa**.
- La estructura y modificaciones de la **cromatina**.



Debido a la gran **dificultad de generar anticuerpos** específicos y sensibles para un FT en concreto. La inmuno-precipitación de la cromatina (ChIP) cuando se trabaja con FTs se suele hacer en un genotipo:

- **Mutante** en el FT de interés para eliminar la interferencia con la proteína nativa.
- Transformante con un **minigene del FT** formado:
 - Promotor del propio FT.
 - Secuencia codificante del FT fusionada a una proteína reportadora o etiqueta con anticuerpo ya disponible i.e. GFP

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**



Chromatin Immunoprecipitation coupled with sequencing



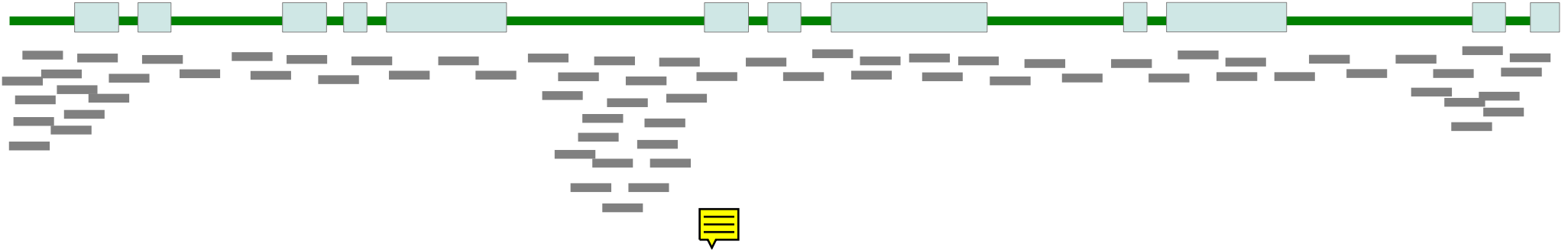
**Fastq
ChIP**



**Mapeo de lecturas
cortas al genoma de
referencia**



Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**



Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**



**Fastq
Control**



Chromatin Immunoprecipitation coupled with sequencing



Fastq ChIP



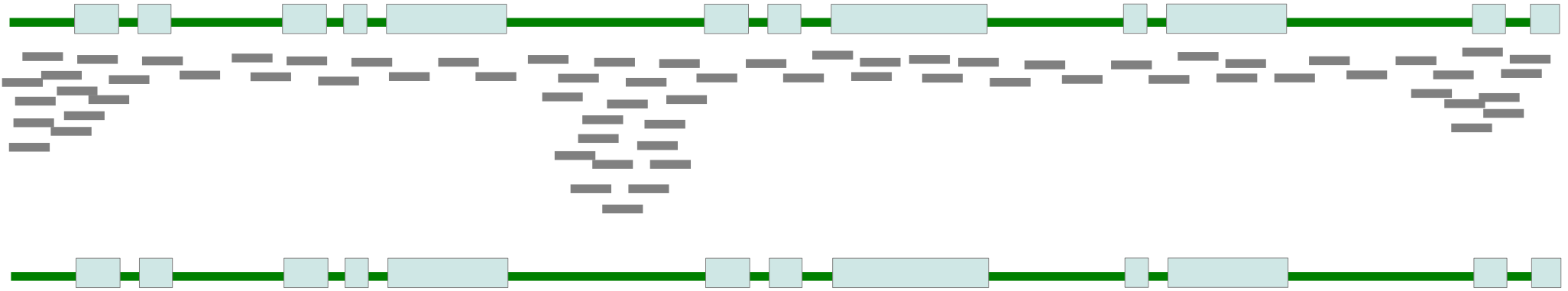
Típicamente se considera uno de los dos siguientes tipos de control:

- **Input:** Extracción de genómico para obtener la distribución esperada del fondo.
- **Mock:** Se sigue exactamente el mismo protocolo que con el ChIP pero no se añade anticuerpo. Se obtendrá la precipitación inespecífica del fondo.

Fastq Control



Chromatin Immunoprecipitation coupled with sequencing



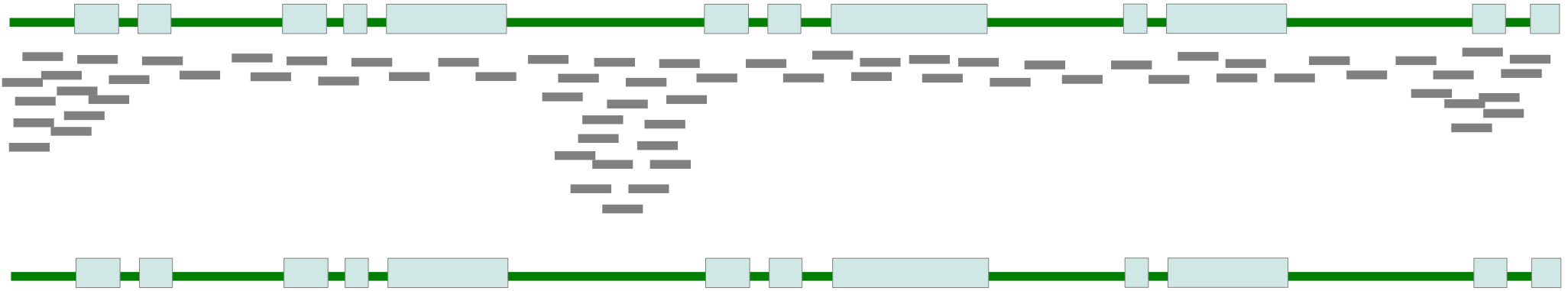
**Fastq
ChIP**



**Fastq
Control**



Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

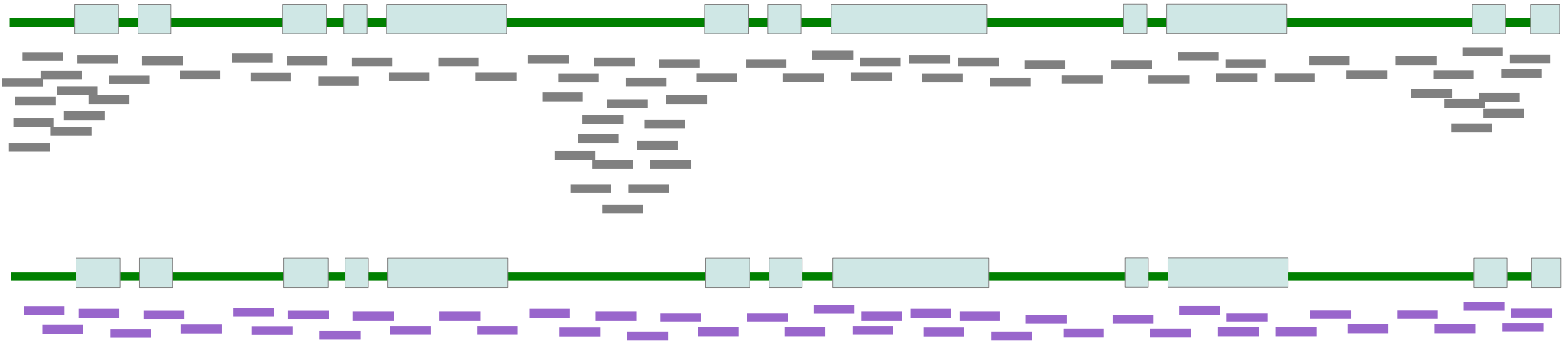


**Mapeo de lecturas
cortas al genoma de
referencia**

**Fastq
Control**



Chromatin Immunoprecipitation coupled with sequencing



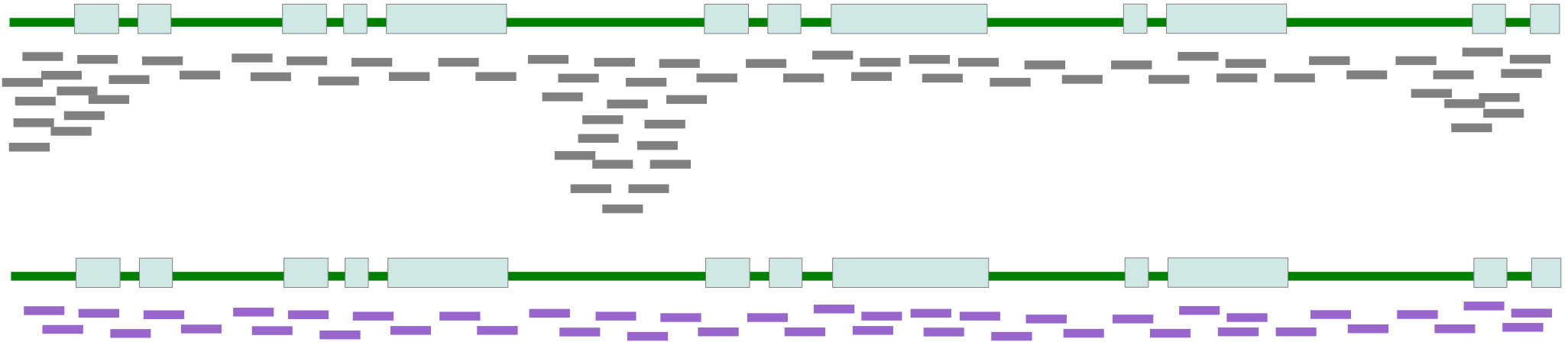
**Fastq
ChIP**



**Fastq
Control**



Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

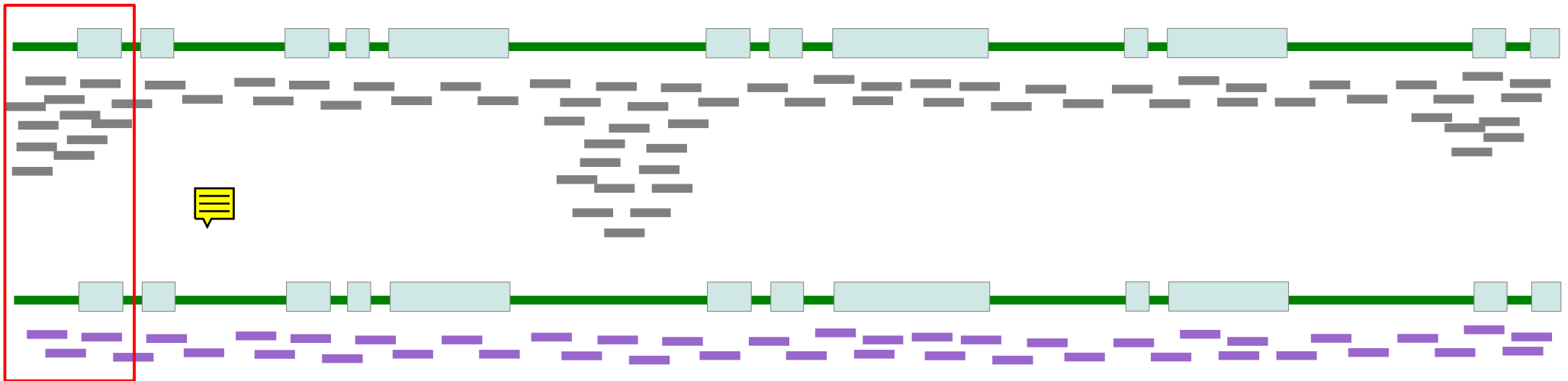


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

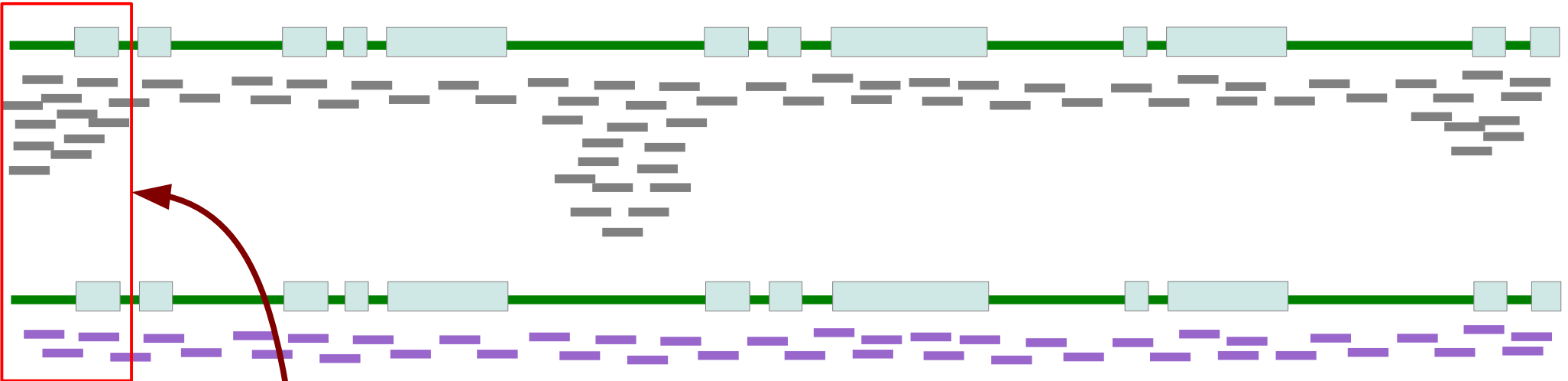


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**



Contraste de hipótesis basado en una **distribución de Poisson** produce un:

- fold-change
- p-valor

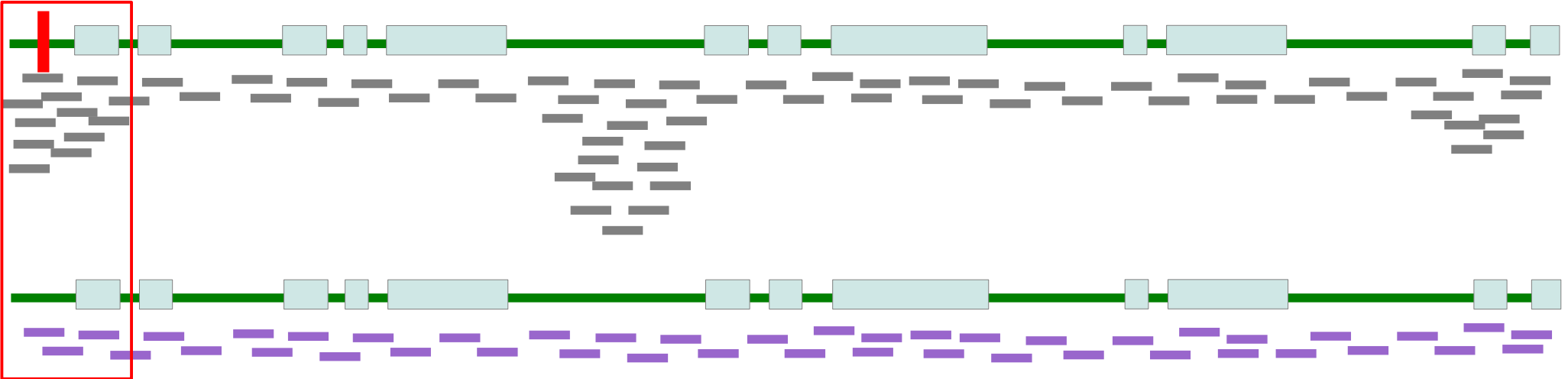
Cuando el fold-change es lo suficientemente alto y el p-valor lo suficientemente bajo se asume que existe un pico y por lo tanto existe evidencia sobre que nuestro FT se una a dicho lugar del DNA.

Determinación de picos o Peak Calling

**Fastq
Control**



Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

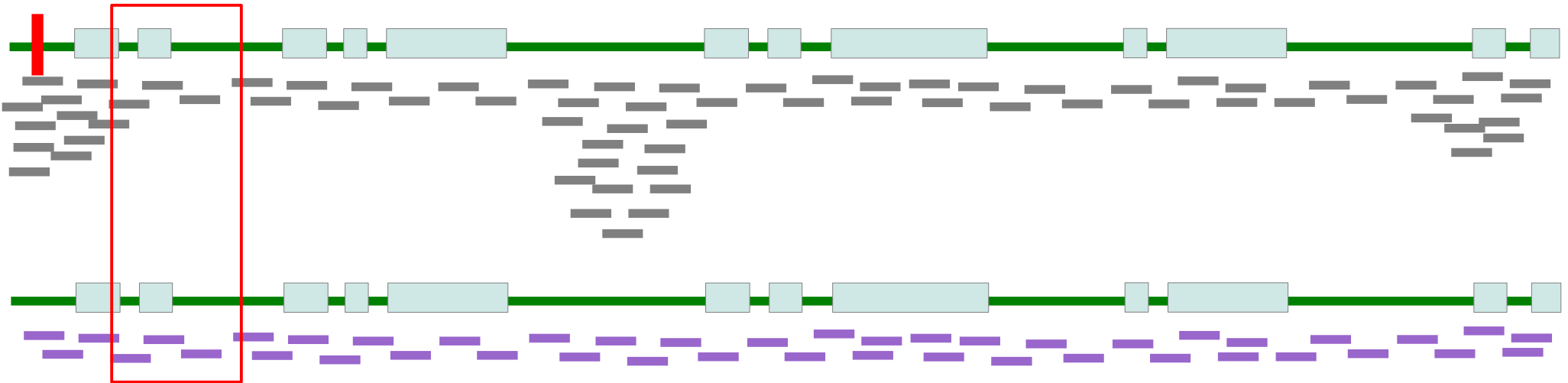


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

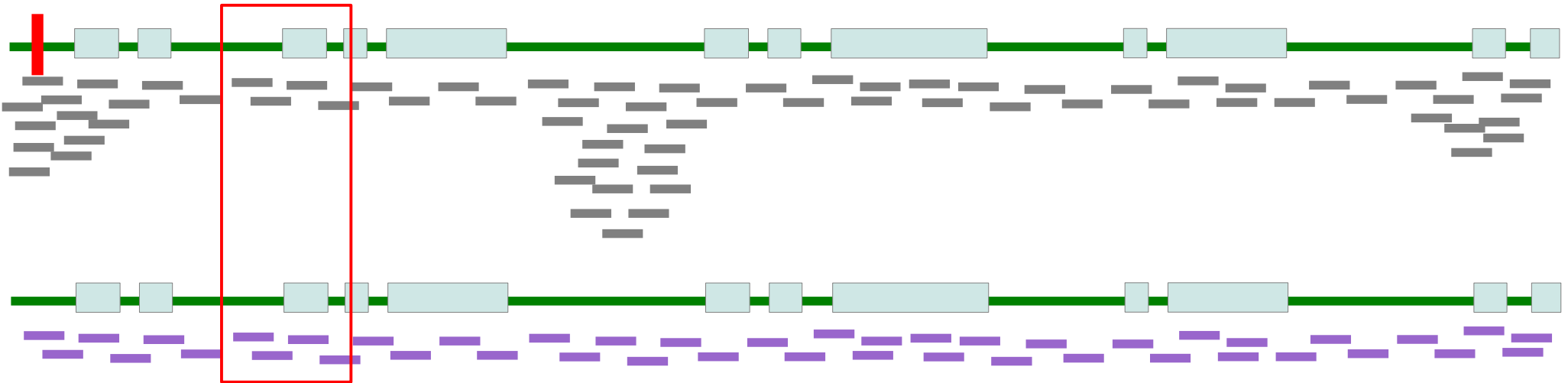


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

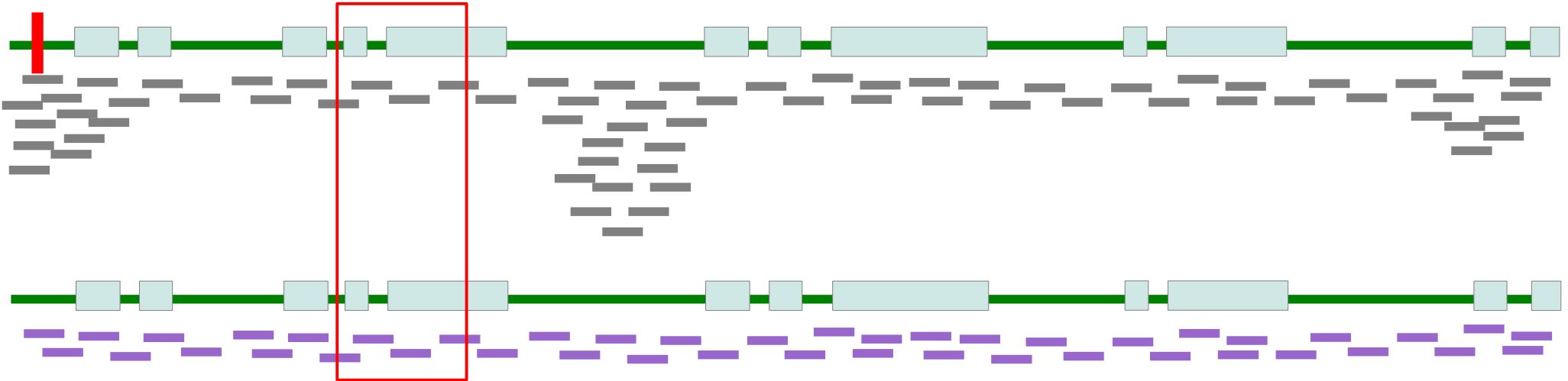


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

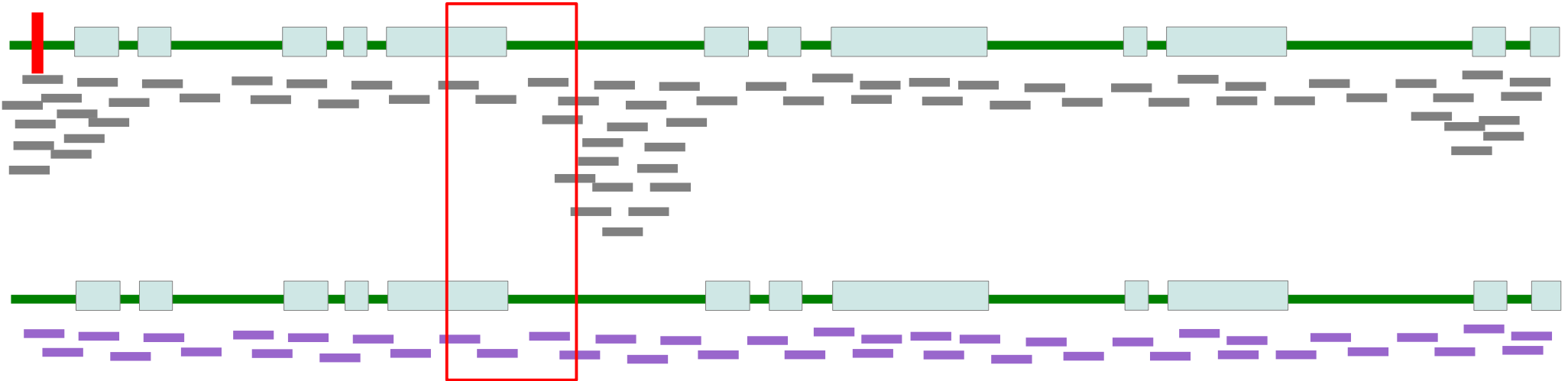


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

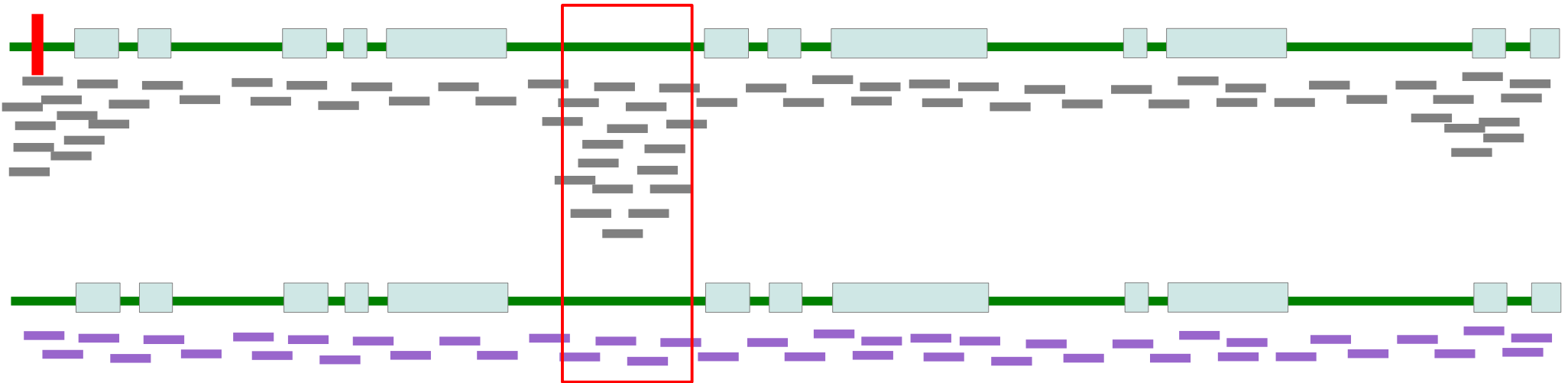


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

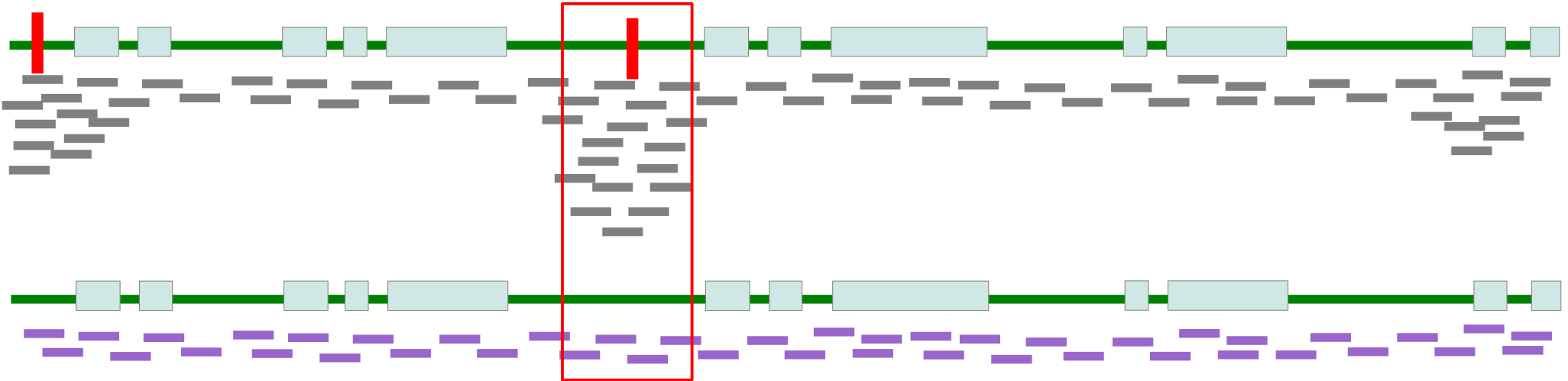


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

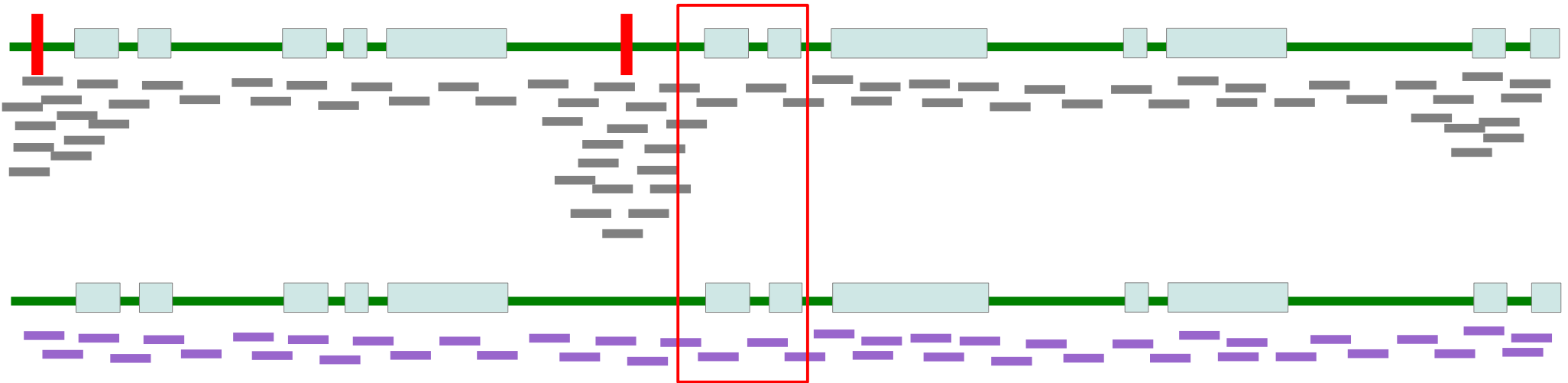


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

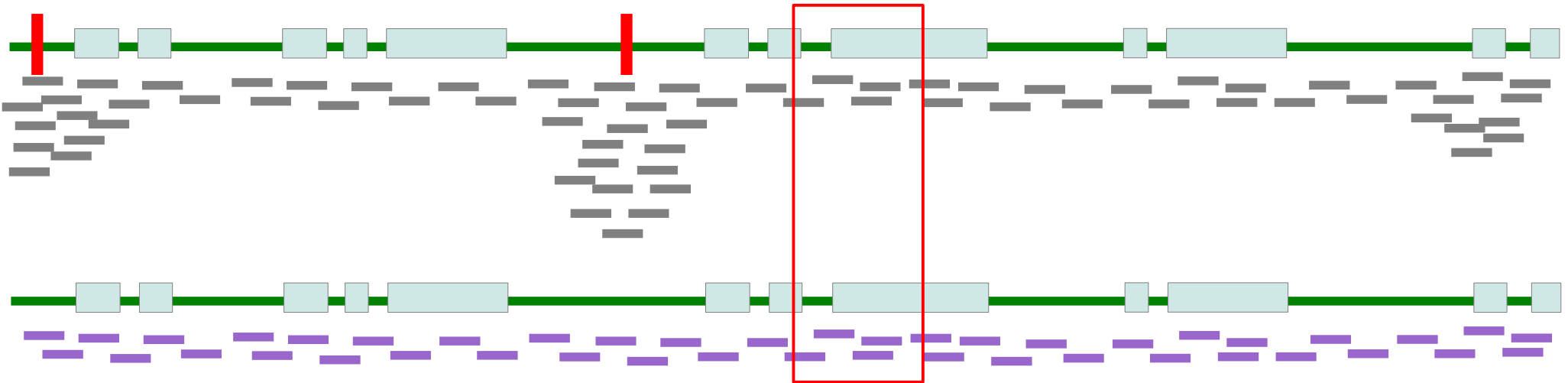


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

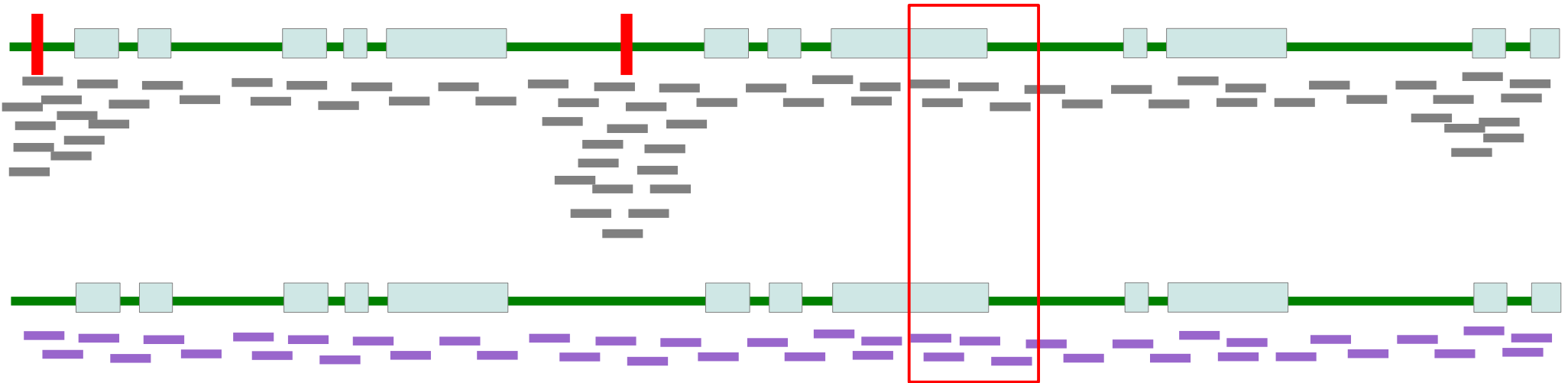


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

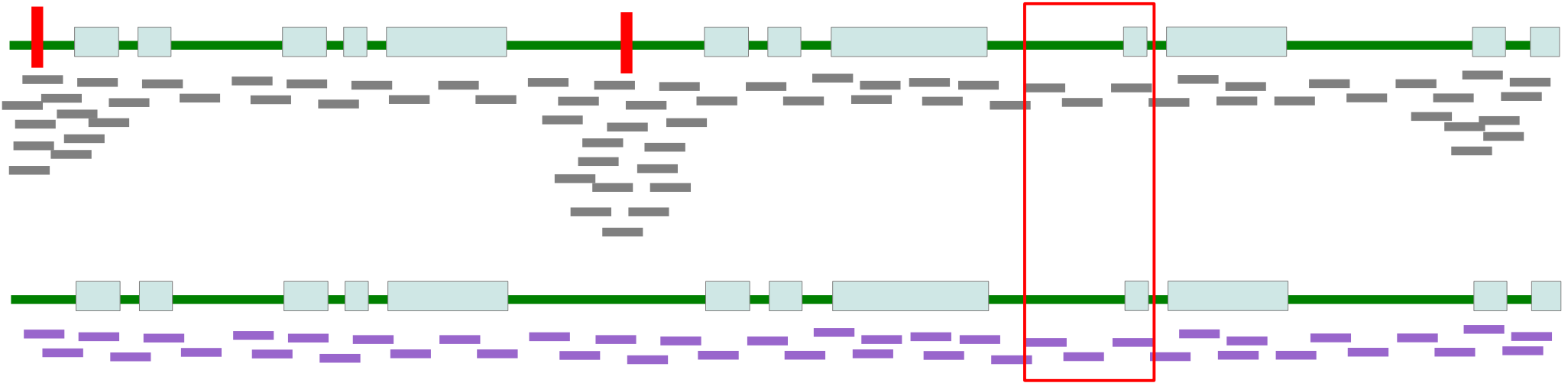


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

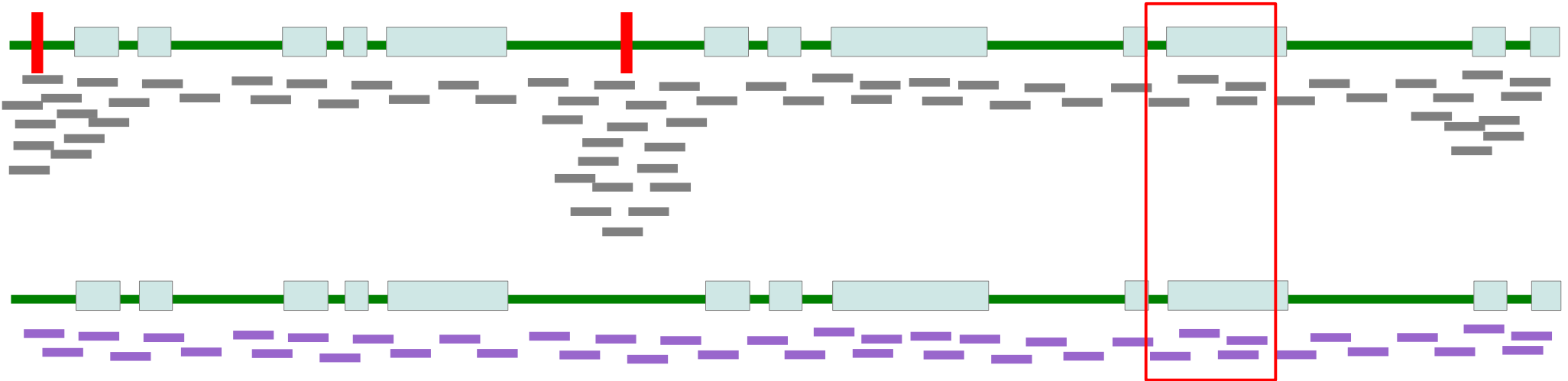


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

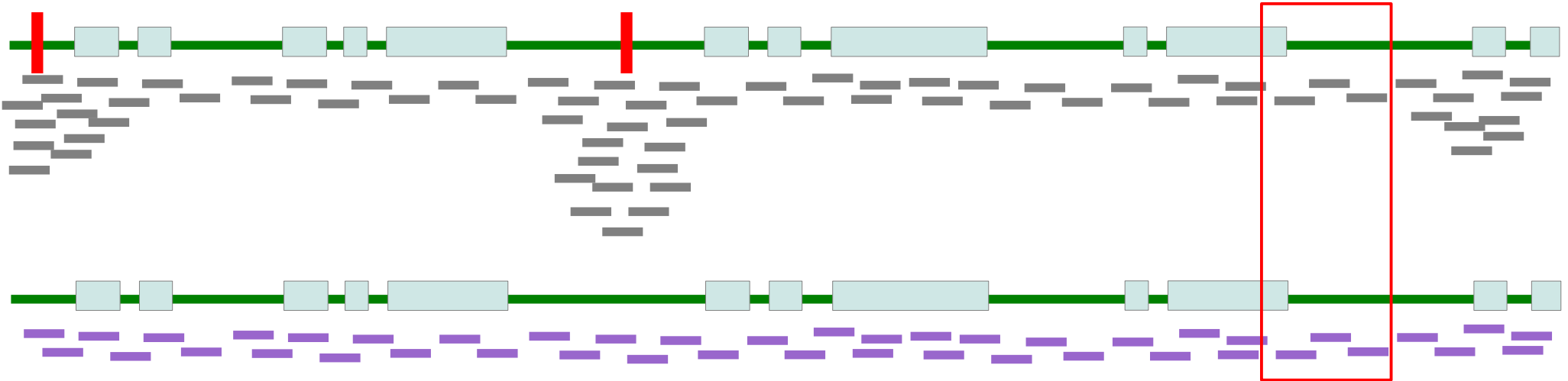


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

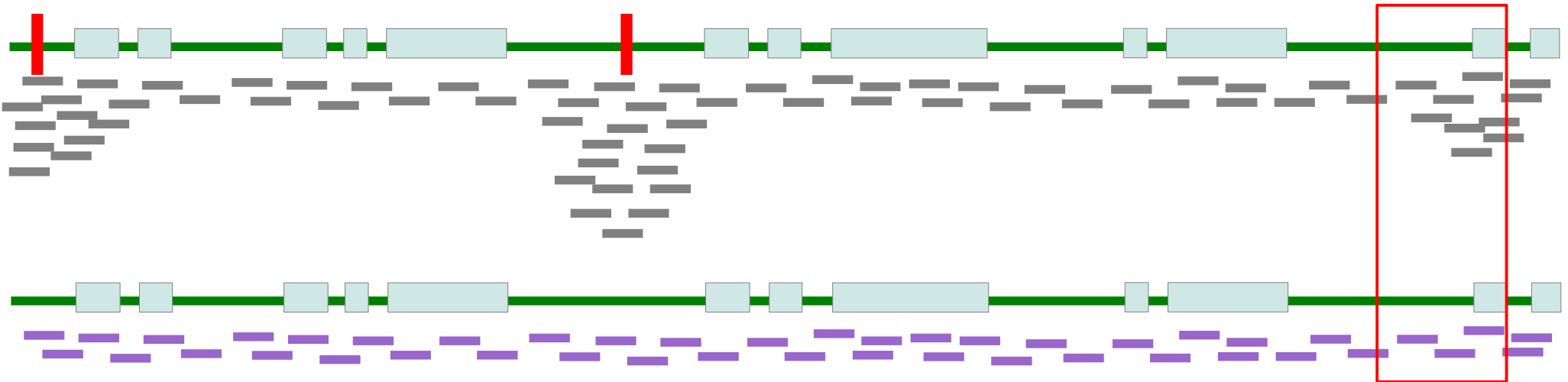


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

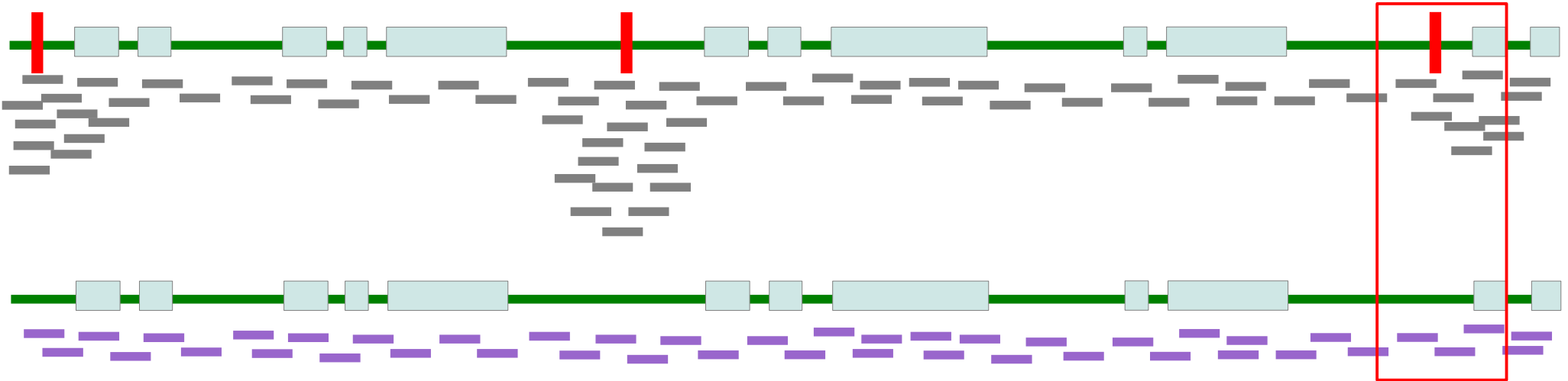


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

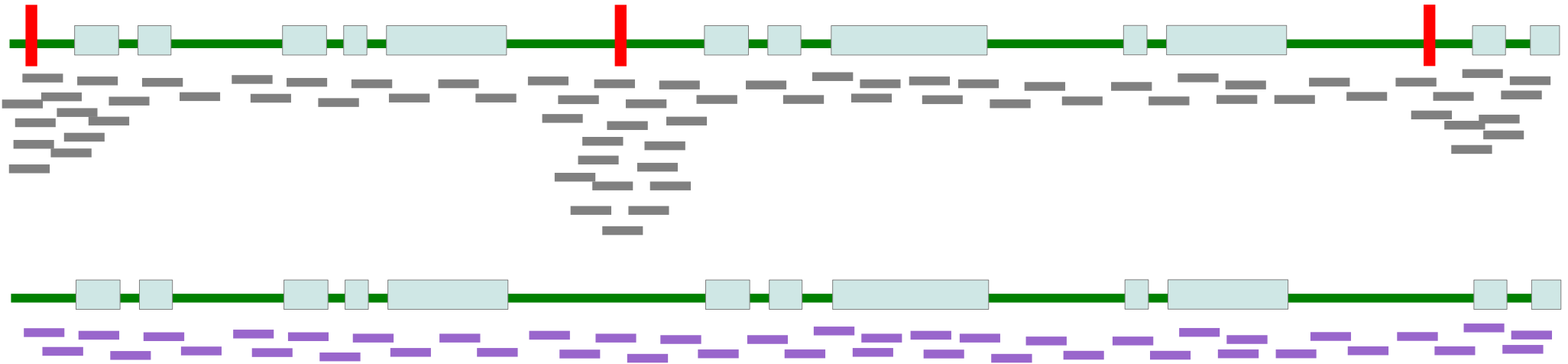


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

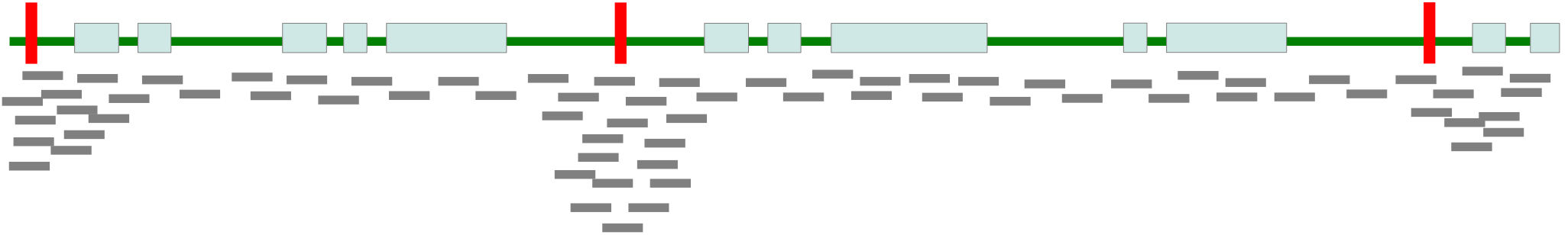


**Fastq
Control**



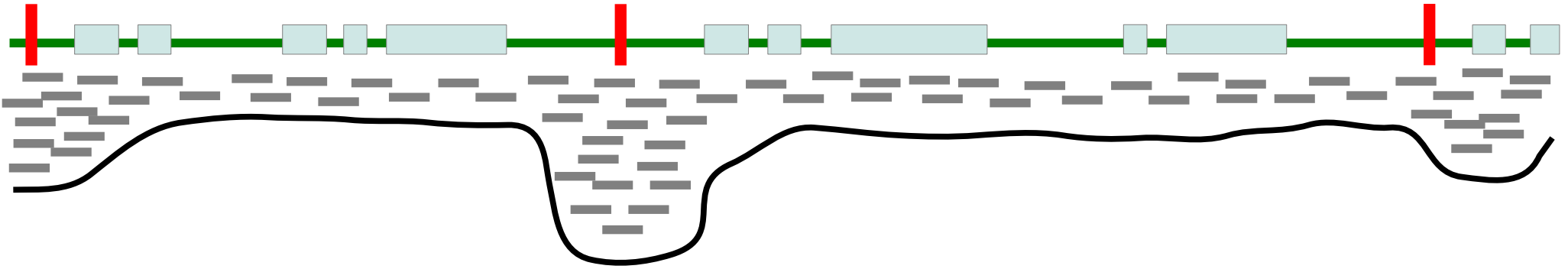
Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



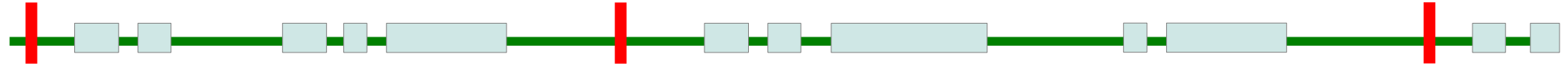
Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



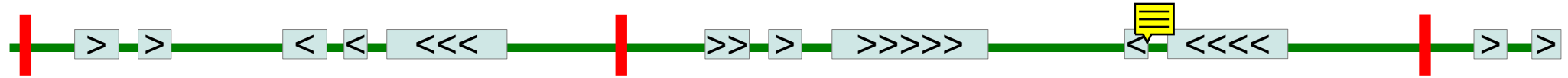
Determinación de picos o Peak Calling

ChIP-seq: Anotación de Picos o Determinación de dianas



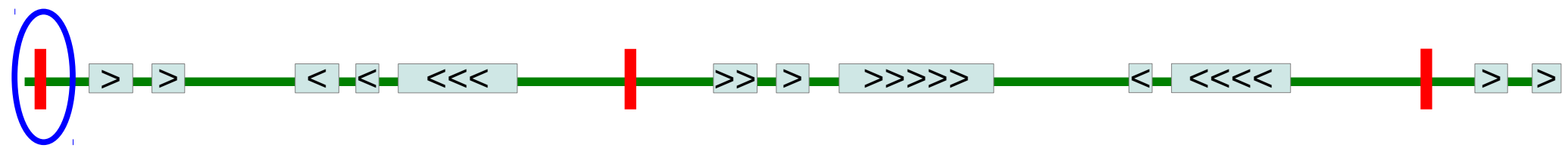
El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

ChIP-seq: Anotación de Picos o Determinación de dianas



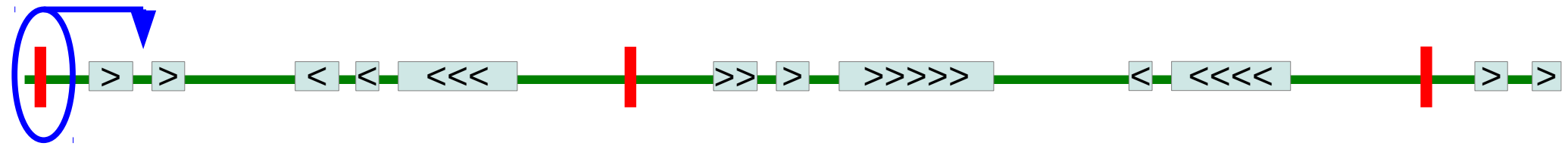
El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

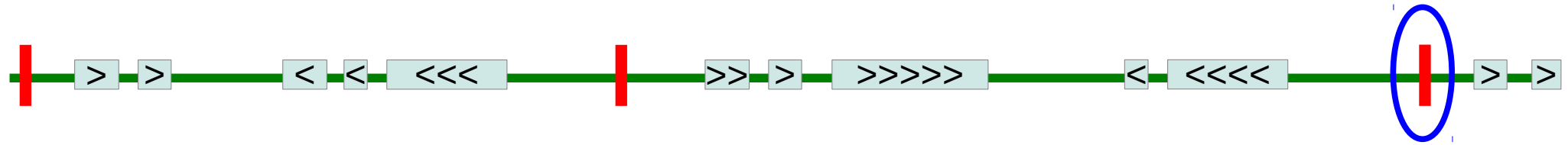
ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

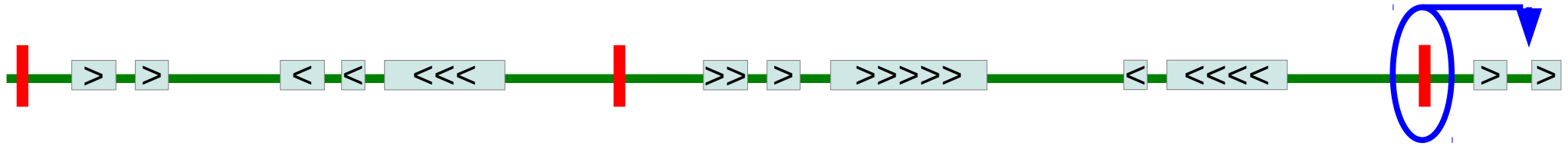
ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

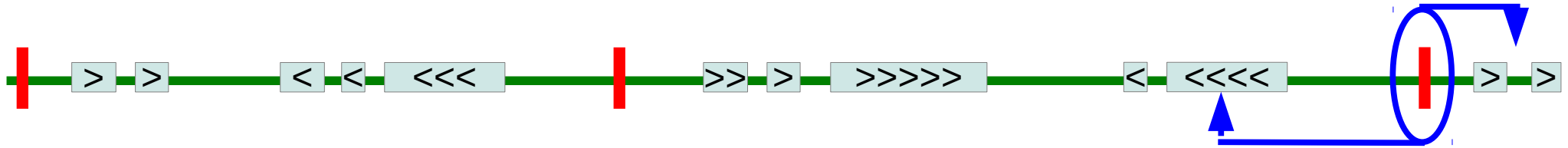
ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

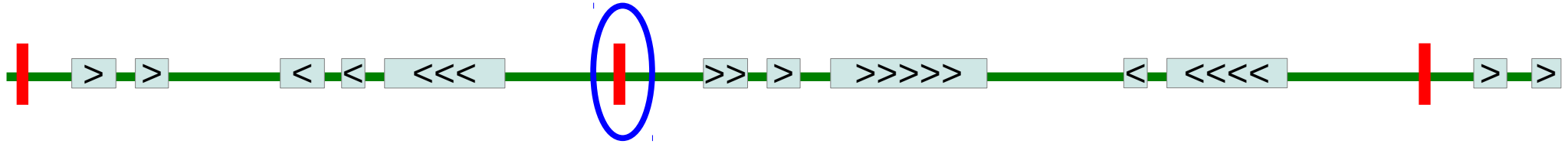
ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

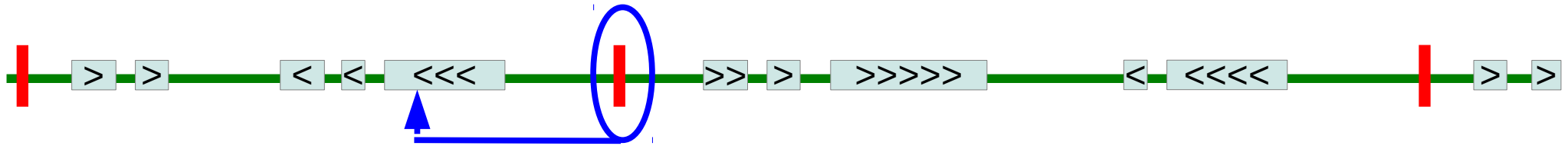
ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

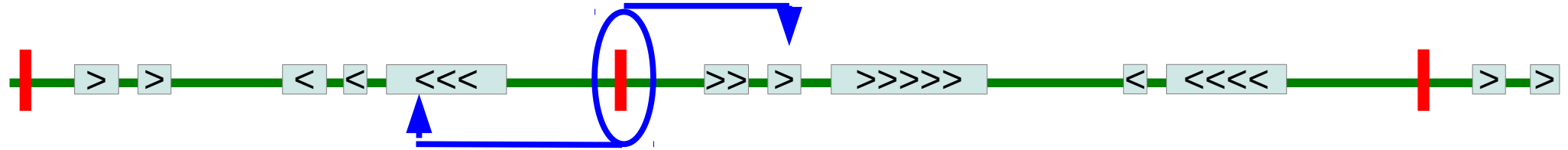
ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

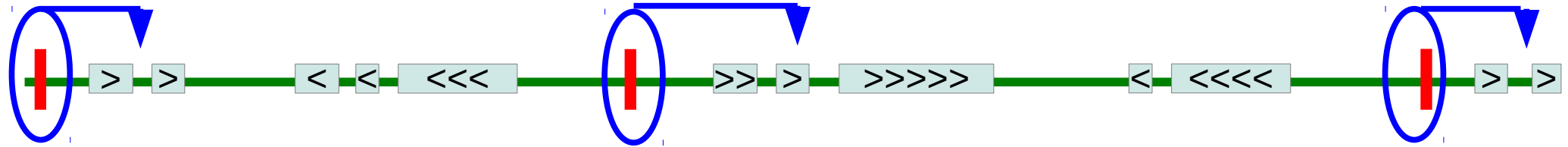
ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

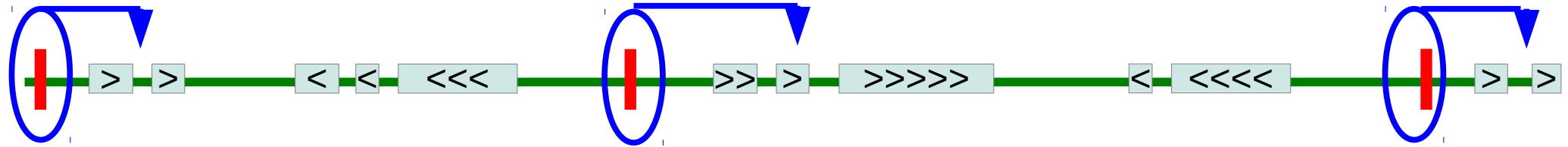
ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

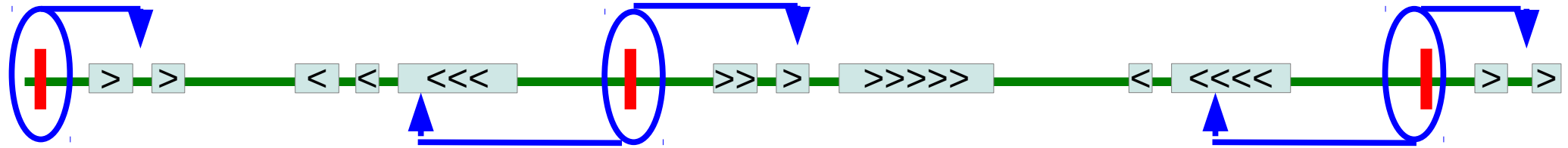
ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

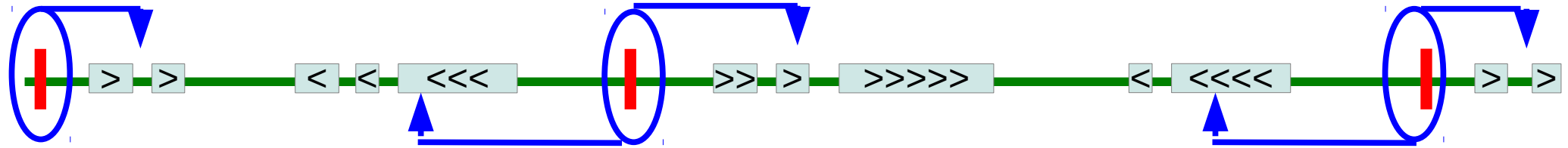
ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

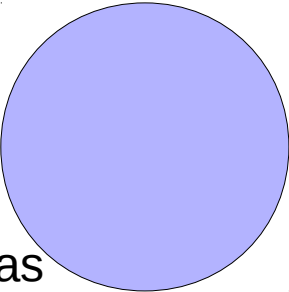
El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

ChIP-seq: Anotación de Picos o Determinación de dianas



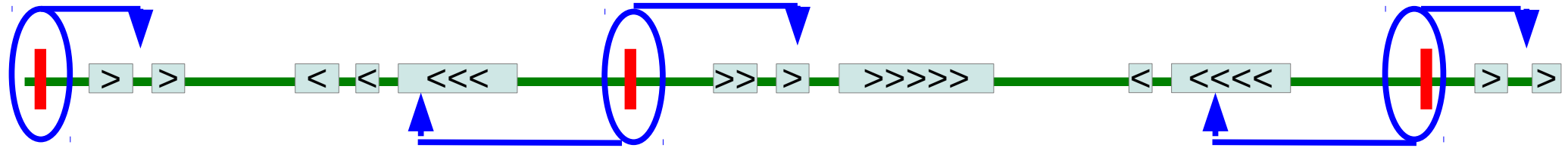
El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.



Genes aguas
abajo de un pico

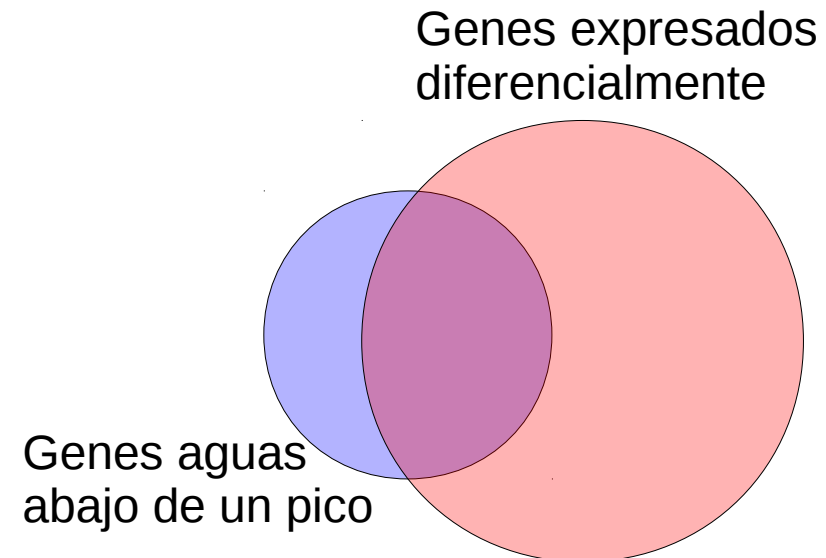
ChIP-seq: Anotación de Picos o Determinación de dianas



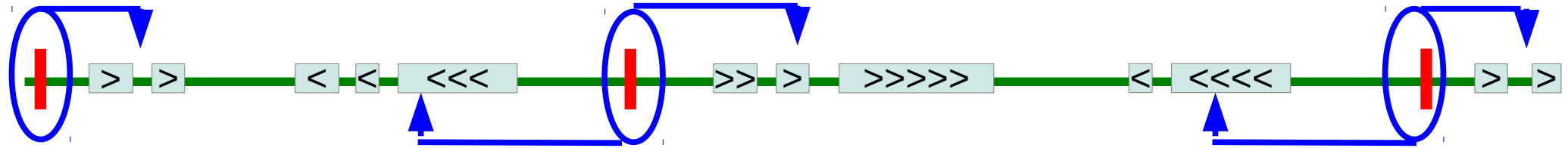
El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

Un método con mayor fiabilidad combina **datos de expresión génica diferencial** (microarrays o RNA-seq) con datos de ChIP-seq.



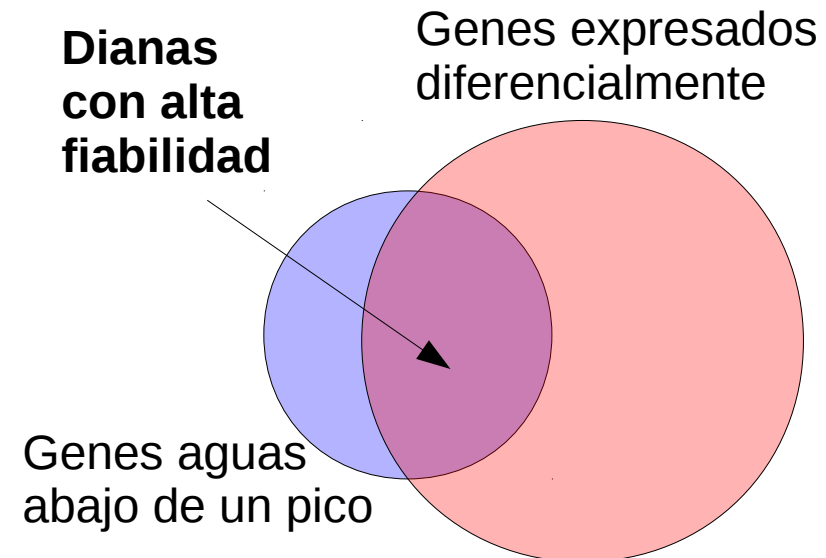
ChIP-seq: Anotación de Picos o Determinación de dianas



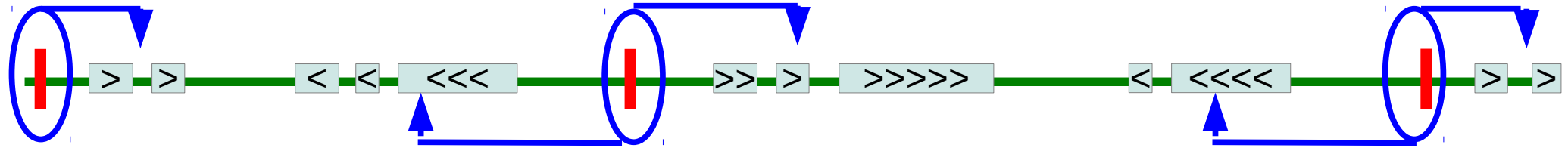
El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

Un método con mayor fiabilidad combina **datos de expresión génica diferencial** (microarrays o RNA-seq) con datos de ChIP-seq.



ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

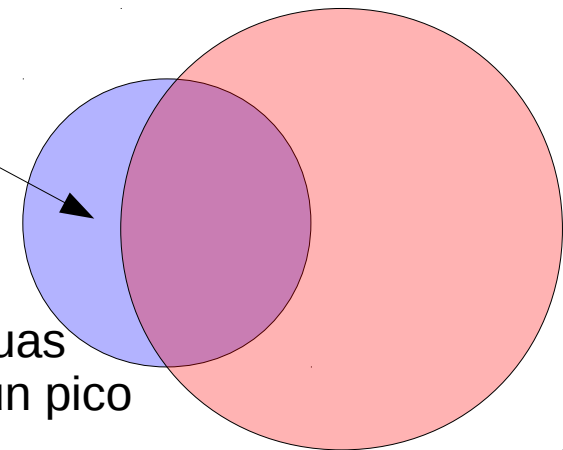
El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

Un método con mayor fiabilidad combina **datos de expresión génica diferencial** (microarrays o RNA-seq) con datos de ChIP-seq.

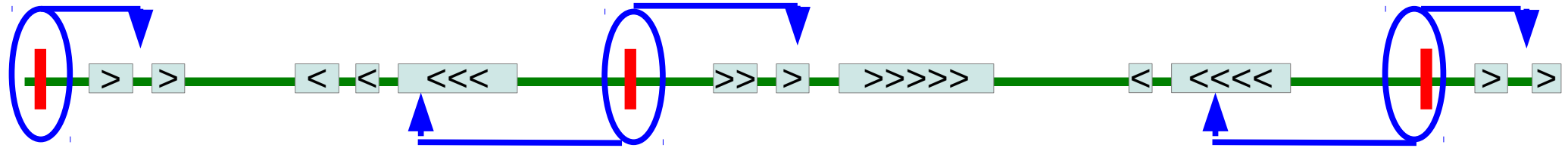
Falsos positivos o dianas que necesitan otros factores para su expresión

Genes expresados diferencialmente

Genes aguas abajo de un pico



ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

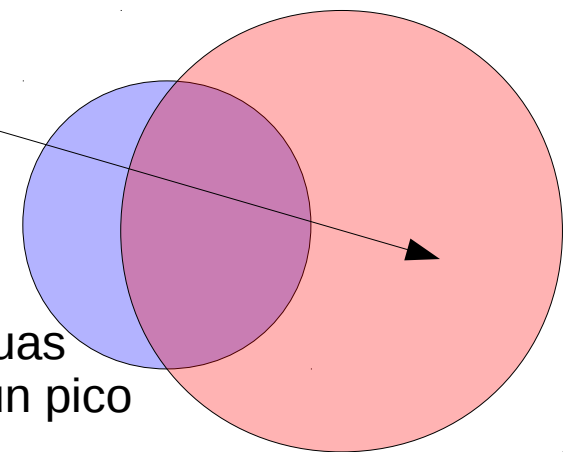
El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

Un método con mayor fiabilidad combina **datos de expresión génica diferencial** (microarrays o RNA-seq) con datos de ChIP-seq.

Dianas indirectas o efectos secundarios

Genes expresados diferencialmente

Genes aguas abajo de un pico



ChIP

Input

Ensembl

File sra
Sample 1

File sra
Sample 2

File sra
Sample 3

File sra
Sample 4

Fasta File
Reference
Genome

File fastq
sample 1

File fastq
sample 2

File fastq
sample 3

File fastq
sample 4

BWT Index
Reference
Genome

SAM File
ChIP

SAM File
Input

BED File
TF Peaks

TF Peaks
Visualization

TF Binding
Sequence

Nearest Downstream
Gene Identification

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**



**Mapeo de lecturas
cortas al genoma de
referencia**



ChIP

Ensembl

File sra
Sample 1

File sra
Sample 2

File fastq
sample 1

File fastq
sample 2

Fasta File
Reference
Genome

BWT Index
Reference
Genome

ChIP

File sra
Sample 1

File sra
Sample 2

wget

File fastq
sample 1

File fastq
sample 2

fastq-dump

Ensembl

Fasta File
Reference
Genome

BWT Index
Reference
Genome

ChIP

Ensembl



```
cd /home/<usuario>/<exp>/genome
bowtie-build <fichero.fa> <fichero>
```

Preparación del espacio de trabajo.

- Formato fasta del genoma de referencia:

>5

```
GGGAACCAGCTACTAGATGGTTCGATTAGTCTTTTCGCCCCCTATACCCA  
AGTCTGAAAAGCGATTTGCACGTCAGCACATCTACGAGCCTCCACCA  
GAGTTTCCTCTGGCTTCACCCTGCTCAGGCATAGTTTACCATCTTTTCG  
GGTCCCAACAGGTATGCTCGCACTCAAACCTTTTCGTAGAAACAACATG
```

```
gunzip chr5.fa.gz  
bowtie-build chr5.fa chr5  
cd ../annotation  
gunzip chr5.gtf.gz
```

~ 1 min

- Formato gtf de la anotación:

```
chromosome_1 phytozomev10 exon 20760 21065 . + .  
ID=Cre01.g000033.t1.1.v5.5.exon.2;Parent=Cre01.g000033.t1.1.v5.5;paci  
d=30788883
```

gen 1

gen 2



ChIP

Ensembl



```
cd /home/<usuario>/<exp>/samples/chip/
```

```
bowtie -S -v 2 --best --strata -m 1 (ruta_indice_genoma_ref) (fichero).fastq > chip.sam
```

Ruta indice de referenica: /home/<usuario>/<exp>/genome/<fichero>

Si hay réplicas los correspondiente ficheros fastq se separan por comas.

ChIP

Ensembl



Alineamientos válidos los mejores con
a lo sumo dos errores

salida en formato SAM

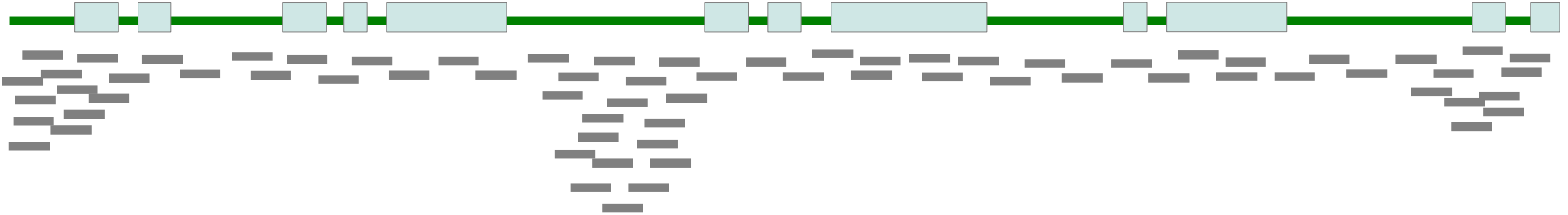
```
cd /home/<usuario>/<exp>/samples/chip/
```

```
bowtie -S -v 2 --best --strata -m 1 (ruta_indice_genoma_ref) (fichero).fastq > chip.sam
```

Ruta indice de referenica: /home/<usuario>/<exp>/genome/<fichero>

Si hay réplicas los correspondiente ficheros fastq se separan por comas.

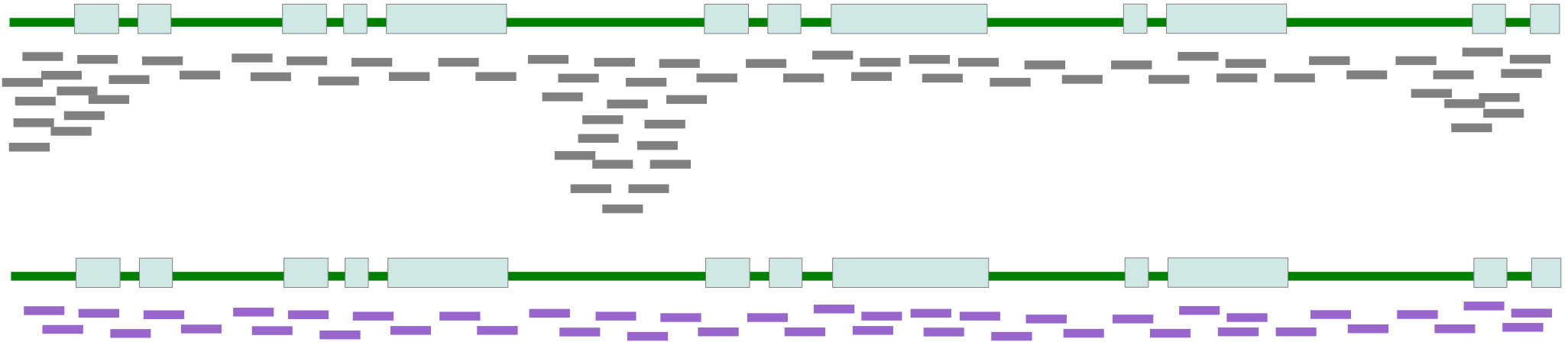
Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**



Chromatin Immunoprecipitation coupled with sequencing



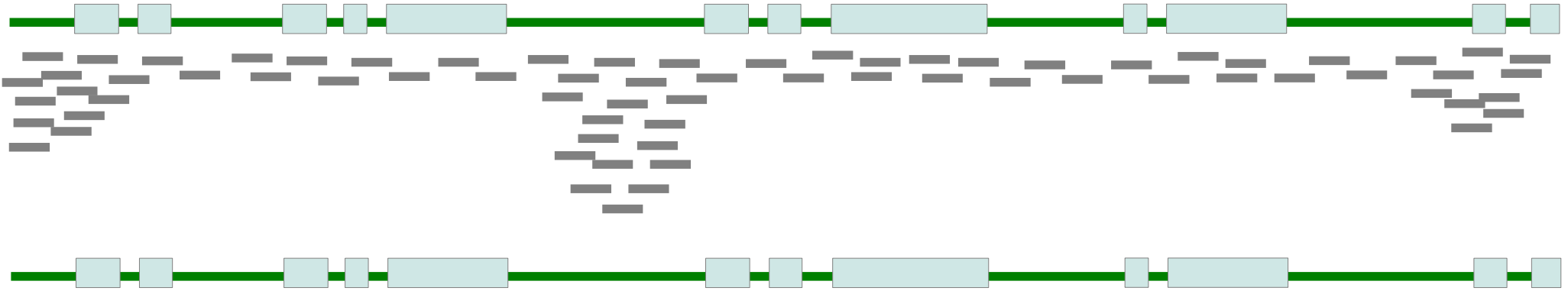
Acceder a la carpeta samples/h2aub_col0
Descomprimir el fichero fastq
Mapear lecturas al genoma de referencia con bowtie

```
gunzip col_0_h2aub_7dag_1_chr5.fastq.gz
```

```
bowtie -S -v 2 --best --strata -m 1 $HOME/myriam_chip_seq/genome/chr5  
col_0_h2aub_7dag_1_chr5.fastq > col_0_h2aub_7dag_1_chr5.sam
```

~ 3 min

Chromatin Immunoprecipitation coupled with sequencing



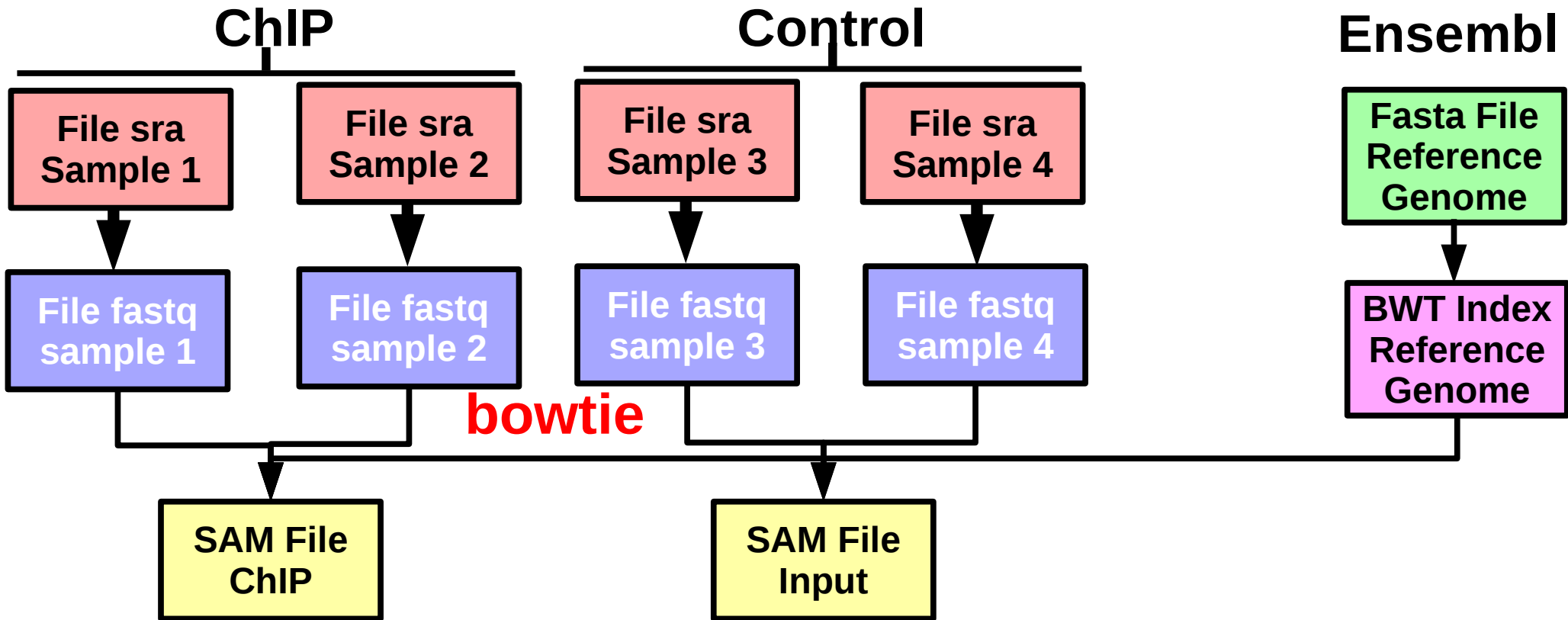
**Fastq
ChIP**



**Mapeo de lecturas
cortas al genoma de
referencia**

**Fastq
Control**





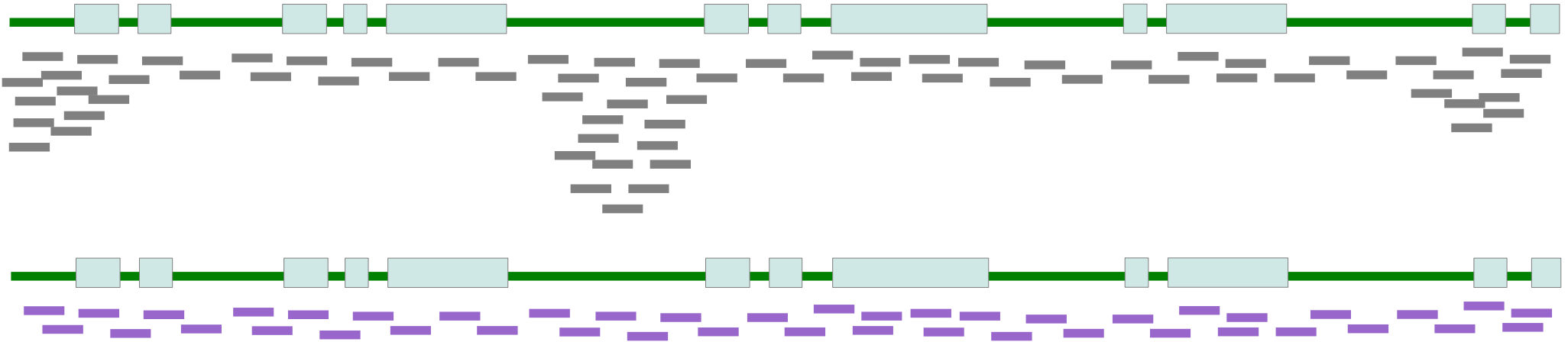
```
cd /home/<usuario>/<exp>/samples/control/
```

```
bowtie -S -v 2 --best --strata -m 1 (ruta_indice_genoma_ref) (fichero).fastq > control.sam
```

Ruta indice de referenica: /home/<usuario>/<exp>/genome/<fichero>

Si hay réplicas los correspondiente ficheros fastq se separan por comas.

Chromatin Immunoprecipitation coupled with sequencing



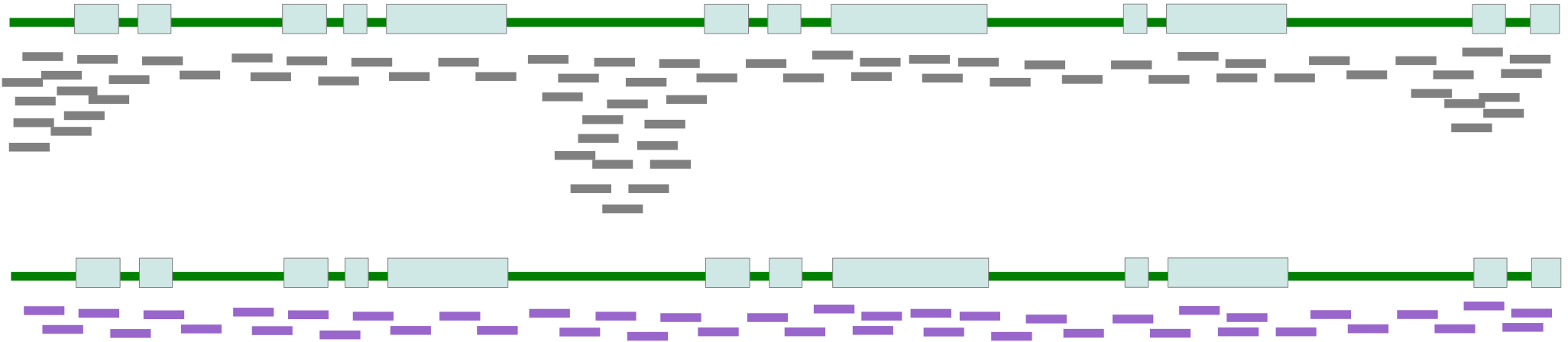
**Fastq
ChIP**



**Fastq
Control**



Chromatin Immunoprecipitation coupled with sequencing



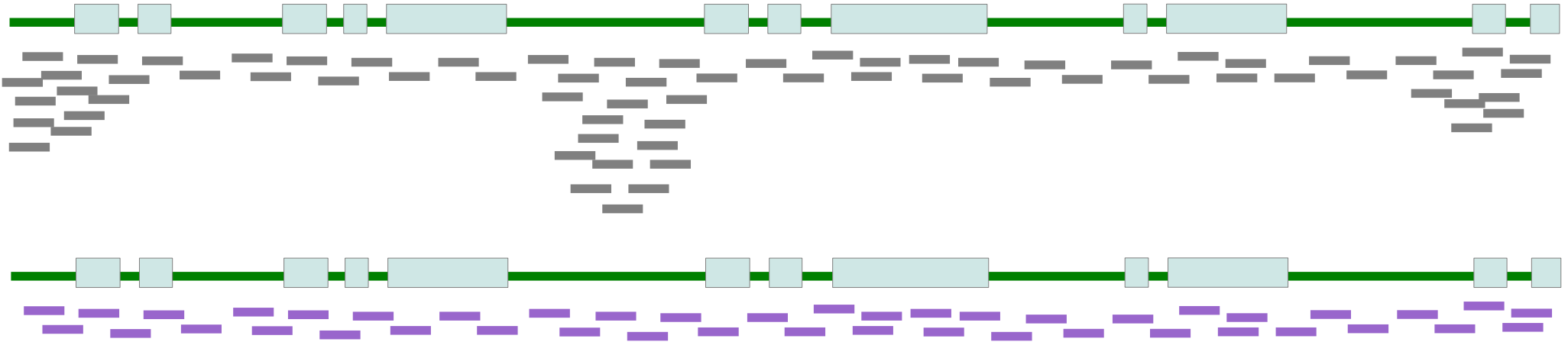
Acceder a la carpeta samples/input
Descomprimir el fichero fastq
Mapear lecturas al genoma de referencia con bowtie

```
gunzip input_chr5.fastq.gz
```

```
bowtie -S -v 2 --best --strata -m 1 $HOME/myriam_chip_seq/genome/chr5  
input_chr5.fastq > input_chr5.sam
```

~ 5 min

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

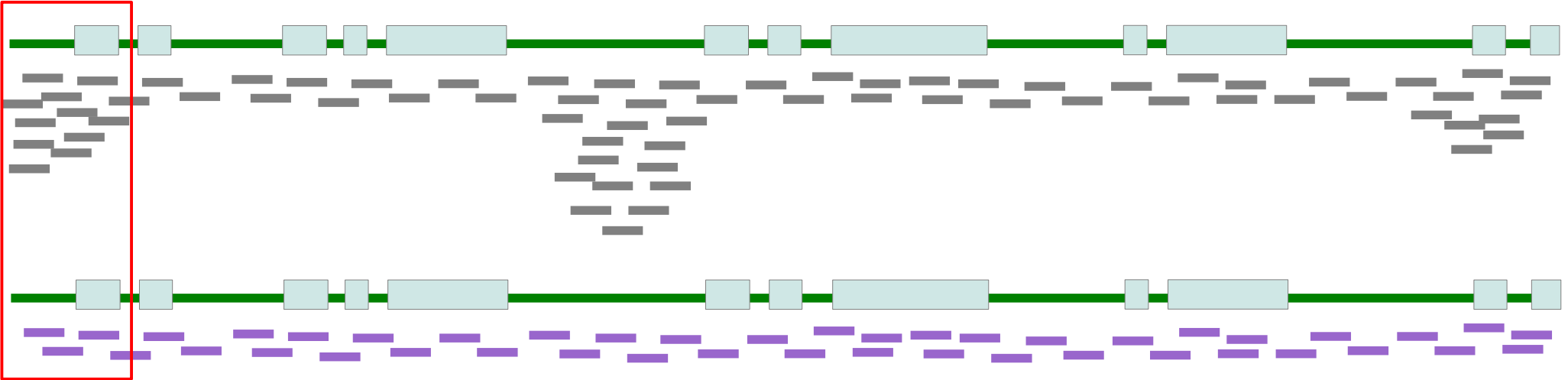


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

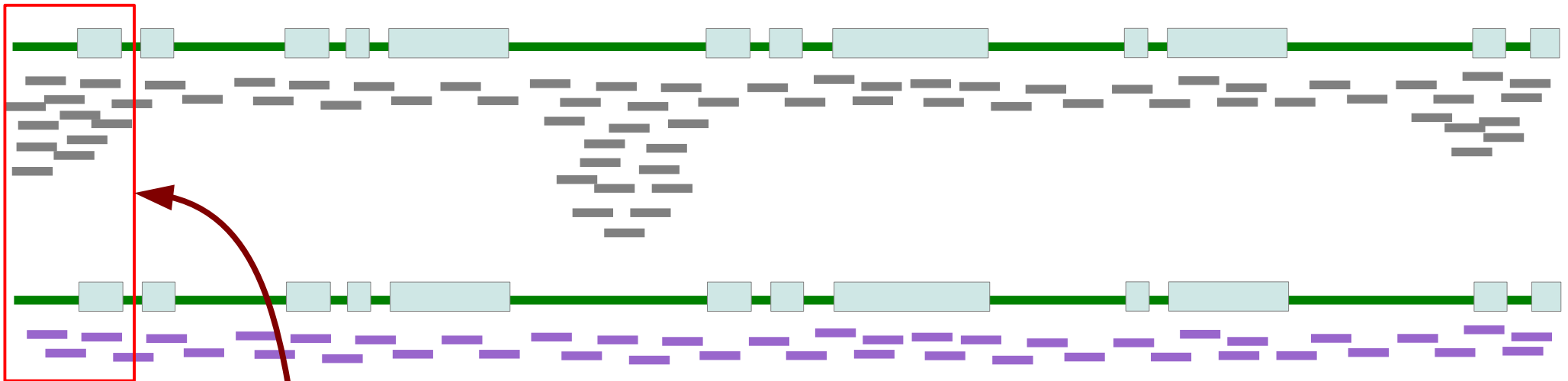


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**



Contraste de hipótesis basado en una **distribución de Poisson** produce un:

- fold-change
- p-valor

Cuando el fold-change es lo suficientemente alto y el p-valor lo suficientemente bajo se asume que existe un pico y por lo tanto existe evidencia sobre que nuestro FT se une a dicho lugar del DNA.

Determinación de picos o Peak Calling

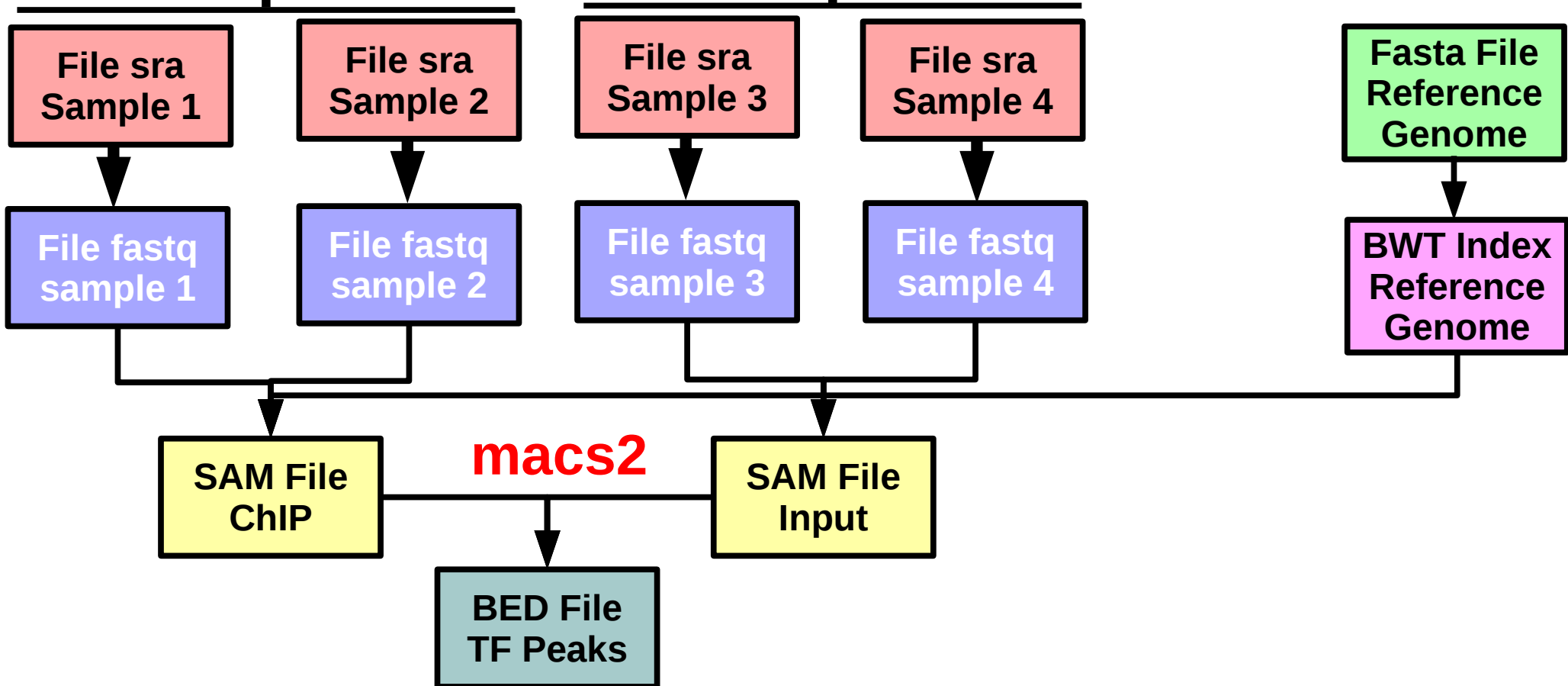
**Fastq
Control**



ChIP

Control

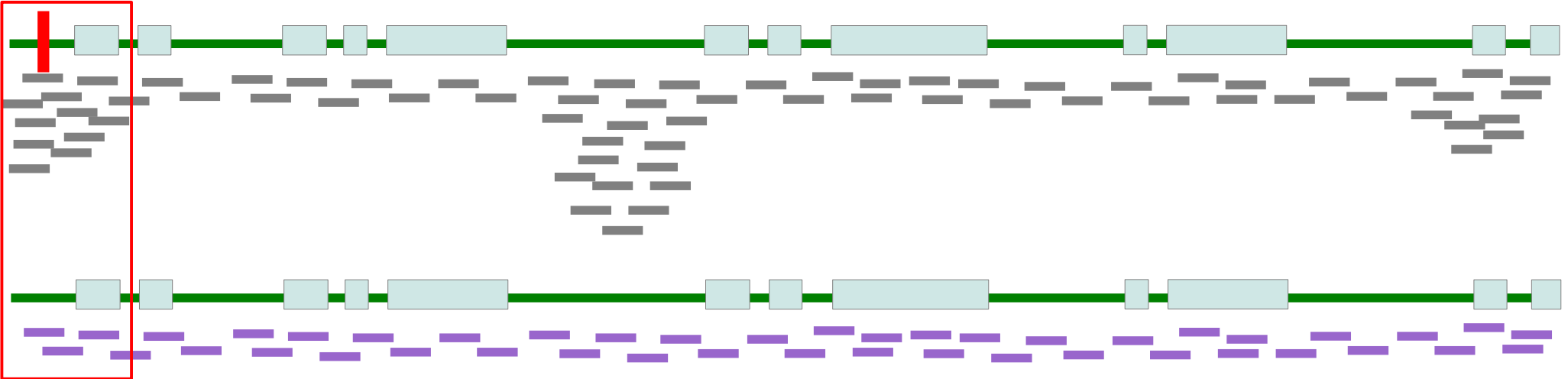
Ensembl



`cd /home/<usuario>/<exp>/results/`

`macs2 callpeak -t <ruta_chip.sam> -c <ruta_control.sam> -f SAM -g <long_genoma>
--outdir <carpeta_resultados> -n <nombre>`

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

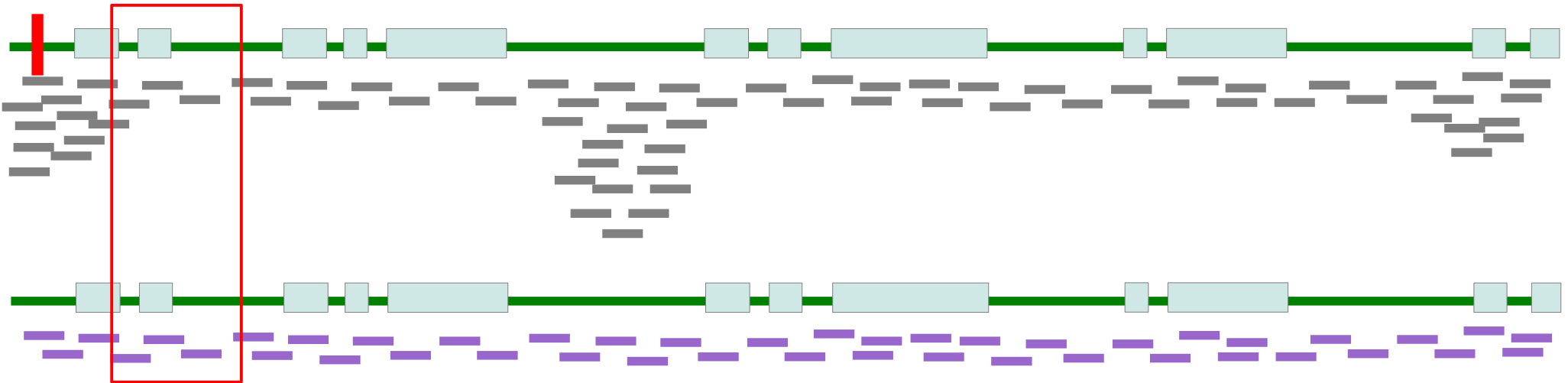


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

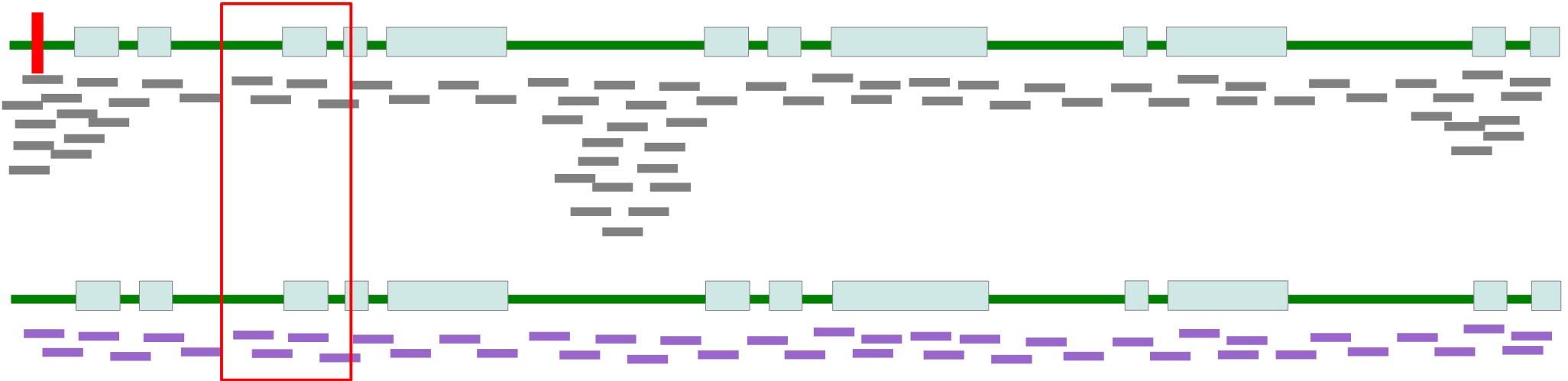


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

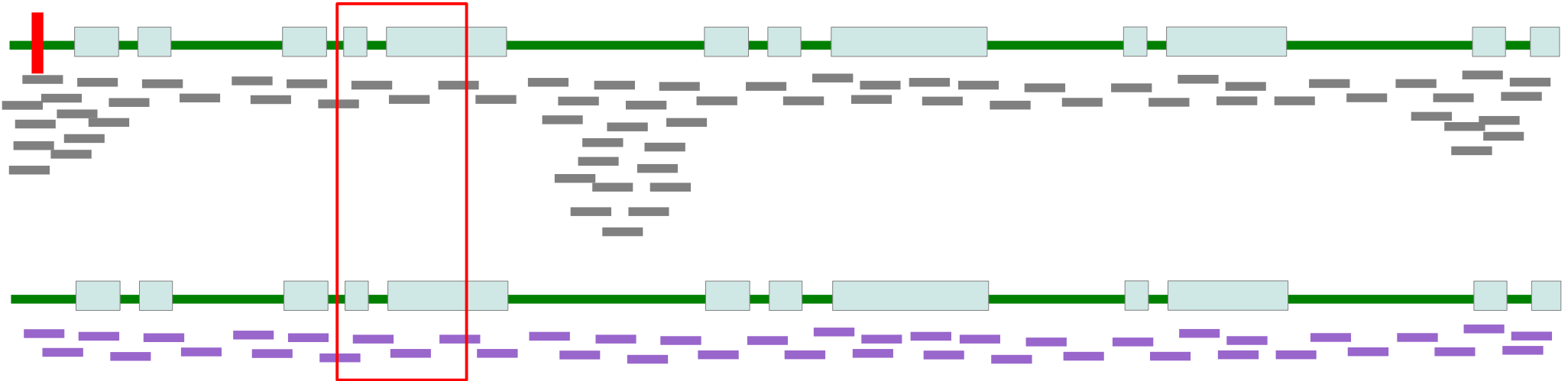


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

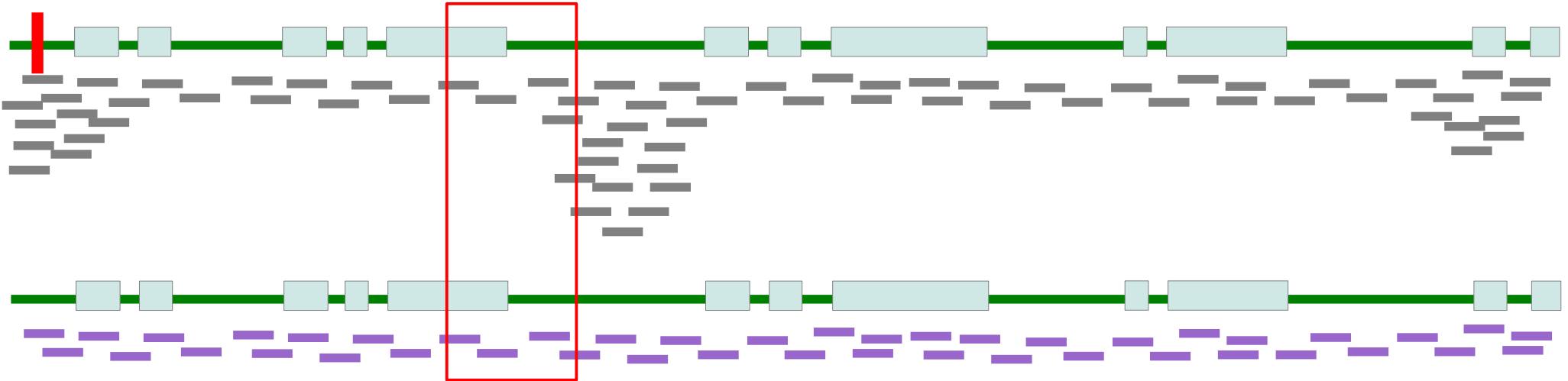


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

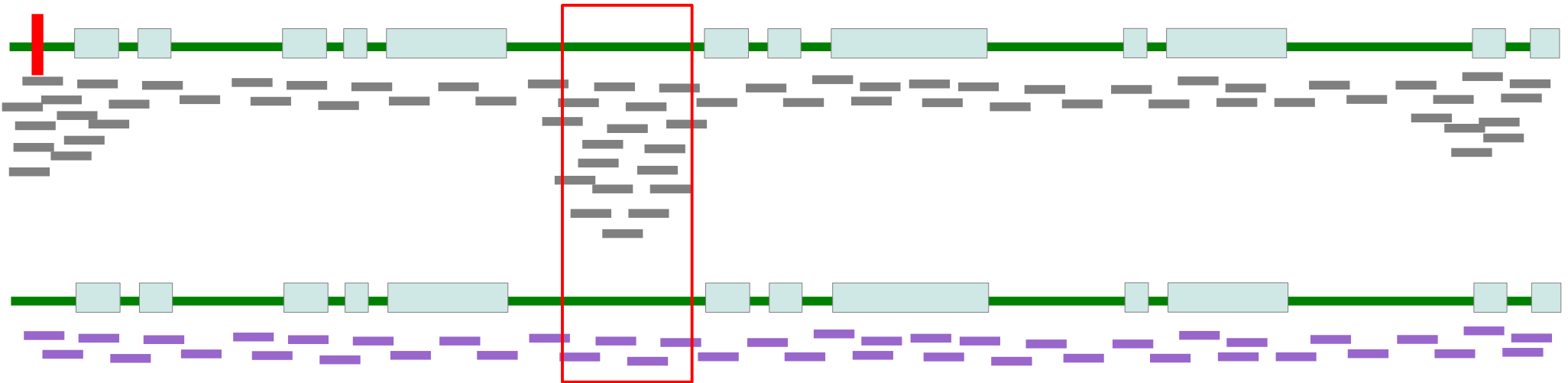


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

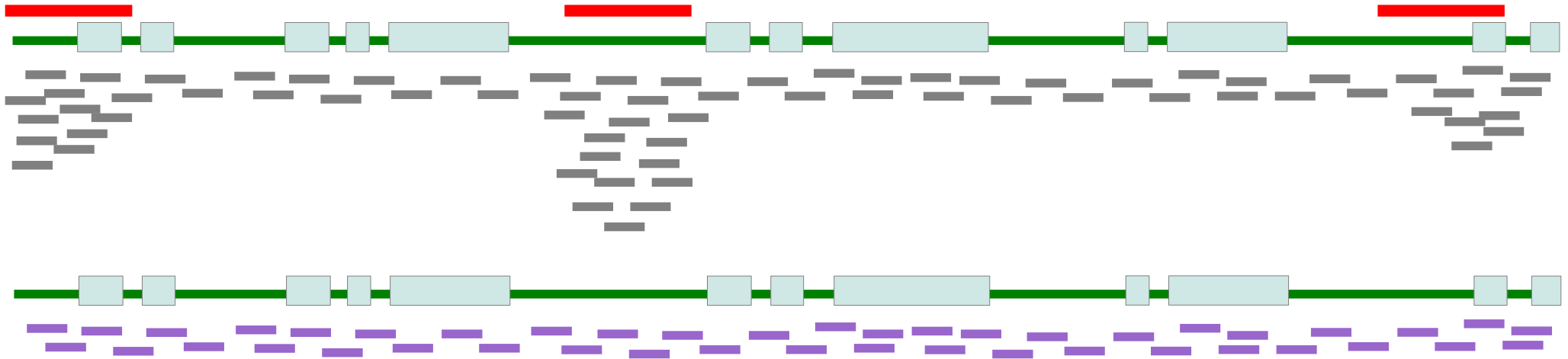


**Fastq
Control**



Determinación de picos o Peak Calling

Determinación de Picos or Peak Calling



Acceder a la carpeta results
Determinar picos usando macs2

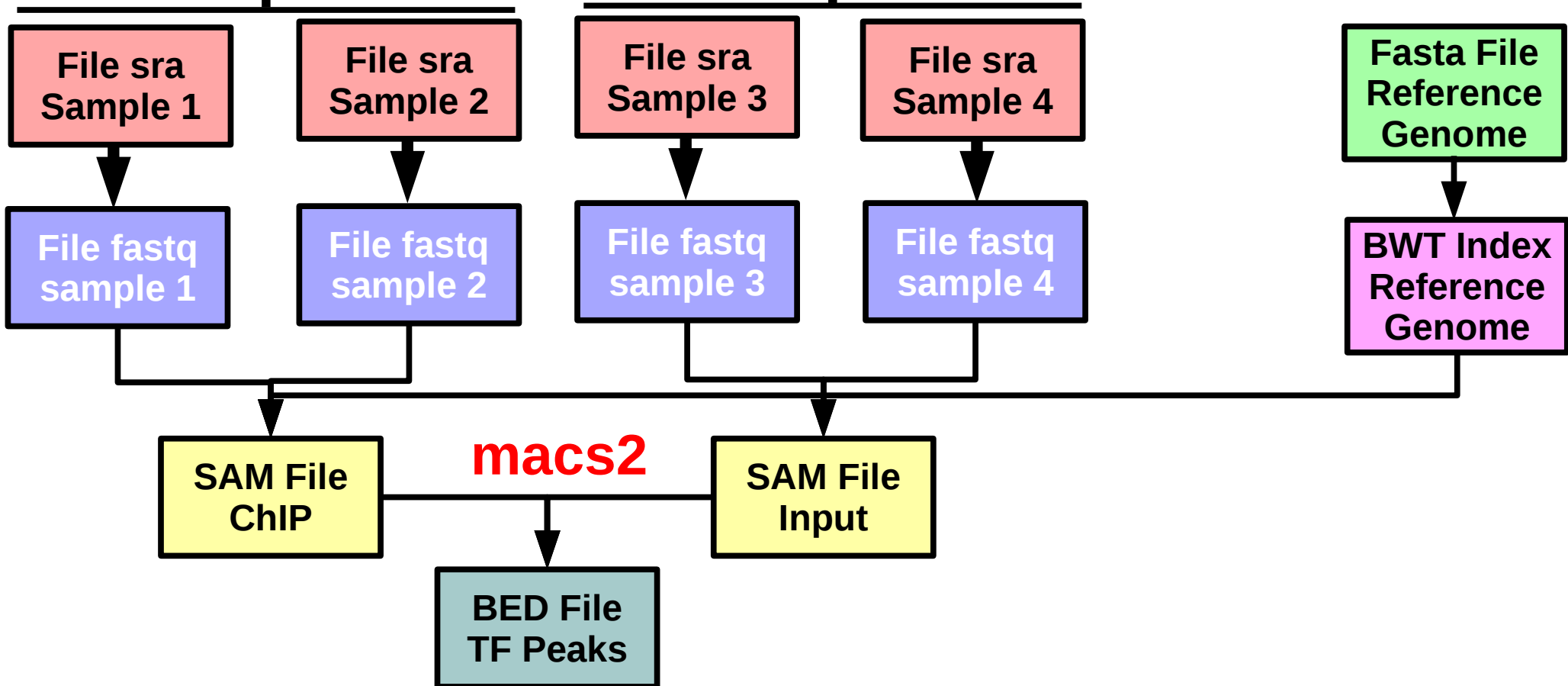
macs2 callpeak

```
-t /home/<grupo>/<exp>/samples/chip/chip.sam  
-c /home/<grupo>/<exp>/samples/input/input.sam  
-f SAM  
--outdir .  
-n prr5
```

ChIP

Control

Ensembl



MACS2 genera varios ficheros de salida. Los más importantes son fichero.narrow_peaks y fichero_summits.bed.

posición		identificador		fold-change
cromosoma				
fran@fran-laptop: ~/pr5/results				
1	13705	13706	pr5_peak_1	18.77059
1	16735	16736	pr5_peak_2	13.96410
1	21016	21017	pr5_peak_3	52.49042
1	37994	37995	pr5_peak_4	45.35525
1	45175	45176	pr5_peak_5	3.44016

ChIP

Control

Ensembl

File sra
Sample 1

File sra
Sample 2

File sra
Sample 3

File sra
Sample 4

Fasta File
Reference
Genome

File fastq
sample 1

File fastq
sample 2

File fastq
sample 3

File fastq
sample 4

BWT Index
Reference
Genome

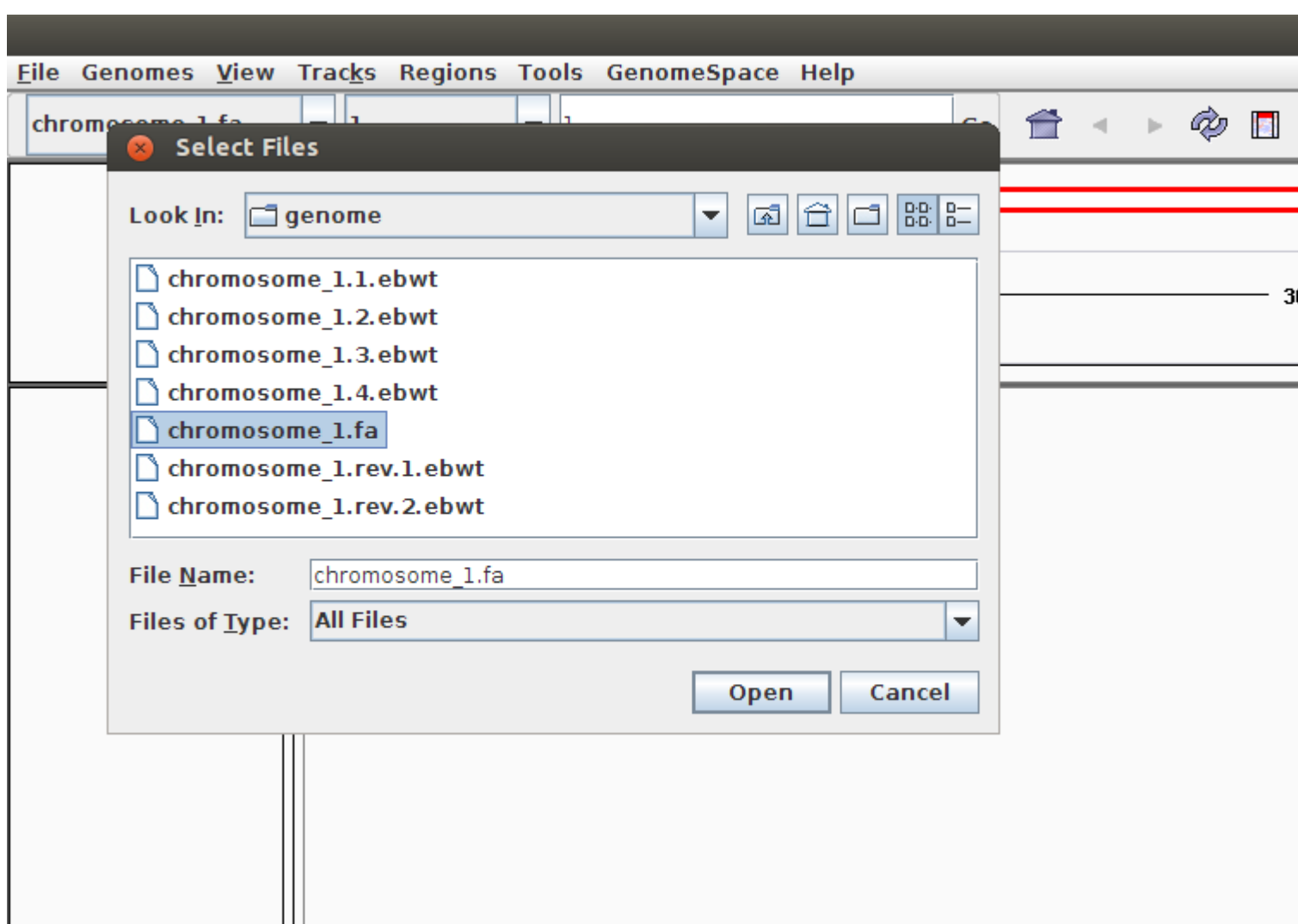
SAM File
ChIP

SAM File
Input

BED File
TF Peaks

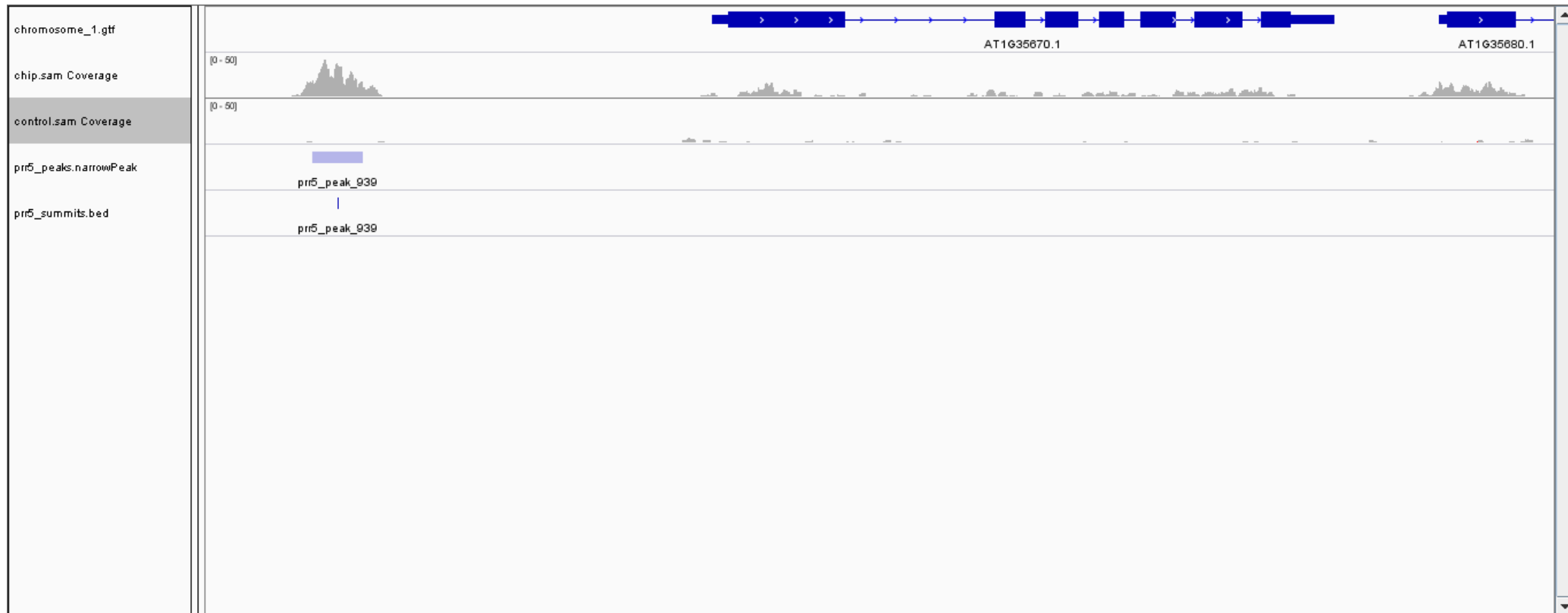
TF Peaks
Visualization

IGV



Es necesario cargar en IGV:

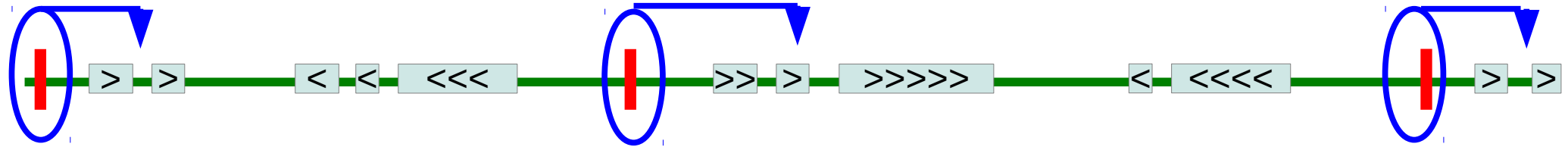
- El genoma de referencia usando: Genomes → Load Genome from File
- La anotación: File → Load From File
- Los ficheros input.sam y control.sam: File → Load From File
- Los ficheros de salida de MACS2 narrowPeaks y .bed: File → Load From File



Es necesario cargar en IGV:

- El genoma de referencia usando: Genomes → Load Genome from File
- La anotación: File → Load From File
- Los ficheros input.sam y control.sam: File → Load From File
- Los ficheros de salida de MACS2 narrowPeaks y .bed: File → Load From File

ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

ChIP

Control

Ensembl

File sra
Sample 1

File sra
Sample 2

File sra
Sample 3

File sra
Sample 4

Fasta File
Reference
Genome

File fastq
sample 1

File fastq
sample 2

File fastq
sample 3

File fastq
sample 4

BWT Index
Reference
Genome

SAM File
ChIP

SAM File
Input

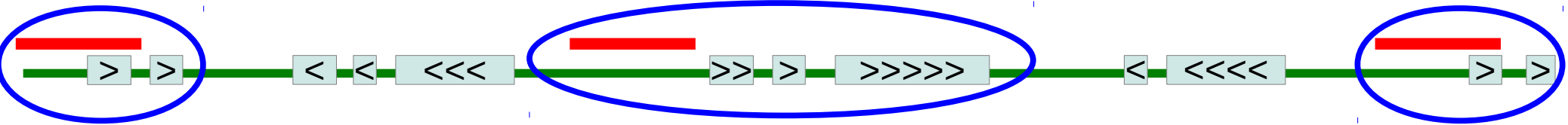
BED File
TF Peaks

TF Peaks
Visualization

PeakAnnotator

Nearest Downstream
Gene Identification

ChIP-seq: Anotación de Picos o Determinación de dianas



El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

Determinar el reguloma usando PeakAnnotator

```
java -jar ../../scripts/PeakAnnotator.jar -u NDG -g all -p prr5_summits.bed  
-a ../../annotation/chromosome_1.gtf -o .
```

Ejecutar el script target_genes.R cambiando los parámetros apropiados.

Distancia
hebra directa

Gen más cercano
Hebra directa

1	Chromosome	Start	End	#Overlaped_Genes	Downstream_FW_Gene	Symbol
2	chr1	13705	13706	1	AT1G01040.1 DCL1 9440	AT1G01020.2 ARV1 4968
3	chr1	16735	16736	0	AT1G01040.1 DCL1 6410	AT1G01030.1 NGA3 3021
4	chr1	21016	21017	0	AT1G01040.1 DCL1 2129	AT1G01030.1 NGA3 7302
5	chr1	37994	37995	0	AT1G01073.1 false 6682	AT1G01060.5 LHY 123

Gen más cercano
Hebra inversa

Distancia
hebra inversa

ChIP

Control

Ensembl

File sra
Sample 1

File sra
Sample 2

File sra
Sample 3

File sra
Sample 4

Fasta File
Reference
Genome

File fastq
sample 1

File fastq
sample 2

File fastq
sample 3

File fastq
sample 4

BWT Index
Reference
Genome

SAM File
ChIP

SAM File
Input

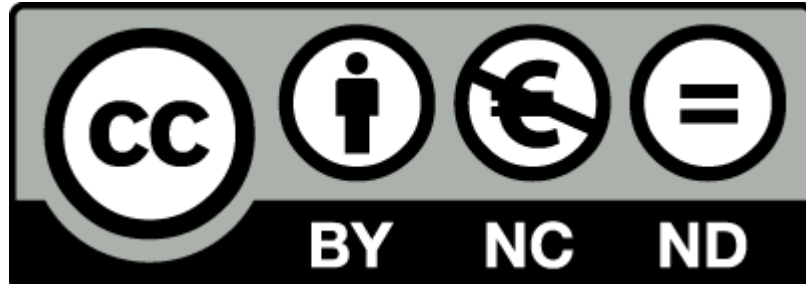
BED File
TF Peaks

TF Peaks
Visualization

Nearest Downstream
Gene Identification

homer

TF Binding
Sequence



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>.