

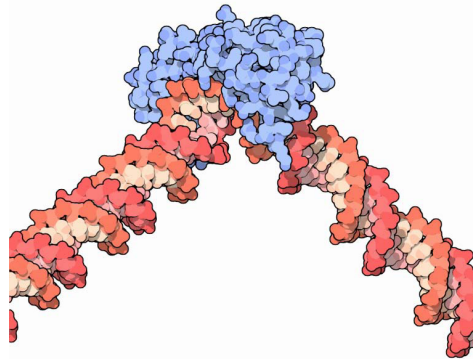
Análisis de datos de ChIP-seq: Identificación de Sitios de Unión de Factores de Transcripción y Marcas Epigenéticas

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

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

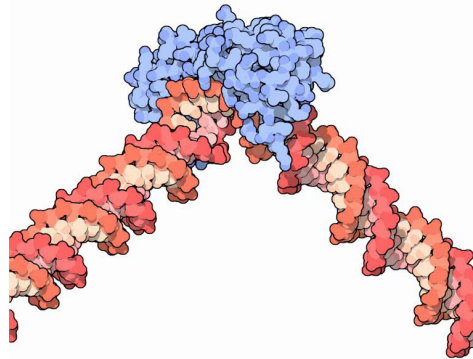
Los Factores de Transcripción se unen a regiones cis para regular la transcripción

- 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**.

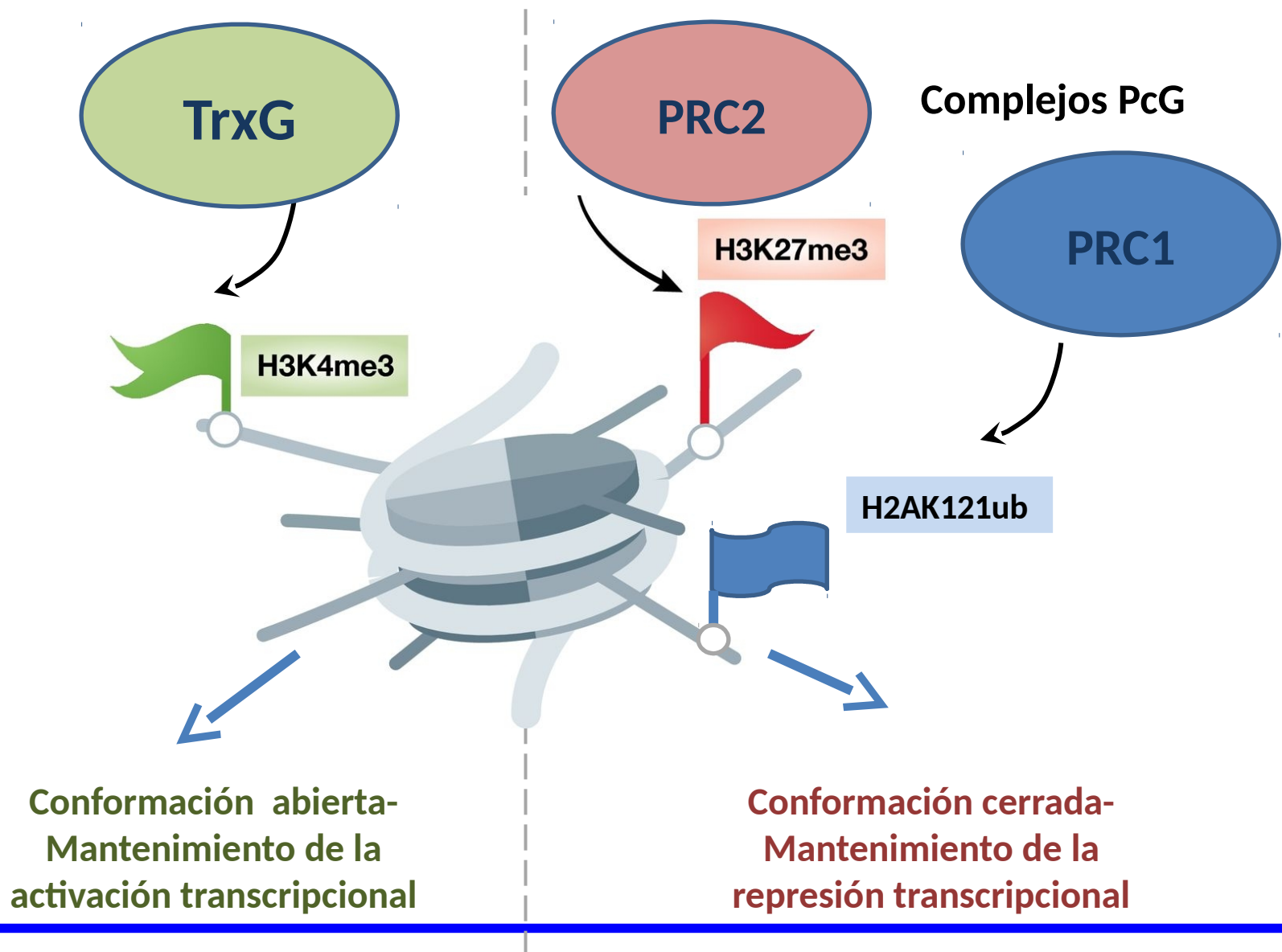


El cistroma o conjunto de regiones cis asociadas a un FT puede ser enormemente plástico

- 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.



Las modificaciones de histonas afectan a la accesabilidad de la cromatina manteniendo la represión/activación de la expresión génica



ChIP-Seq determina el cistroma de un FT

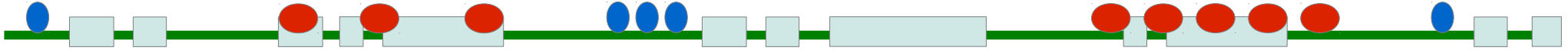
- Una técnica ómica que permite determinar los sitios de interacción entre proteínas y DNA en unas condiciones específicas se denomina **ChIP-Seq** (Chromatin Immunoprecipitation coupled with Sequencing).
- Esta técnica combina dos metodologías ya establecidas:
 - **ChIP**: Chromatin Immuno-Precipitation
 - **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

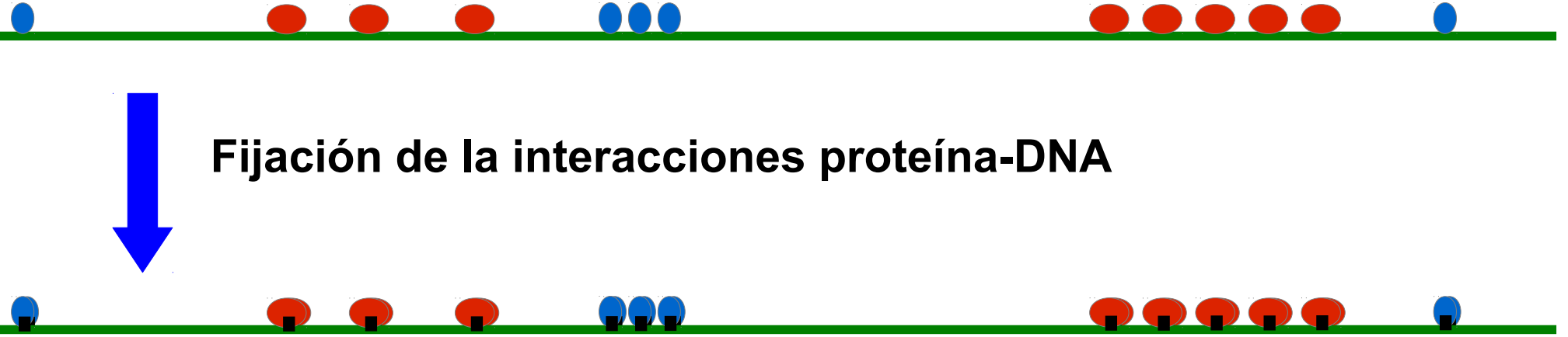


Chromatin Immunoprecipitation coupled with sequencing



Chromatin Immunoprecipitation coupled with sequencing

Fijación de la interacciones proteína-DNA



Chromatin Immunoprecipitation coupled with sequencing



Fijación de la interacciones proteína-DNA



Fragmentación de DNA



Chromatin Immunoprecipitation coupled with sequencing



Chromatin Immunoprecipitation coupled with sequencing



Microperlas con anticuerpos específicos contra FT



Chromatin Immunoprecipitation coupled with sequencing

Microperlas con anticuerpos específicos contra FT

Inmuno-precipitación

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

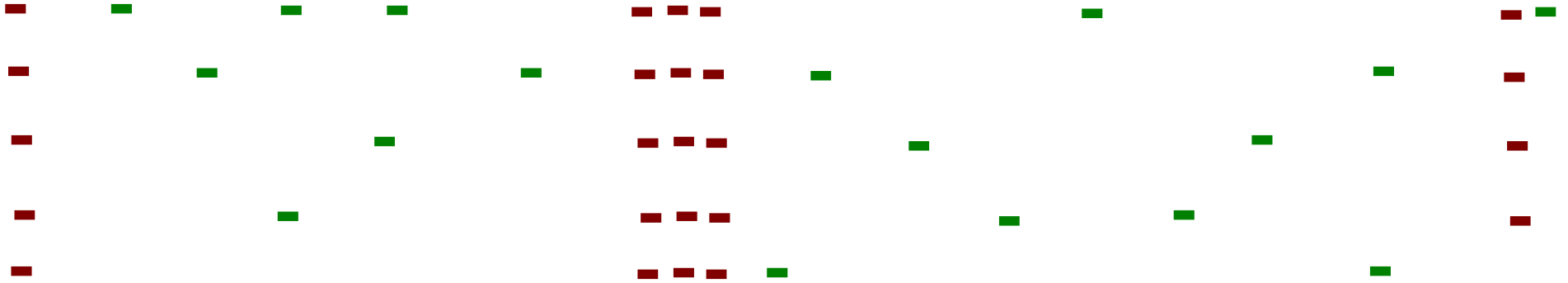


...

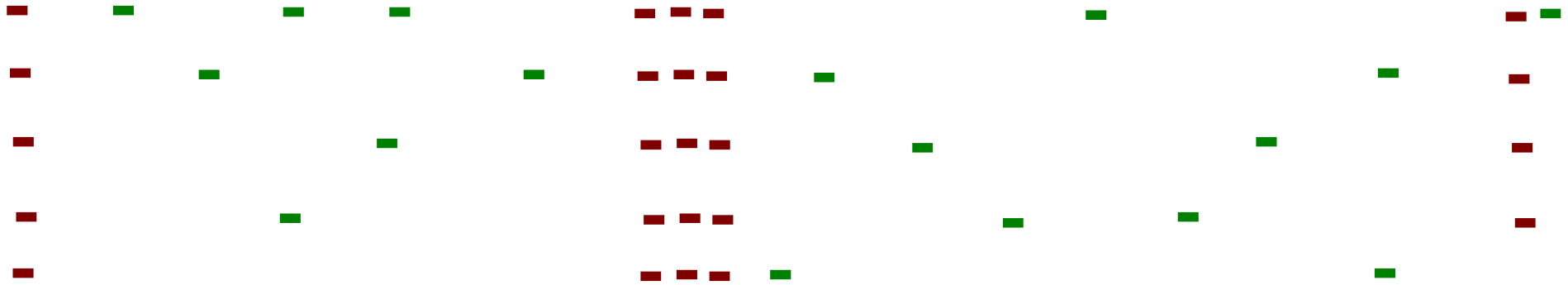
Célula N



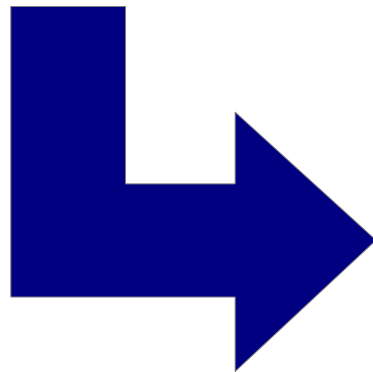
Chromatin Immunoprecipitation coupled with sequencing



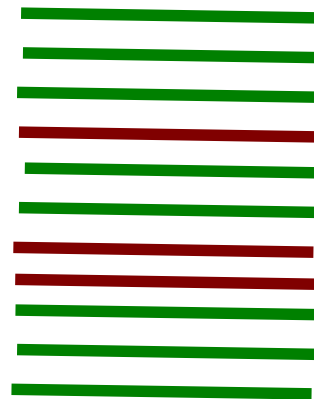
Chromatin Immunoprecipitation coupled with sequencing



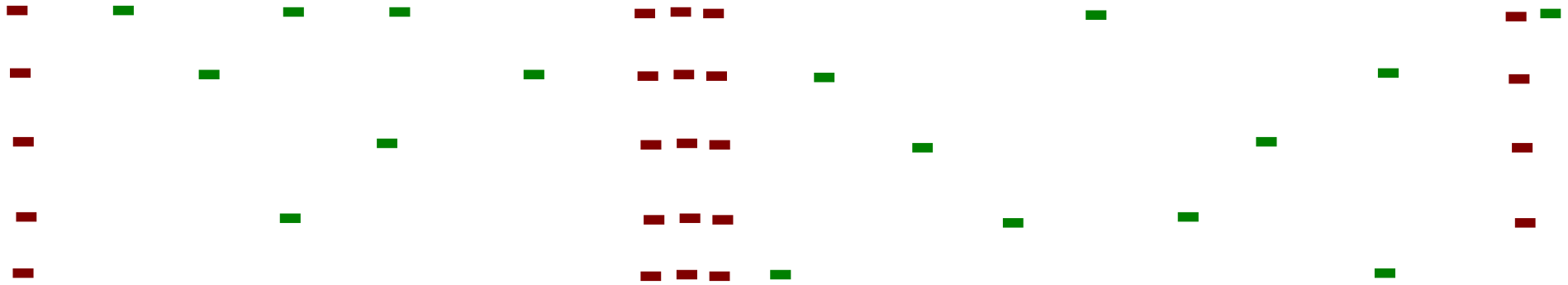
Secuenciación
de altas prestaciones



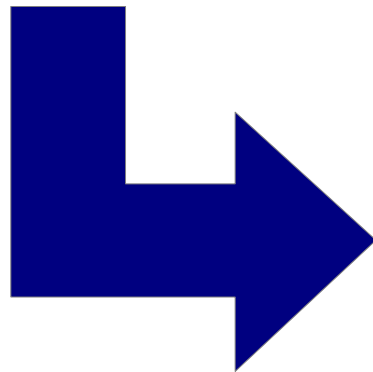
Fichero fastq



Chromatin Immunoprecipitation coupled with sequencing



**Secuenciación
de altas prestaciones**



Fichero fastq

A representation of a Fastq file, showing multiple lines of text (reads) separated by blank lines. The lines are gray and represent the sequence data output from the sequencing process.

Análisis de datos de ChIP-seq

El análisis de datos de ChIP-seq para el determinar los sitios del genoma donde se une un factor de transcripción en concreto o donde se encuentra localizada un modificación de histonas se divide en los siguientes pasos:

- **Paso 1:** Preparación del espacio de trabajo.
- **Paso 2:** Análisis de calidad de la lecturas.
- **Paso 3:** Mapeado de lecturas cortas al genoma de referencia.
- **Paso 4:** Determinación de picos.
- **Paso 5:** Asociación de picos a genes diana.

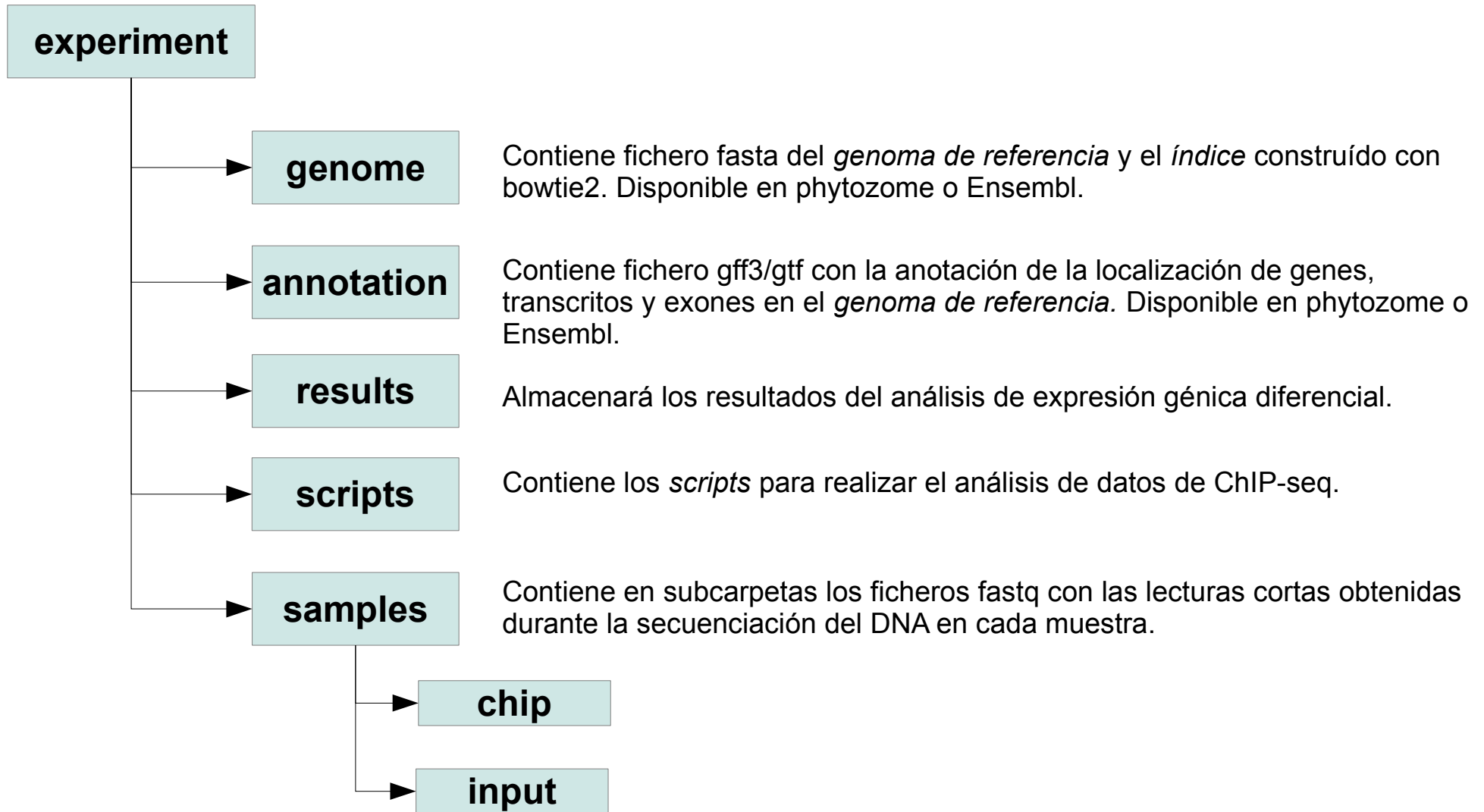
Análisis de datos de ChIP-seq

El análisis de datos de ChIP-seq para el determinar los sitios del genoma donde se une un factor de transcripción en concreto o donde se encuentra localizada un modificación de histonas se divide en los siguientes pasos:

- **Paso 1:** Preparación del espacio de trabajo.
- **Paso 2:** Análisis de calidad de la lecturas.
- **Paso 3:** Mapeado de lecturas cortas al genoma de referencia.
- **Paso 4:** Determinación de picos.
- **Paso 5:** Asociación de picos a genes diana.

Paso 1: Preparación del espacio de trabajo.

- Es crítico **mantener un espacio de trabajo ordenado**. Se recomienda la siguiente distribución:



Análisis de datos de ChIP-seq

El análisis de datos de ChIP-seq para el determinar los sitios del genoma donde se une un factor de transcripción en concreto o donde se encuentra localizada un modificación de histonas se divide en los siguientes pasos:

- **Paso 1:** Preparación del espacio de trabajo.
- **Paso 2:** Análisis de calidad de la lecturas.
- **Paso 3:** Mapeado de lecturas cortas al genoma de referencia.
- **Paso 4:** Determinación de picos.
- **Paso 5:** Asociación de picos a genes diana.

Análisis de datos de ChIP-seq

El análisis de datos de ChIP-seq para el determinar los sitios del genoma donde se une un factor de transcripción en concreto o donde se encuentra localizada un modificación de histonas se divide en los siguientes pasos:

- **Paso 1:** Preparación del espacio de trabajo.
- **Paso 2:** Análisis de calidad de la lecturas.
- **Paso 3:** Mapeado de lecturas cortas al genoma de referencia.
- **Paso 4:** Determinación de picos.
- **Paso 5:** Asociación de picos a genes diana.

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**



Chromatin Immunoprecipitation coupled with sequencing



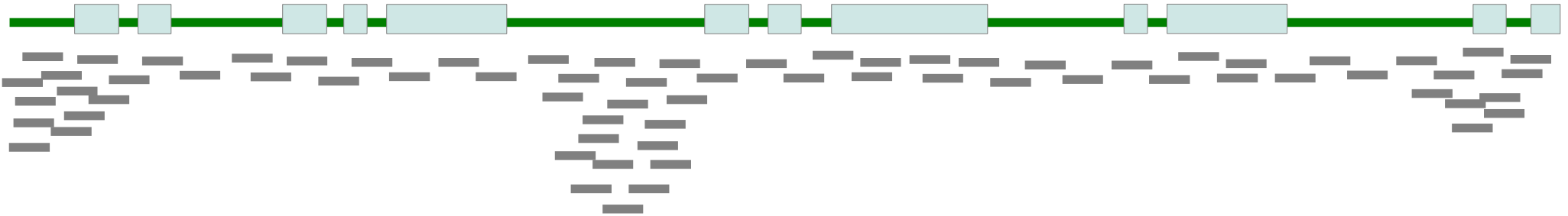
**Fastq
ChIP**



**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



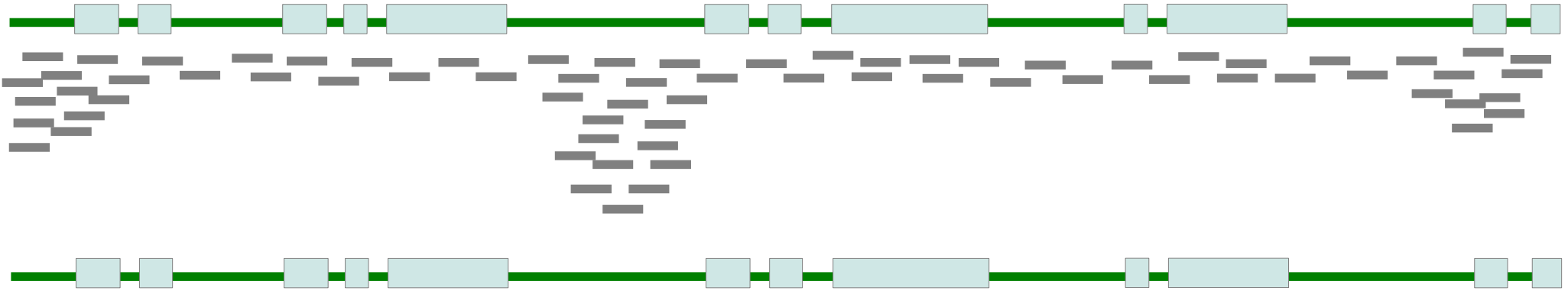
Típicamente se considera uno de los dos siguientes tipos de control:

- **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



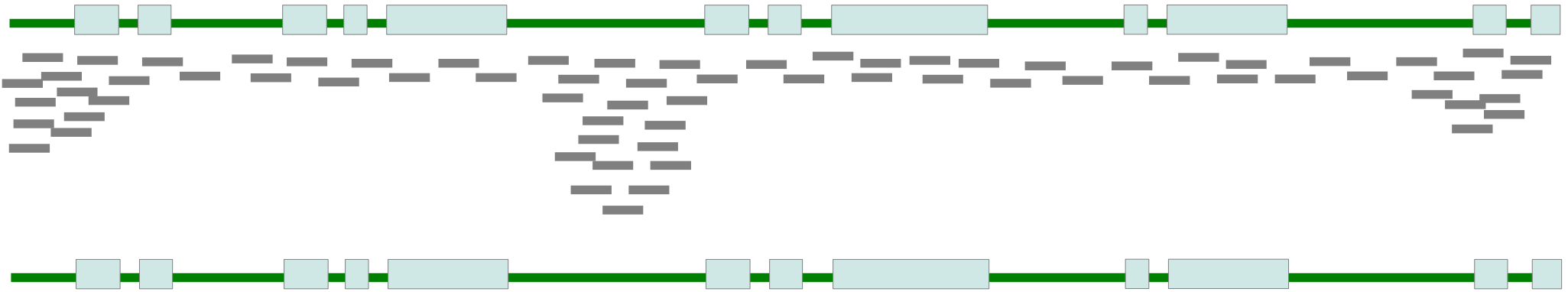
**Fastq
ChIP**



**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**

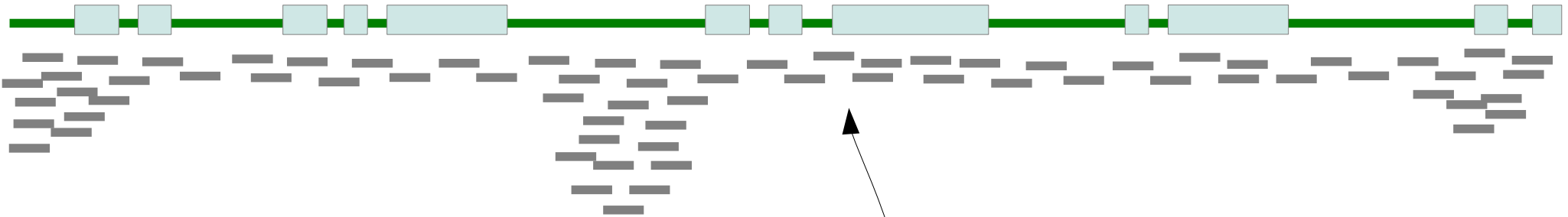


**Mapeo de lecturas
cortas al genoma de
referencia**

```
cd Desktop/prr5/genome
bowtie2-build chromosome1.fa index

cd ../samples/chip
bowtie2 -x ../../genome/index -U chip_prr5_chr1.fastq -S chip.sam
```

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

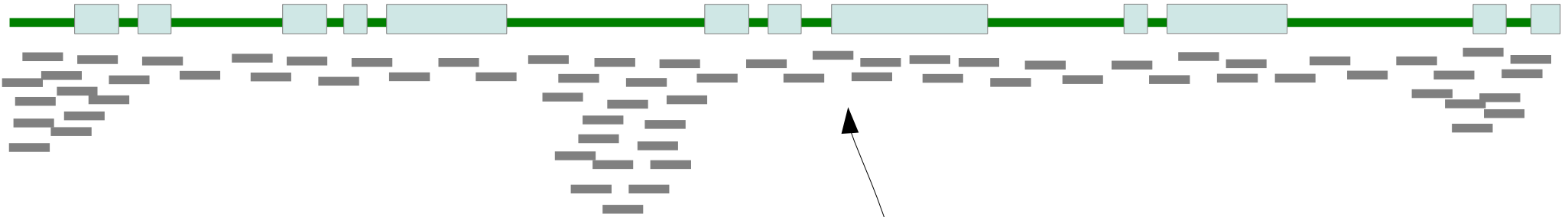


**Mapeo de lecturas
cortas al genoma de
referencia**

```
cd Desktop/prr5/genome  
bowtie2-build chromosome1.fa index
```

```
cd ../samples/chip  
bowtie2 -x ../../genome/index -U chip_prr5_chr1.fastq -S chip.sam
```

Chromatin Immunoprecipitation coupled with sequencing



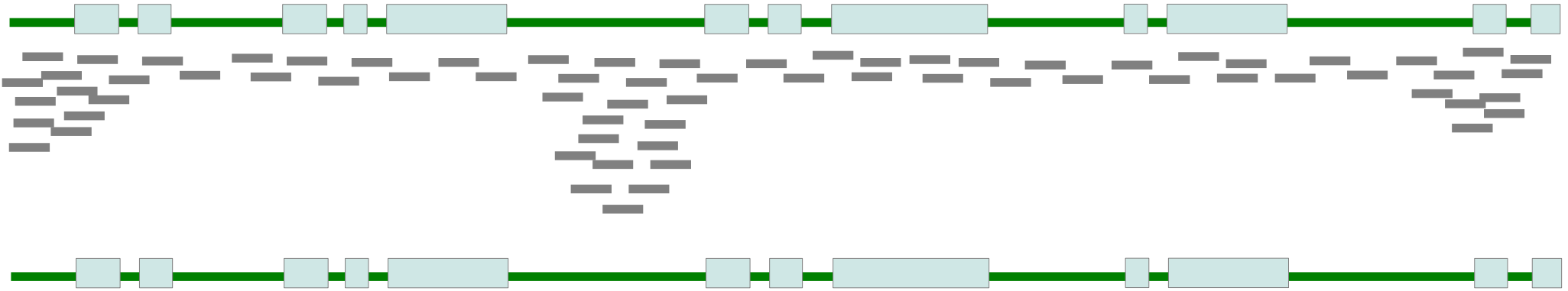
**Fastq
ChIP**



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

```
samtools sort -o chip.bam chip.sam  
samtools index chip.bam
```

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**



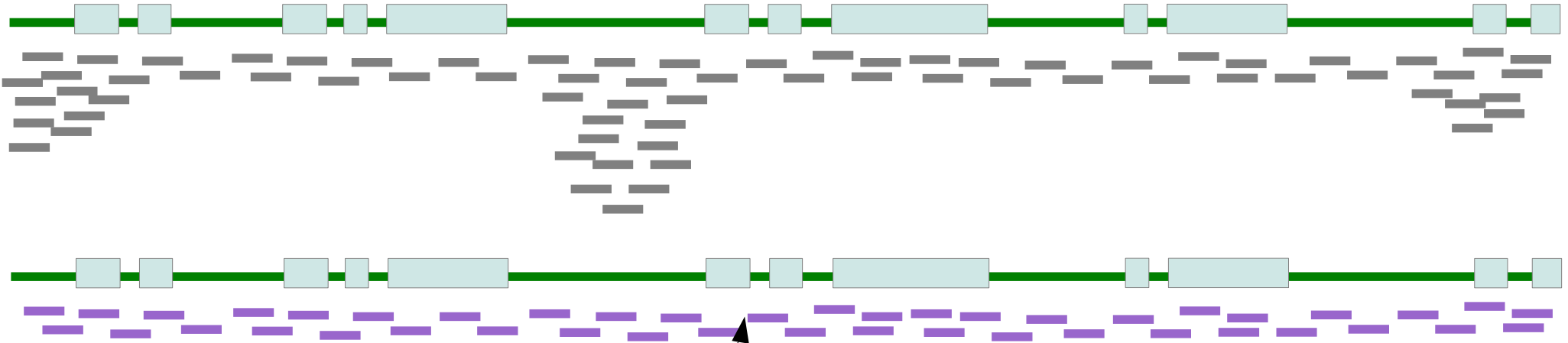
Mapeo de lecturas
cortas al genoma de
referencia

```
cd ../input  
bowtie2 -x ../../genome/index -U input_prr5_chr1.fastq -S input.sam
```

**Fastq
Control**



Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**



Mapeo de lecturas
cortas al genoma de
referencia

**Fastq
Control**



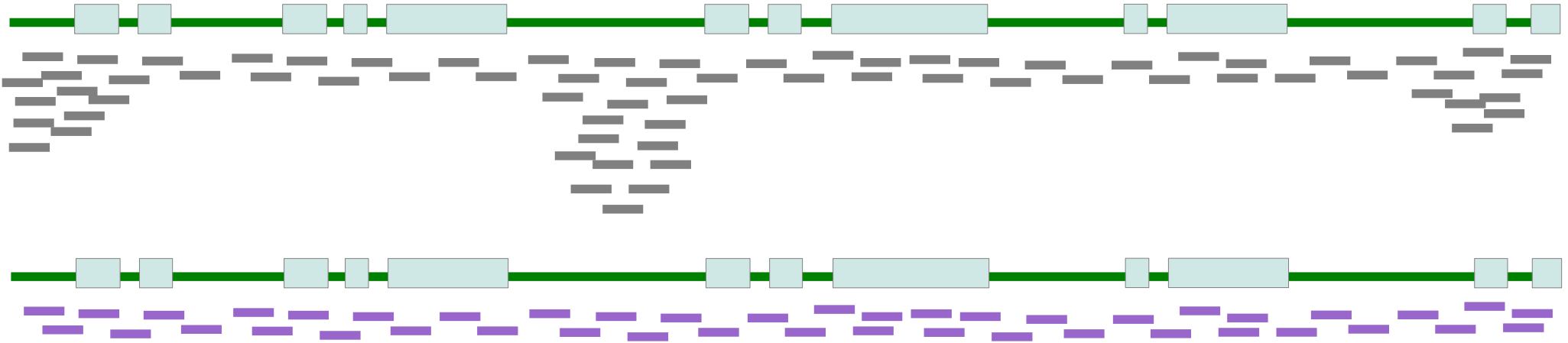
```
cd ../input  
bowtie2 -x ../../genome/index -U input_prr5_chr1.fastq -S input.sam
```

Análisis de datos de ChIP-seq

El análisis de datos de ChIP-seq para el determinar los sitios del genoma donde se une un factor de transcripción en concreto o donde se encuentra localizada un modificación de histonas se divide en los siguientes pasos:

- **Paso 1:** Preparación del espacio de trabajo.
- **Paso 2:** Análisis de calidad de la lecturas.
- **Paso 3:** Mapeado de lecturas cortas al genoma de referencia.
- **Paso 4:** Determinación de picos.
- **Paso 5:** Asociación de picos a genes diana.

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

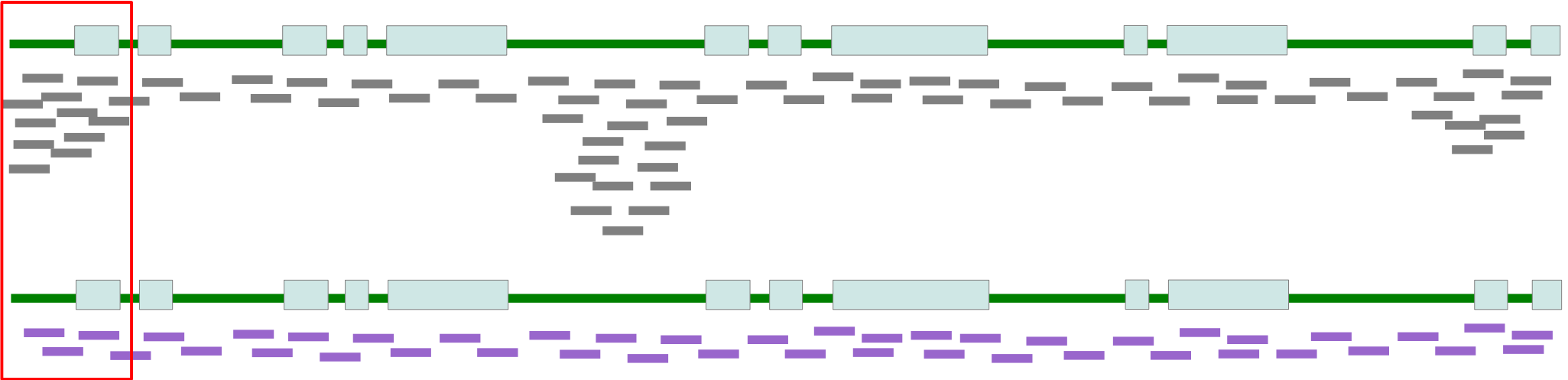


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

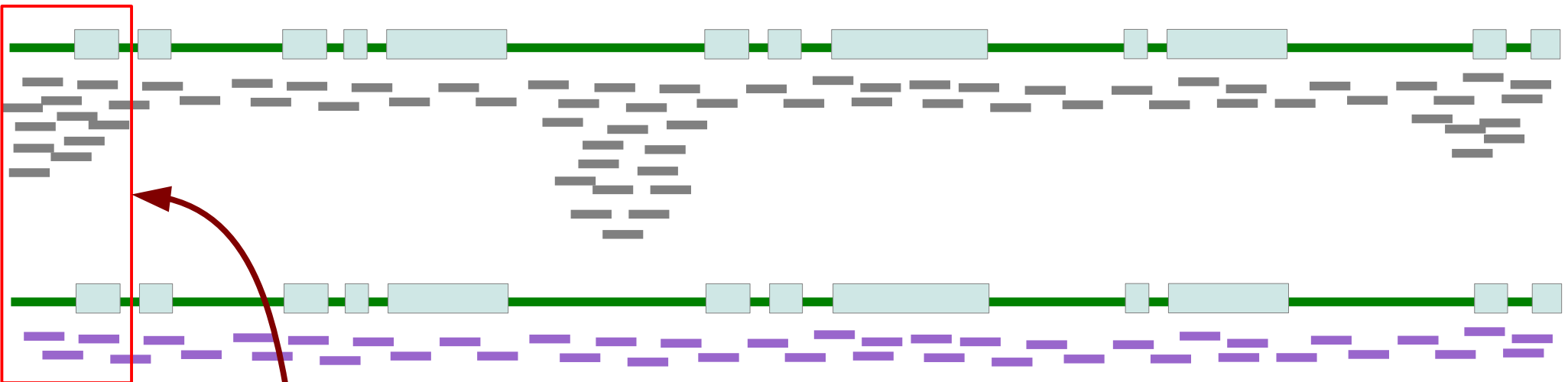


**Fastq
Control**



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

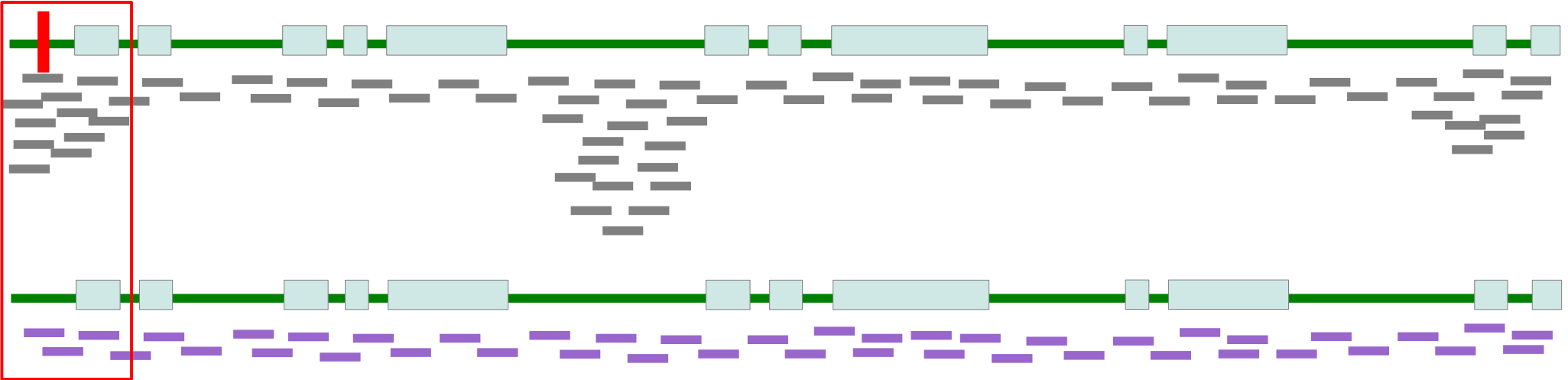
Cuando el fold-change es lo suficientemente alto y el p-valor lo suficientemente bajo se asume que existe un pico y por lo tanto existe evidencia sobre que nuestro FT se une a dicho lugar del DNA.

Determinación de picos o Peak Calling

**Fastq
Control**



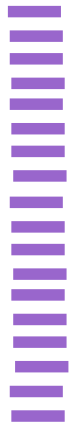
Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

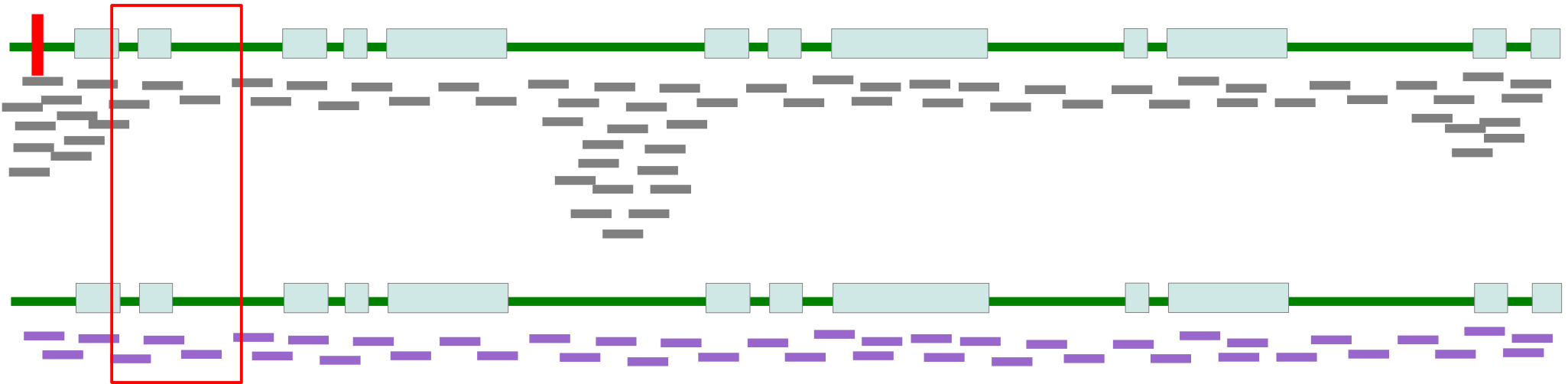


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

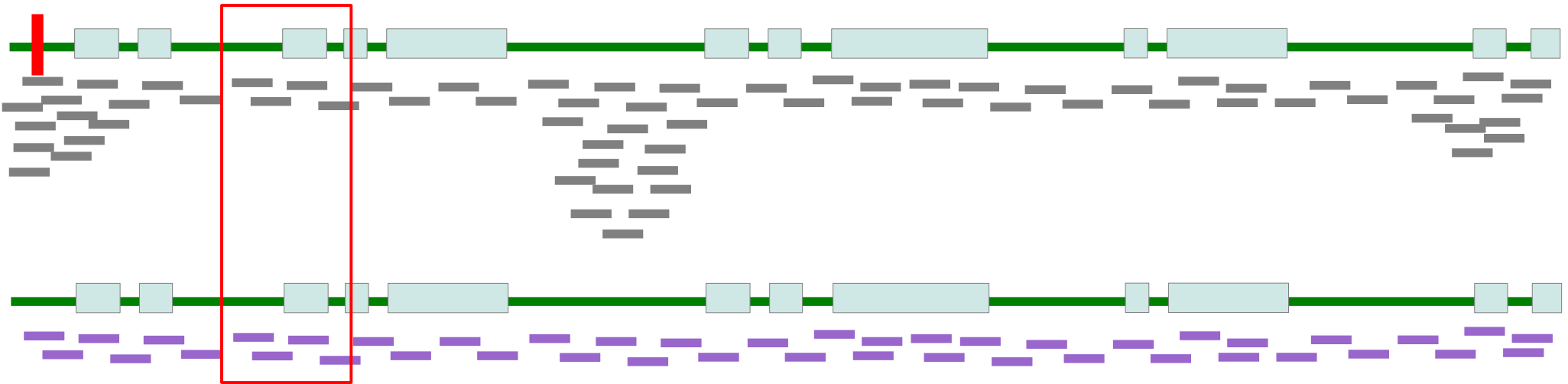


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

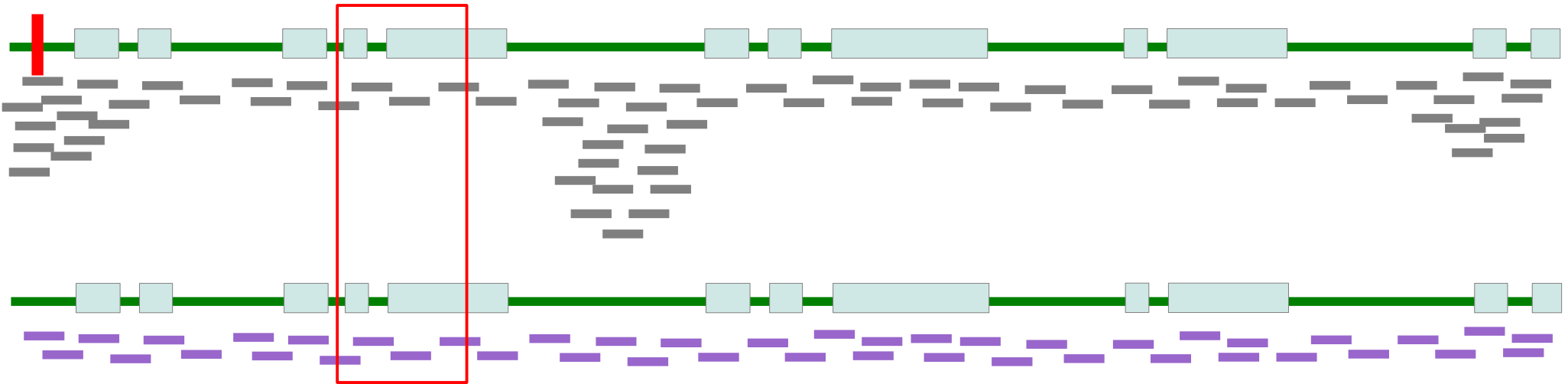


**Fastq
Control**



Determinación de picos o Peak Calling

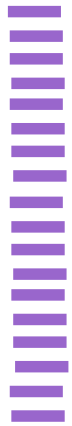
Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

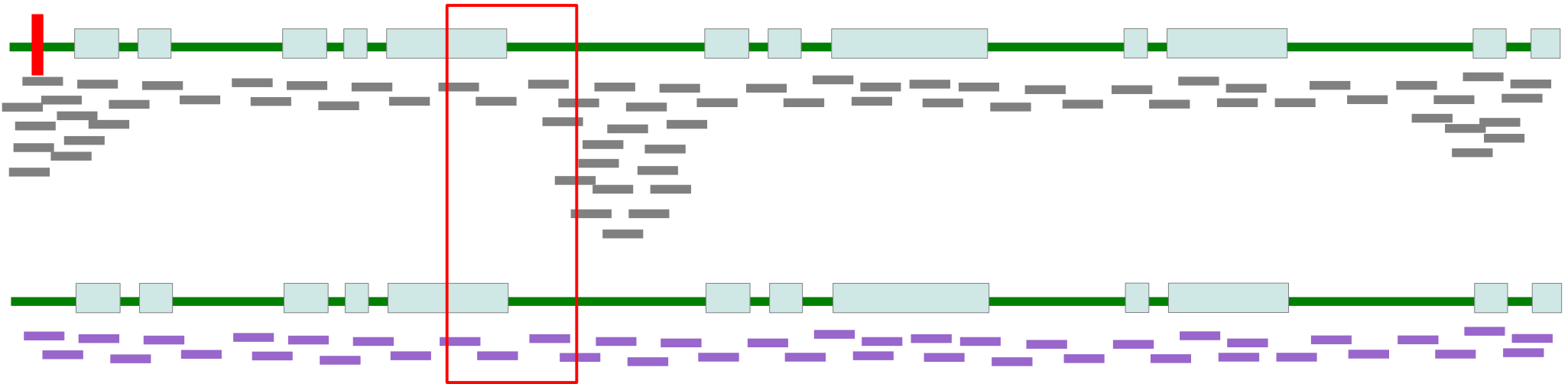


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

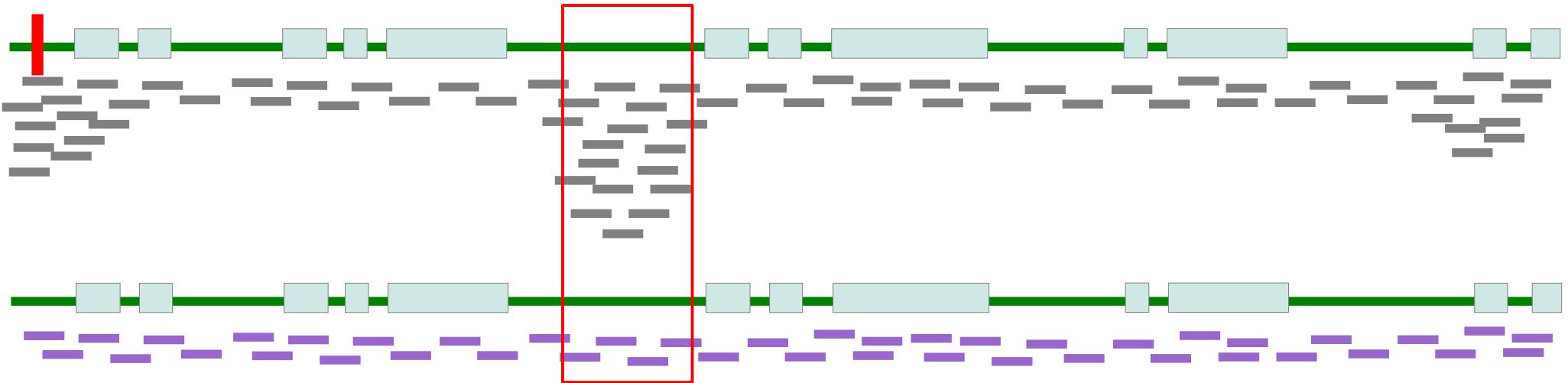


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

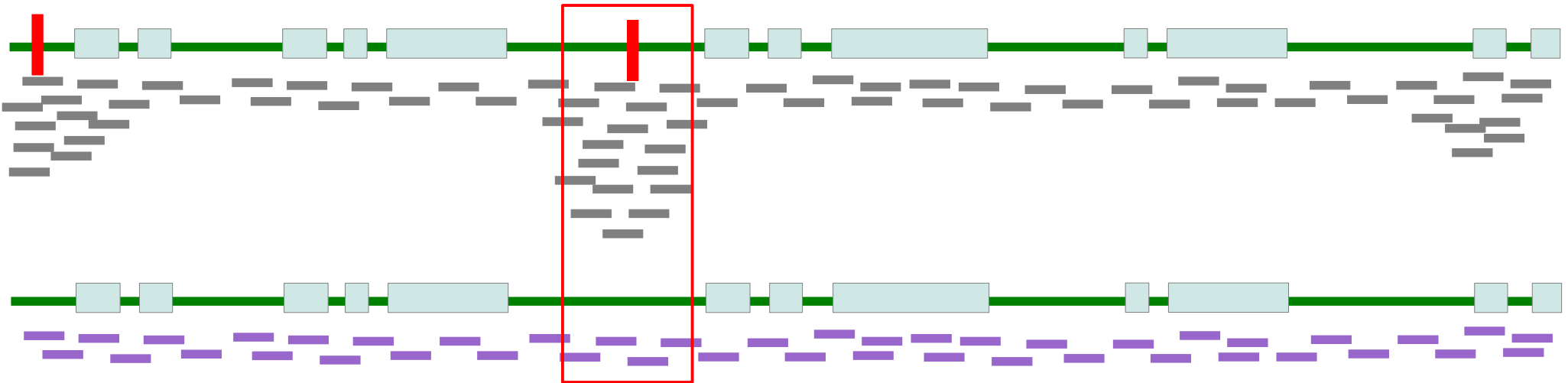


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

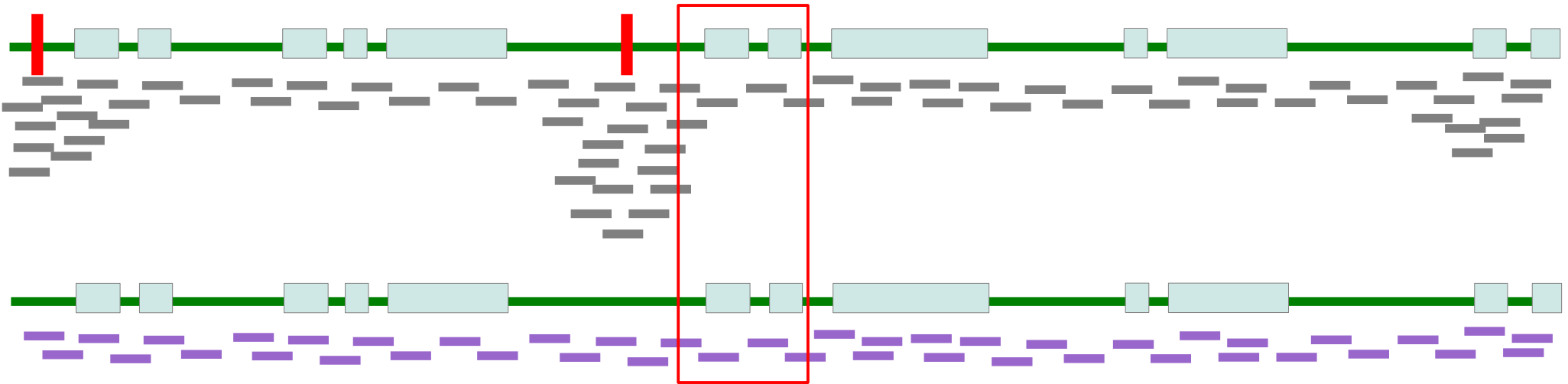


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

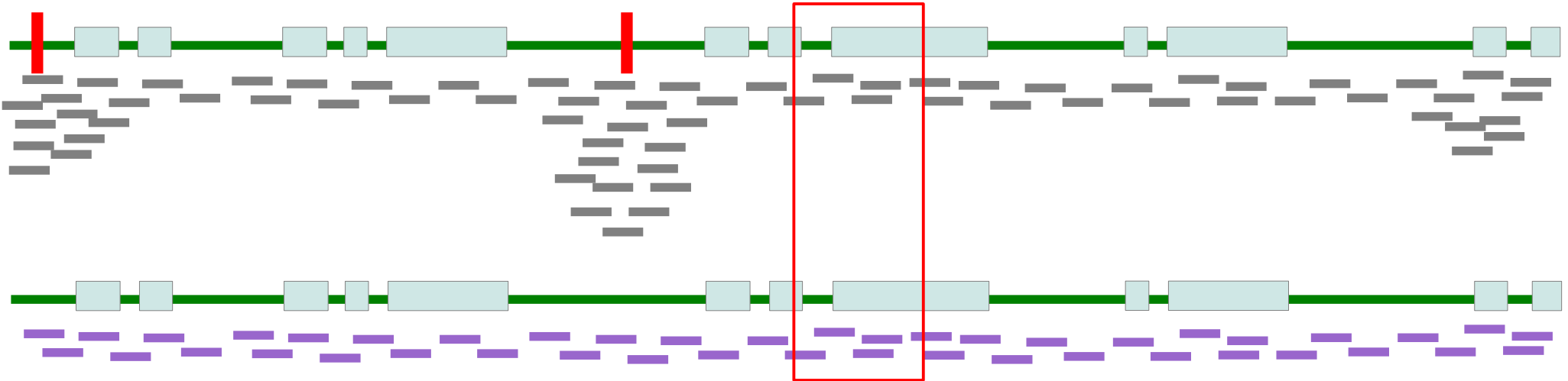


**Fastq
Control**



Determinación de picos o Peak Calling

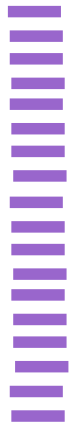
Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

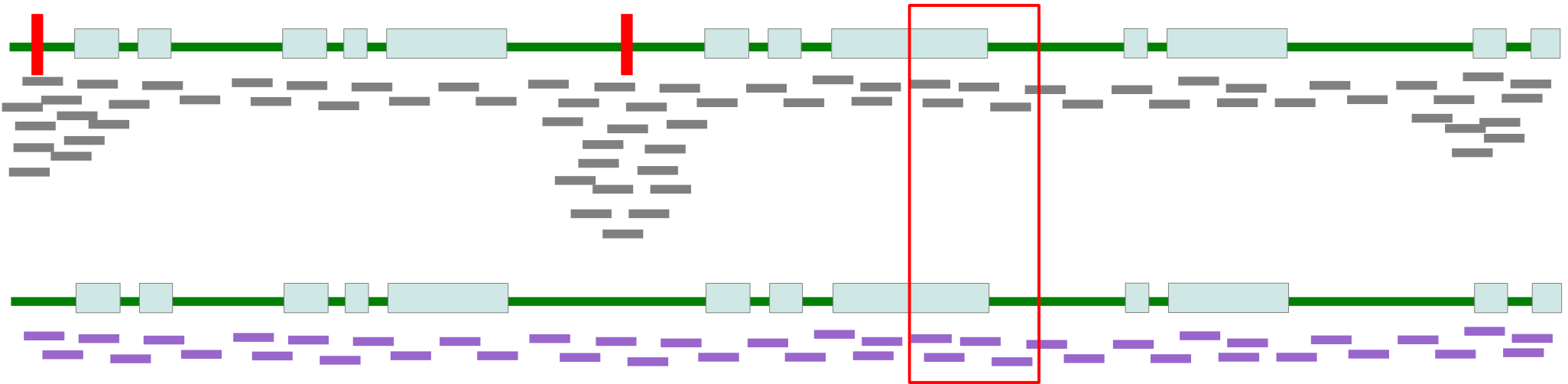


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

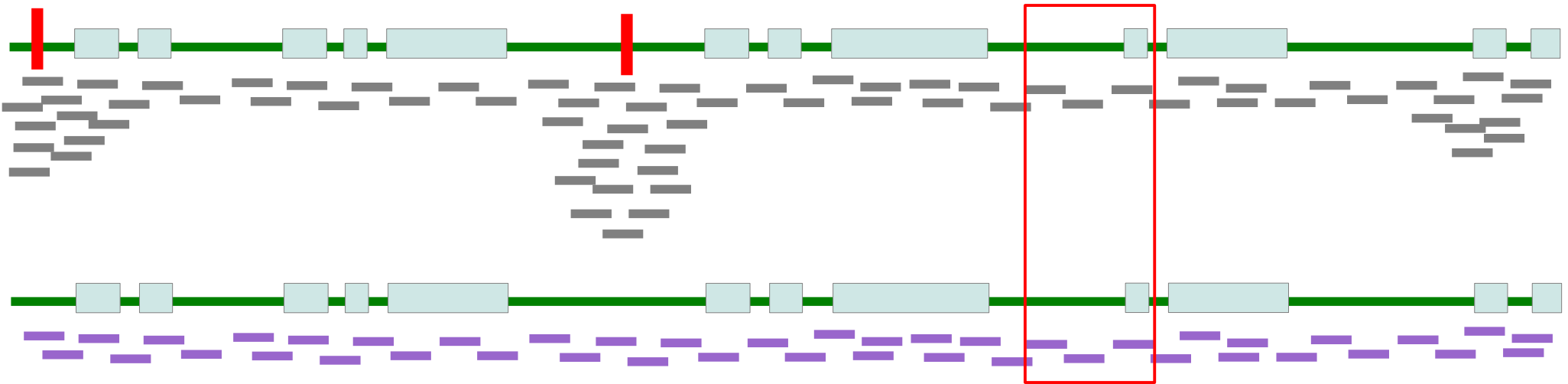


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

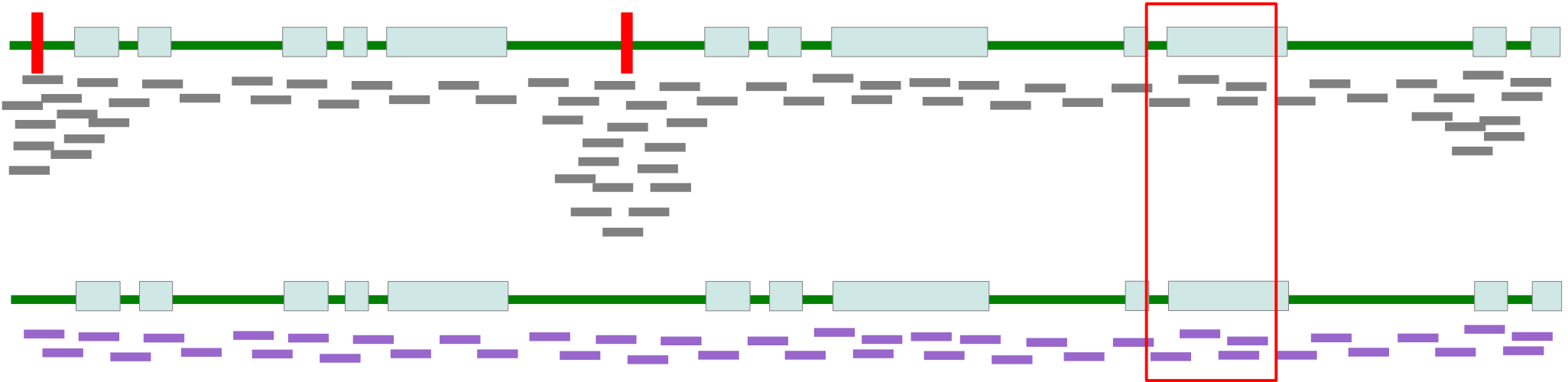


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

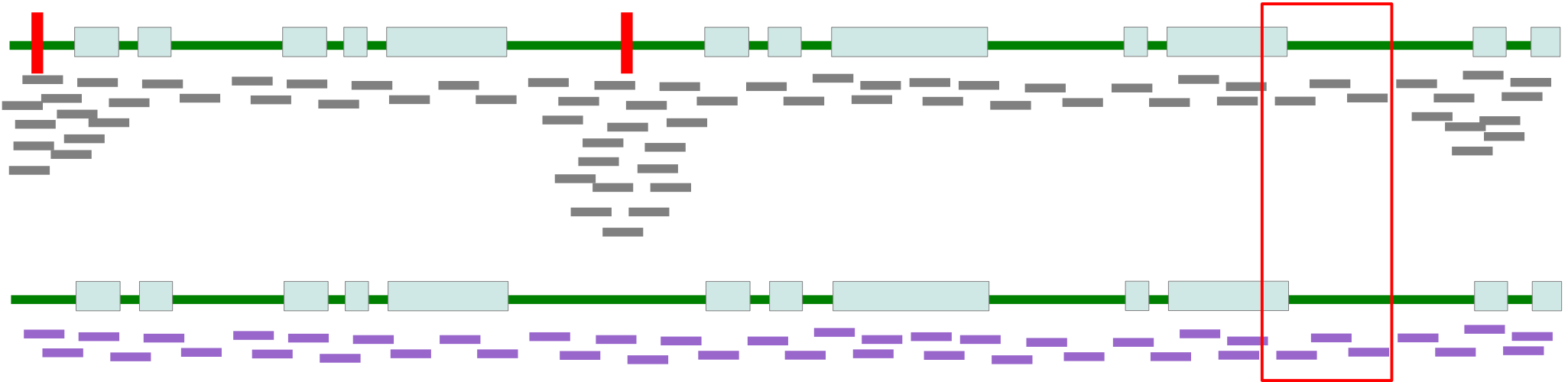


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

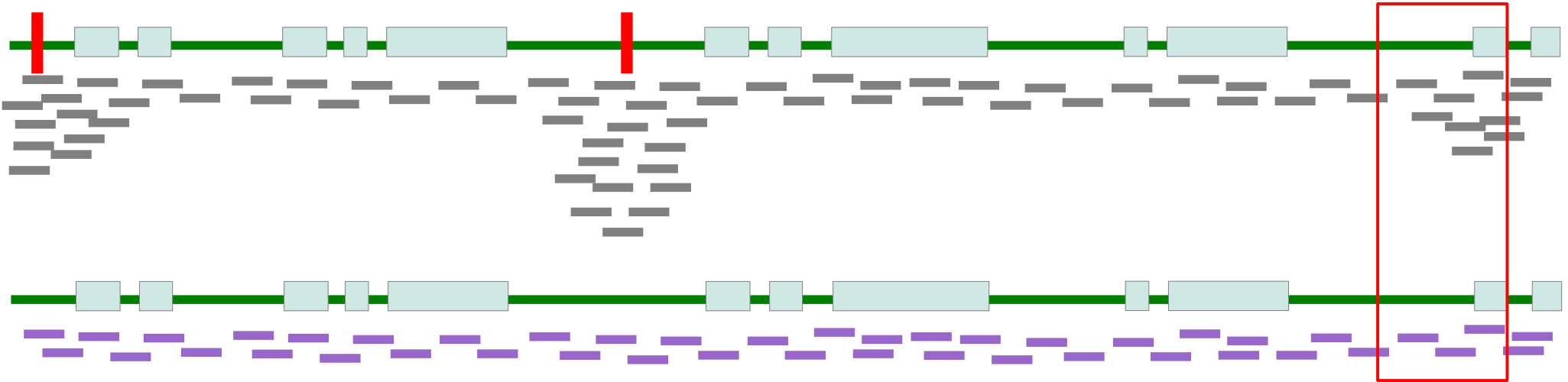


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

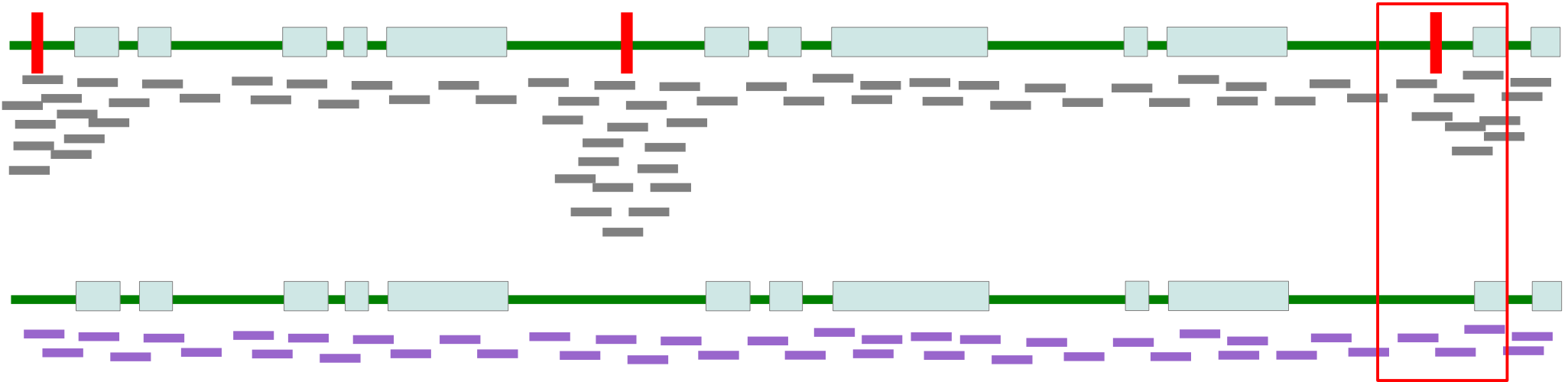


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

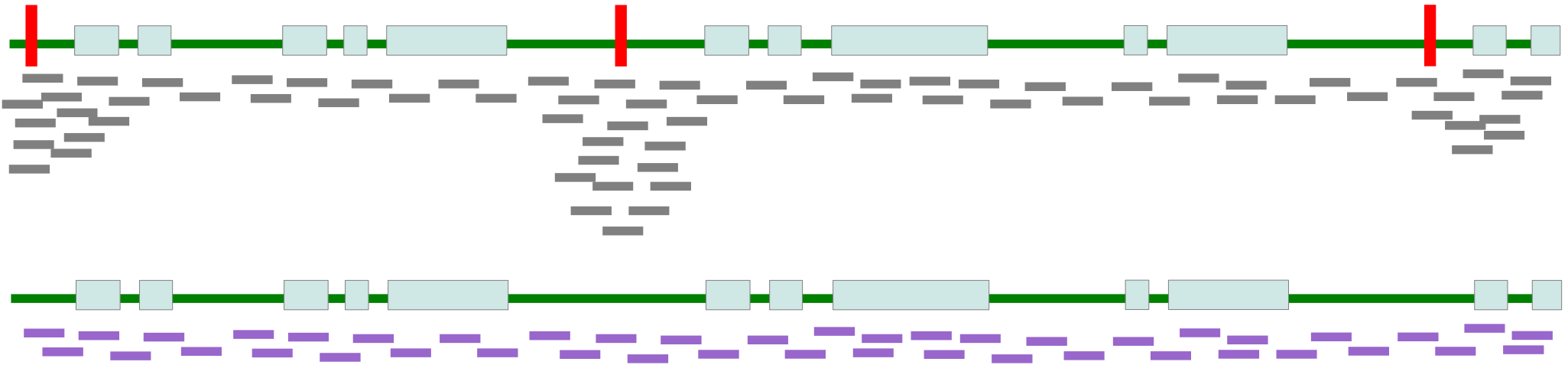


**Fastq
Control**



Determinación de picos o Peak Calling

Chromatin Immunoprecipitation coupled with sequencing



**Fastq
ChIP**

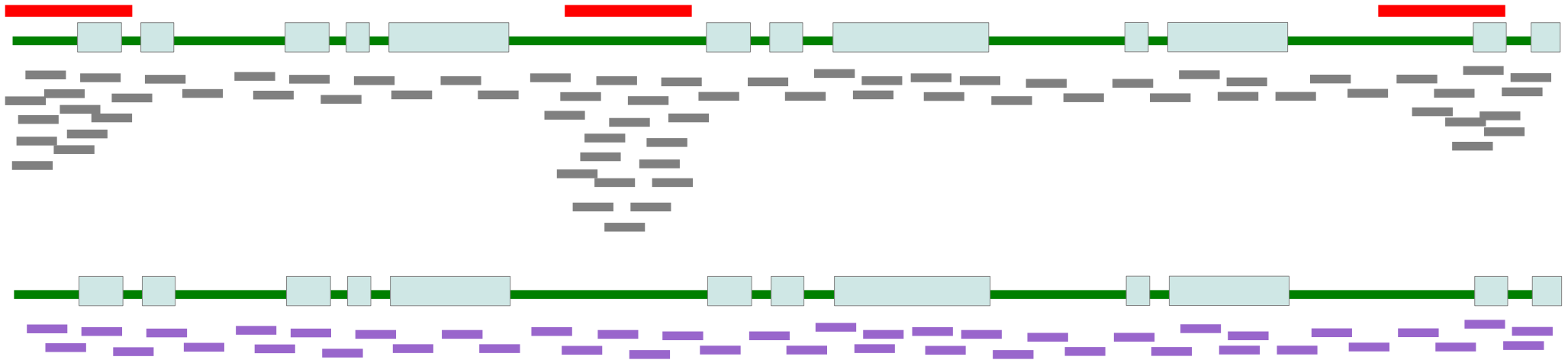


**Fastq
Control**



Determinación de picos o Peak Calling

Determinación de Picos or Peak Calling



Acceder a la carpeta results
Determinar picos usando macs2

```
macs2 callpeak  
-t ../samples/chip/chip.bam  
-c ../samples/input/input.bam  
-f BAM  
--outdir .  
-n prr5
```

Visualización de los Picos en IGV

Create .genome file

Unique identifier: atha_chr1

Descriptive name: atha_chr1

FASTA file: /home/fran/Desktop/prr5/genome/chromosome1.fa Browse

Optional

Cytoband file: Browse

Gene file: /home/fran/Desktop/prr5/annotation/chromosome1.gtf Browse

Alias file: Browse

OK Cancel

Es necesario abrir IGV, crear un genoma a partir de los ficheros fasta y gtf además de cargar los ficheros chip.bam, input.bam y narrowPeaks

- Genomes → Create .genome File (rellenar los campos como se indica arriba)
- Los ficheros input.bam y control.bam: File → Load From File
- Los ficheros de salida de MACS2 narrowPeaks y .bed: File → Load From File

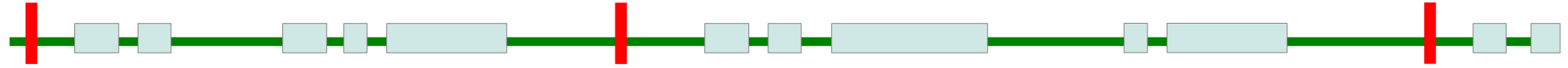


Análisis de datos de ChIP-seq

El análisis de datos de ChIP-seq para el determinar los sitios del genoma donde se une un factor de transcripción en concreto o donde se encuentra localizada un modificación de histonas se divide en los siguientes pasos:

- **Paso 1:** Preparación del espacio de trabajo.
- **Paso 2:** Análisis de calidad de la lecturas.
- **Paso 3:** Mapeado de lecturas cortas al genoma de referencia.
- **Paso 4:** Determinación de picos.
- **Paso 5:** Asociación de picos a genes diana.

ChIP-seq: Anotación de Picos o Determinación de dianas



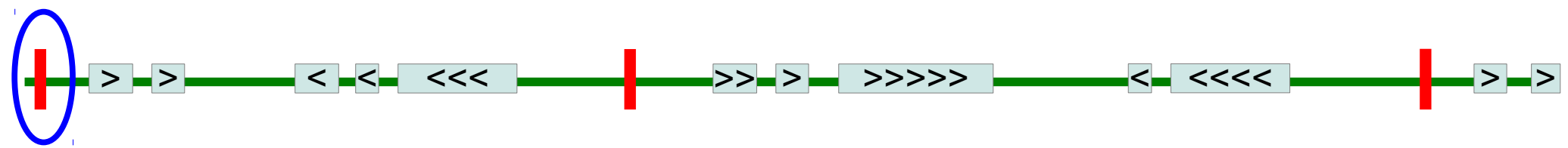
El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

ChIP-seq: Anotación de Picos o Determinación de dianas



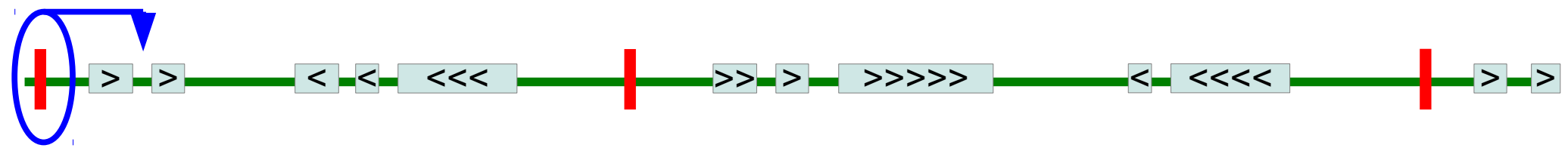
El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

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.

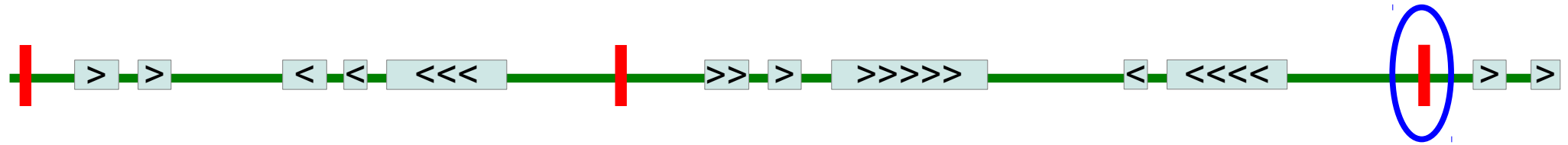
ChIP-seq: Anotación de Picos o Determinación de dianas



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)**.

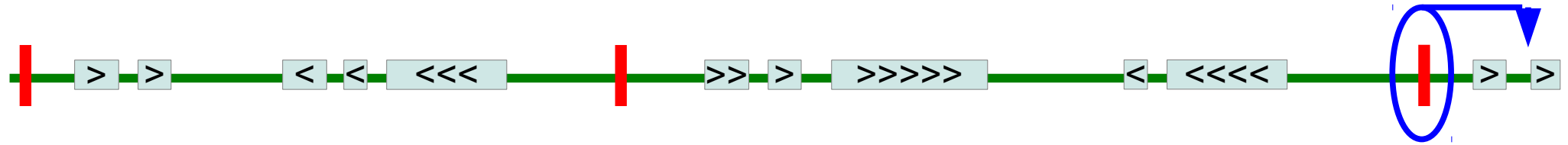
ChIP-seq: Anotación de Picos o Determinación de dianas



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)**.

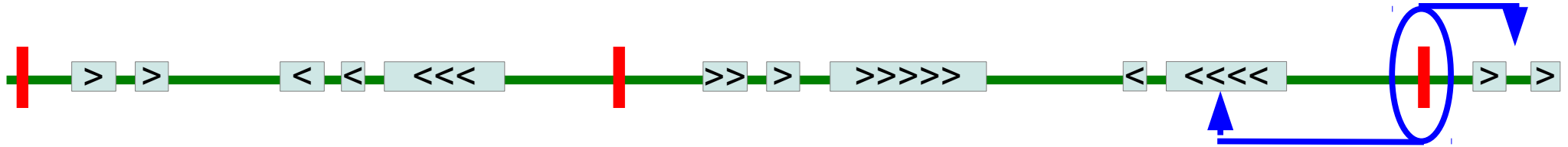
ChIP-seq: Anotación de Picos o Determinación de dianas



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)**.

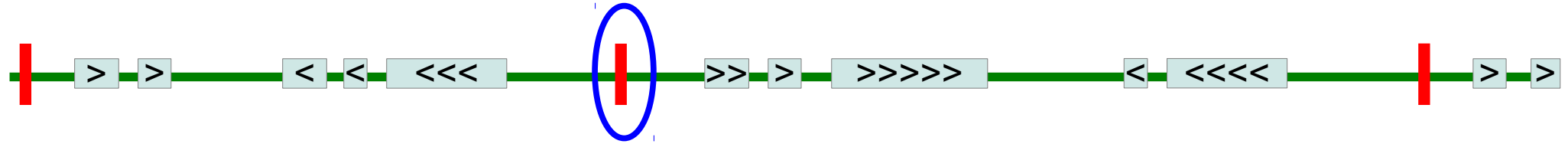
ChIP-seq: Anotación de Picos o Determinación de dianas



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)**.

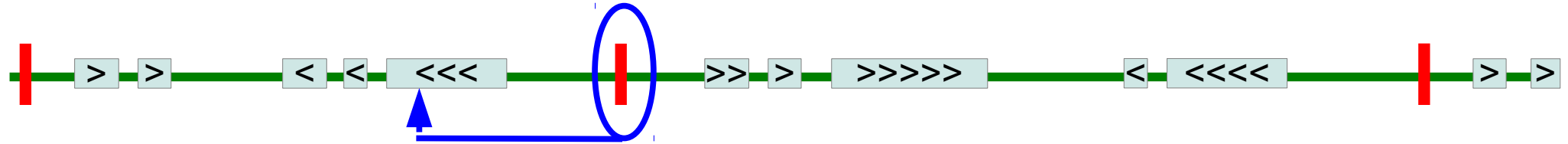
ChIP-seq: Anotación de Picos o Determinación de dianas



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)**.

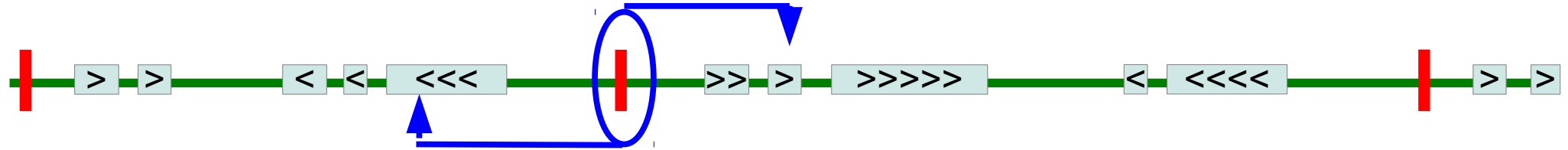
ChIP-seq: Anotación de Picos o Determinación de dianas



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)**.

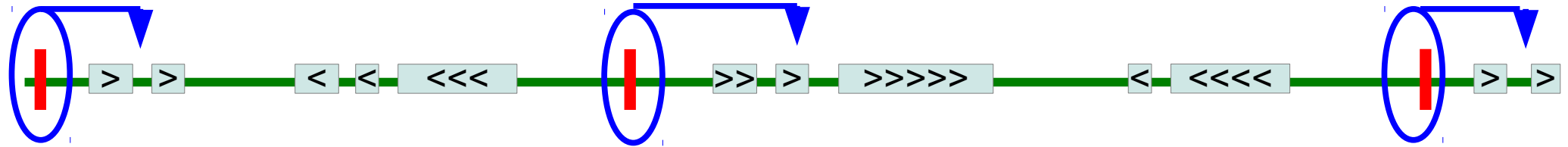
ChIP-seq: Anotación de Picos o Determinación de dianas



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)**.

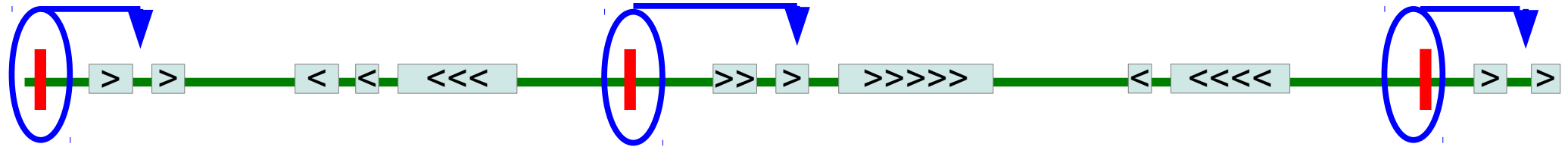
ChIP-seq: Anotación de Picos o Determinación de dianas



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)**.

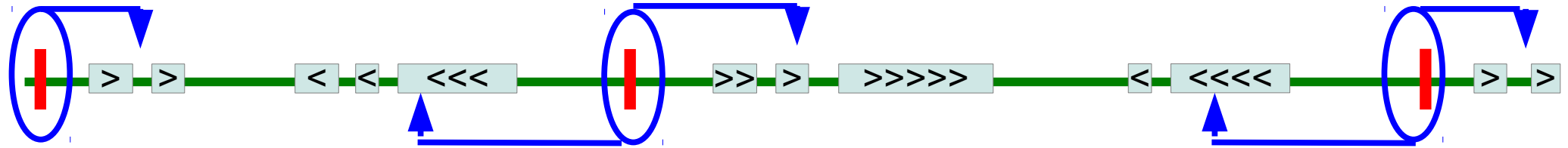
ChIP-seq: Anotación de Picos o Determinación de dianas



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)**.

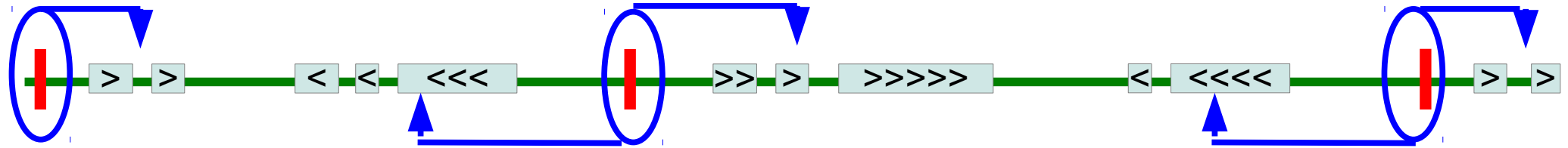
ChIP-seq: Anotación de Picos o Determinación de dianas



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

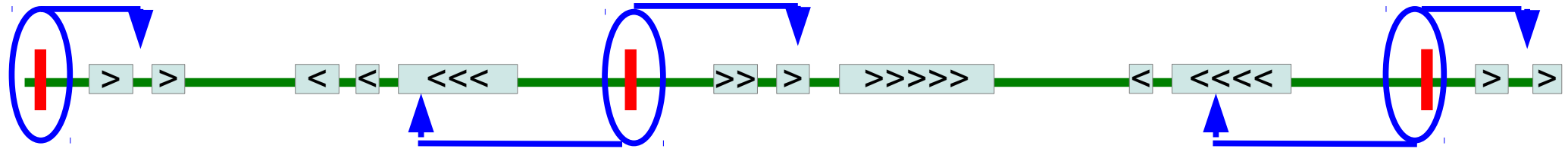


El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

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

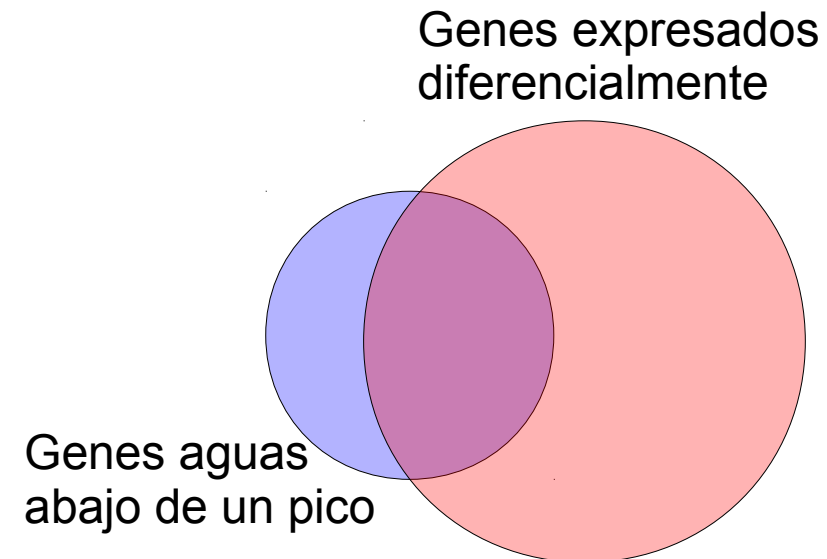
ChIP-seq: Anotación de Picos o Determinación de dianas



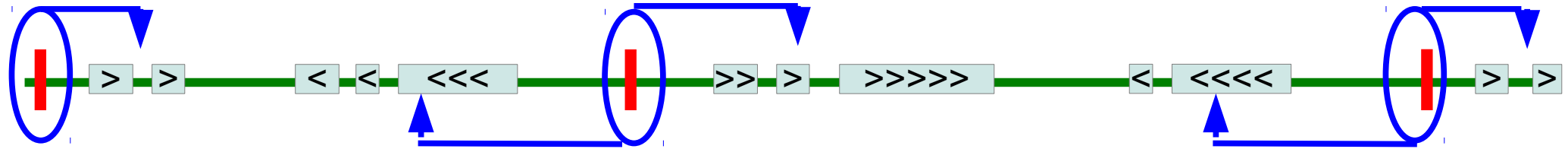
El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

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.



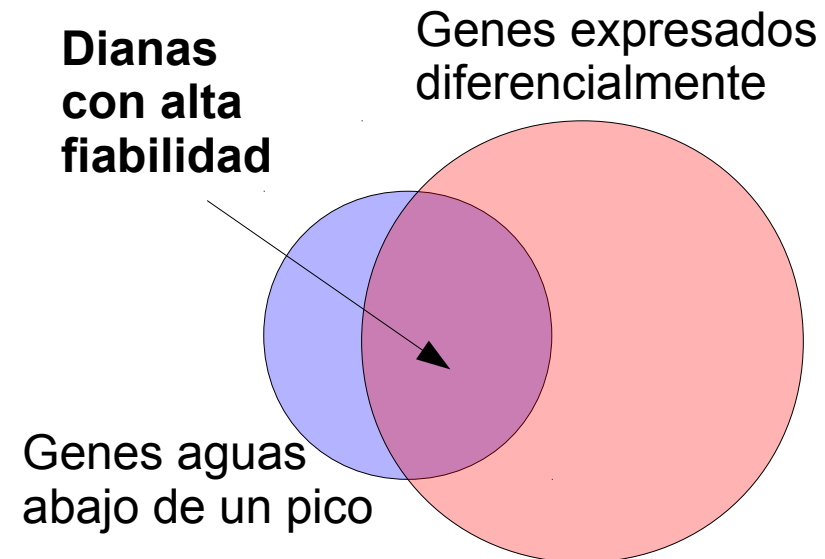
ChIP-seq: Anotación de Picos o Determinación de dianas



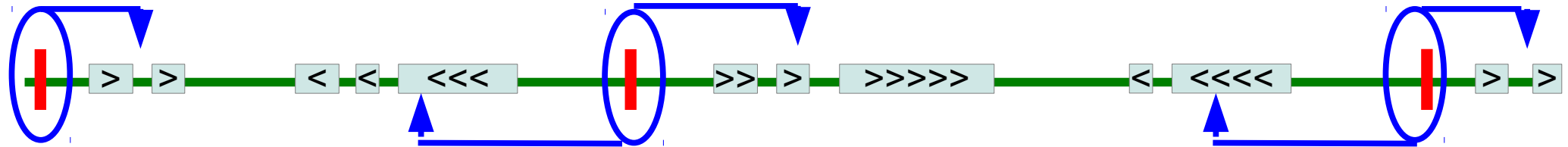
El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

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.

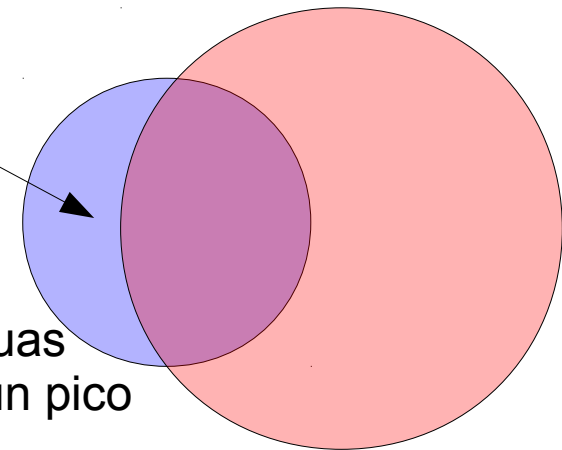
El método más extendido para anotar picos o asignar el gen diana asociado a un pico es el del gen más cercano aguas abajo o **nearest downstream gene (NDG)**.

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

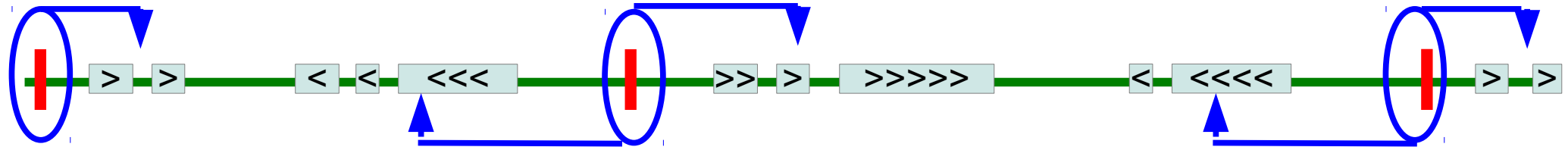
Falsos positivos o dianas que necesitan otros factores para su expresión

Genes expresados diferencialmente

Genes aguas abajo de un pico



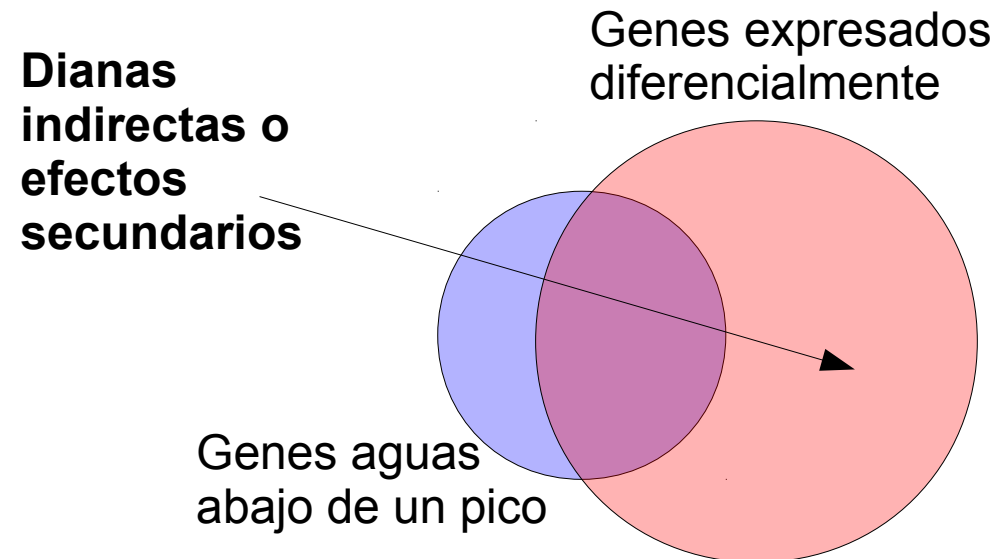
ChIP-seq: Anotación de Picos o Determinación de dianas

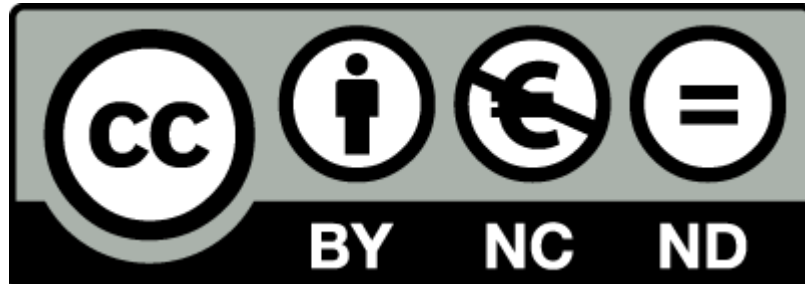


El conjunto global de genes regulados por un FT en unas condiciones específicas se denomina **reguloma**. La determinación del reguloma a partir del cistroma NO es directa.

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.





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/>.