

# iGenCo: Técnicas ómicas en el diagnóstico de enfermedades raras

Parc Científic de Barcelona, 16-17 de Novembre

**La Marató**





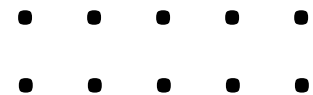
# ANÁLISIS DEL EPIGENOMA

Laura Balagué, Natàlia Carreras i  
Xavier Escribà

La Marató 

ISGlobal



# TABLA DE CONTENIDOS

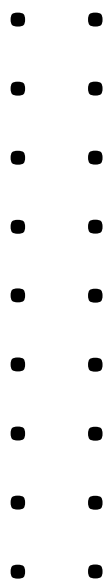
**LAS EPIMUTACIONES** **01**  
Qué son y por qué las estudiamos

**MÉTODOS DE DETECCIÓN PREVIOS** **02**  
Cuáles son y qué limitaciones tienen

**PAQUETE *EPIMUTATIONS*** **03**  
Cómo funciona

**MANOS A LA OBRA** **04**  
Caso práctico

**KAHOOT** **05**  
¿Quién se ha enterado de algo?



# LAS EPIMUTACIONES

Las **epimutaciones** se definen como alteraciones raras del patrón de metilación en un loci específico.

Pueden usarse para **resolver casos clínicos ambiguos** que no pueden ser diagnosticados mediante técnicas convencionales (secuenciación de exomas o genomas completos).

# MÉTODOS DE DETECCIÓN PREVIOS

01

## MANOVA

Análisis multivariante  
de la varianza

Aref-Eshghi et al. 2019

02

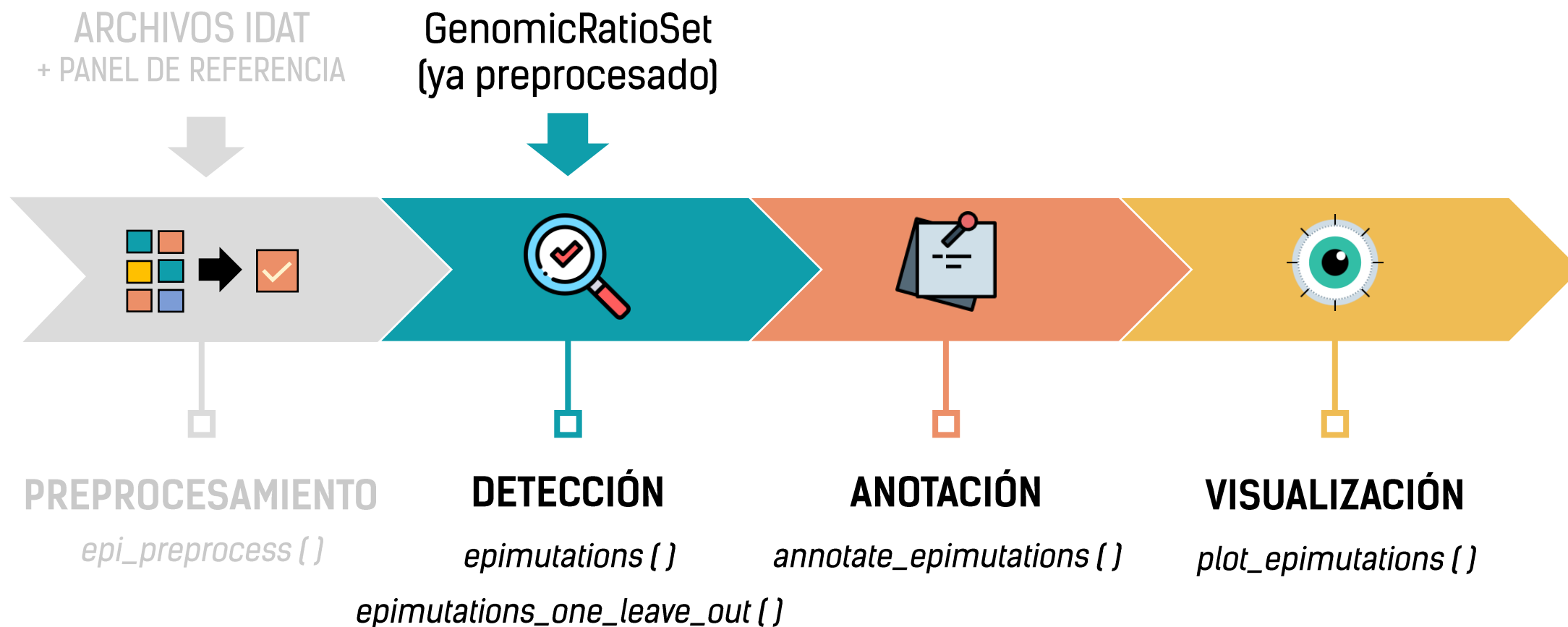
## QUANTILE

Distribución de  
cuantiles

Garg et al. 2020

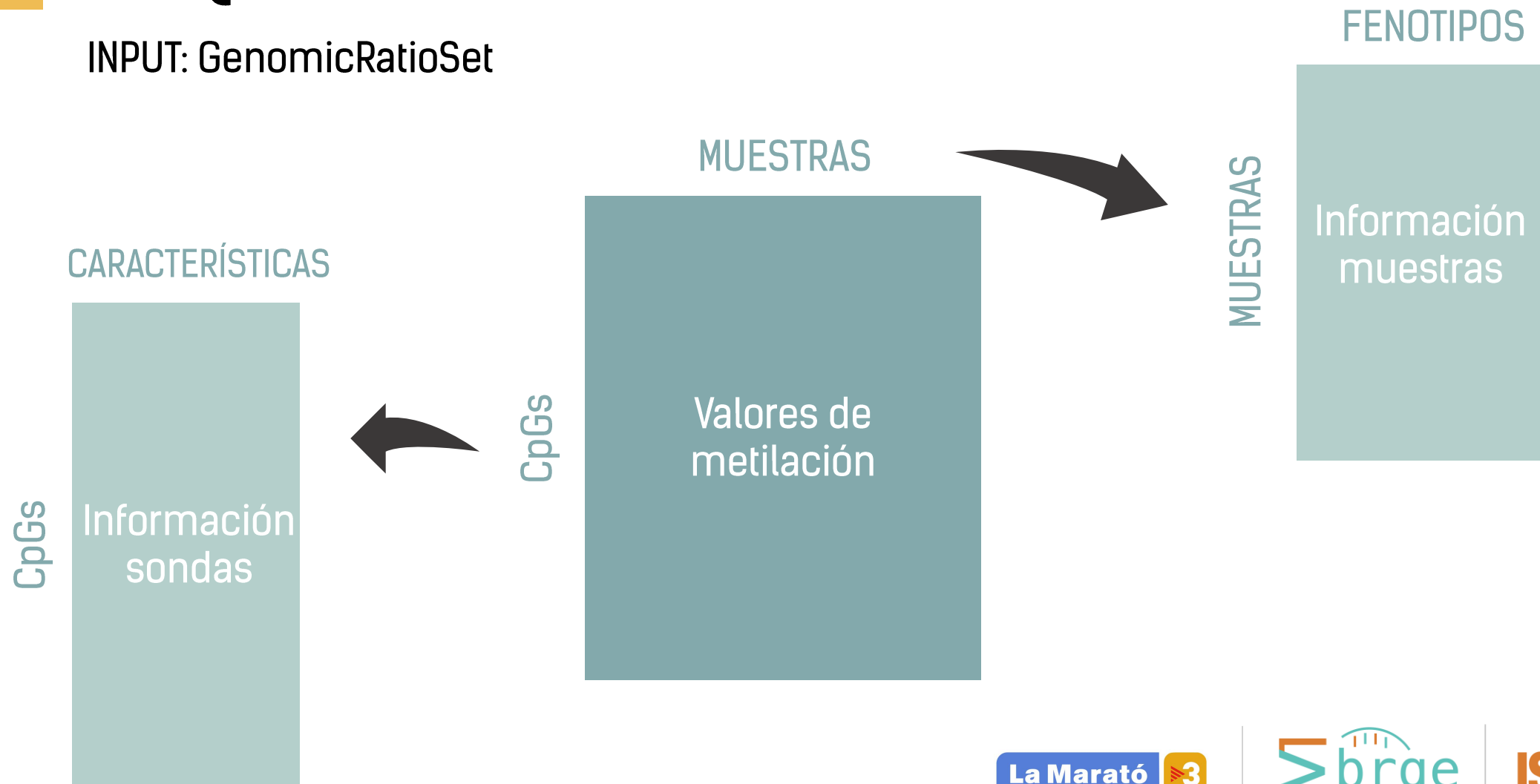
- Métodos no implementados en R
- No hay comparación de los métodos

# PAQUETE *EPIMUTACIONES*



# PAQUETE *EPIMUTACIONES*

INPUT: GenomicRatioSet



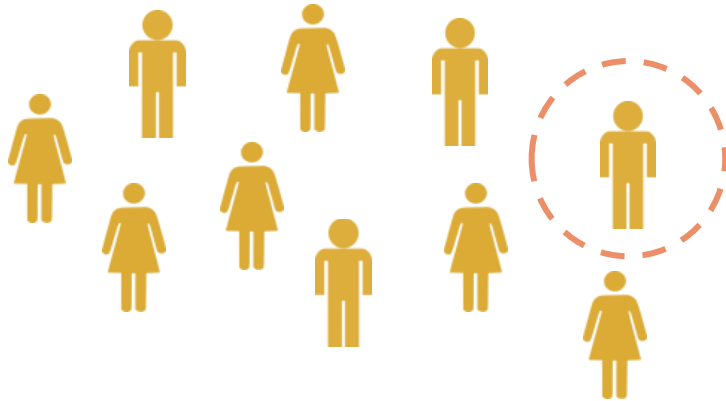


- 
- 
- 
- 

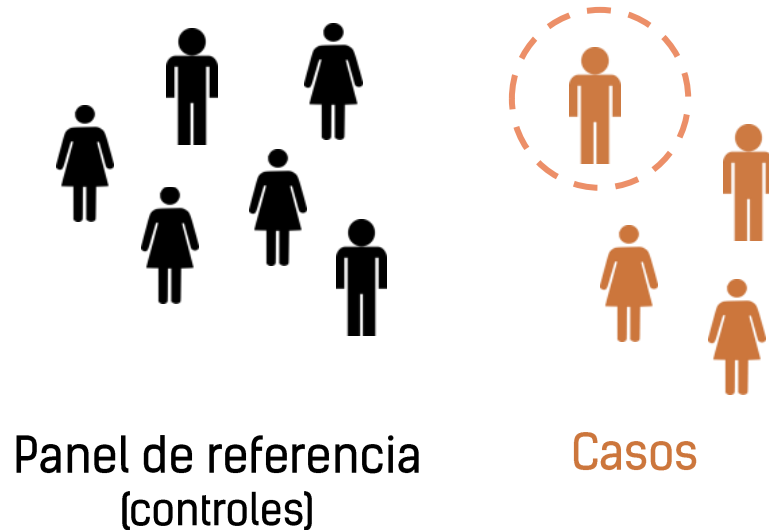
# PAQUETE *EPIMUTACIONES*

DISTRIBUCIÓN DE LOS DATOS

## A *Leave one out*



## B Caso - Control



- 
- 
- 
- 
-



# MÉTODOS DE DETECCIÓN

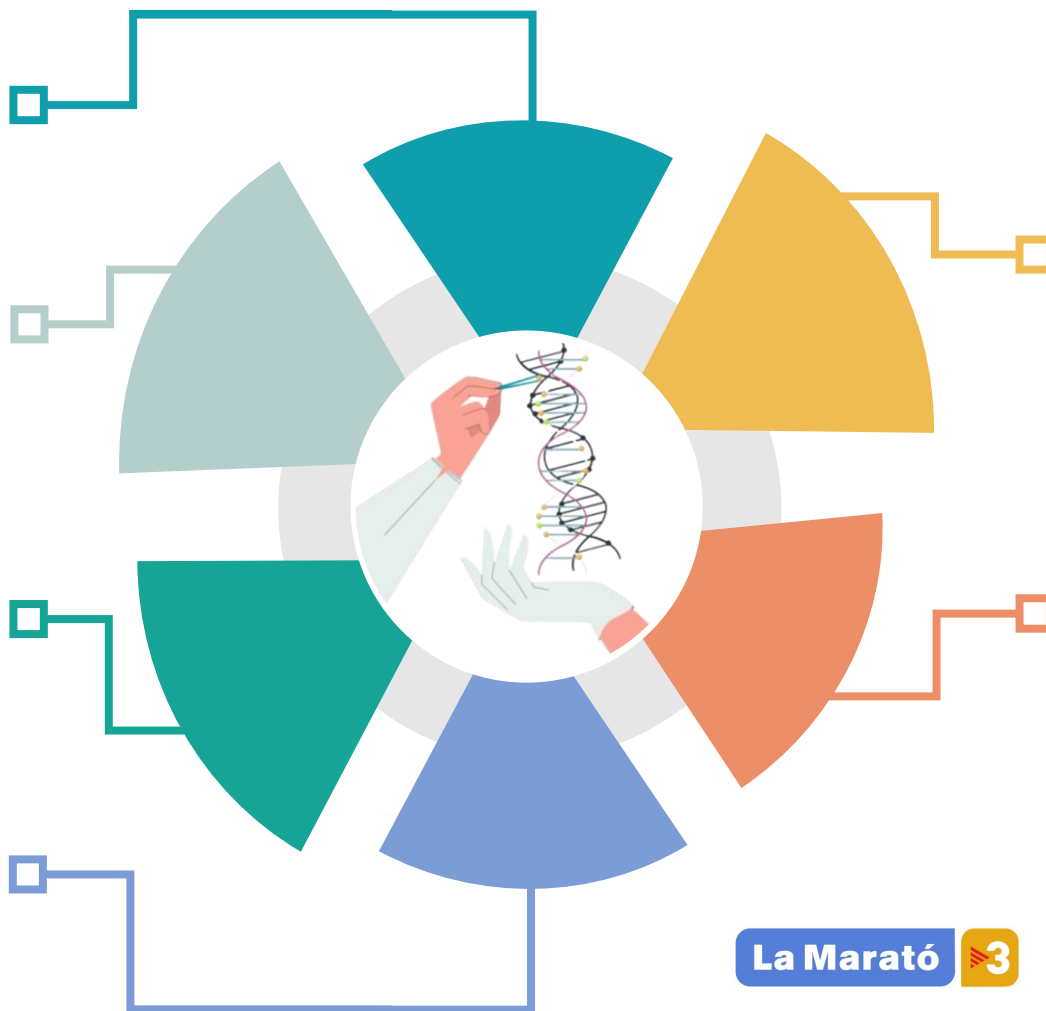
BUMPHUNTER + OUTLIER TEST

★ **MANOVA**  
Análisis multivariante  
de la varianza

**MLM**  
Modelo lineal  
multivariante

**iForest**  
*Isolation forest*

**MAH-DIST**  
Distancia de  
Mahalanobis



**QUANTILE** ★  
Distribución de  
cuantiles

**BETA**  
Distribución de  
valores betas

OUTLIERS + CLUSTERING

La Marató 3

brge

ISGlobal

# MÉTODOS DE DETECCIÓN

BUMPHUNTER + OUTLIER TEST

★ **MANOVA**  
Análisis multivariante  
de la varianza

**MLM**  
Modelo lineal  
multivariante

**iForest**  
*Isolation forest*

**MAH-DIST**  
Distancia de  
Mahalanobis



La Marató 3

LB brge

ISGlobal

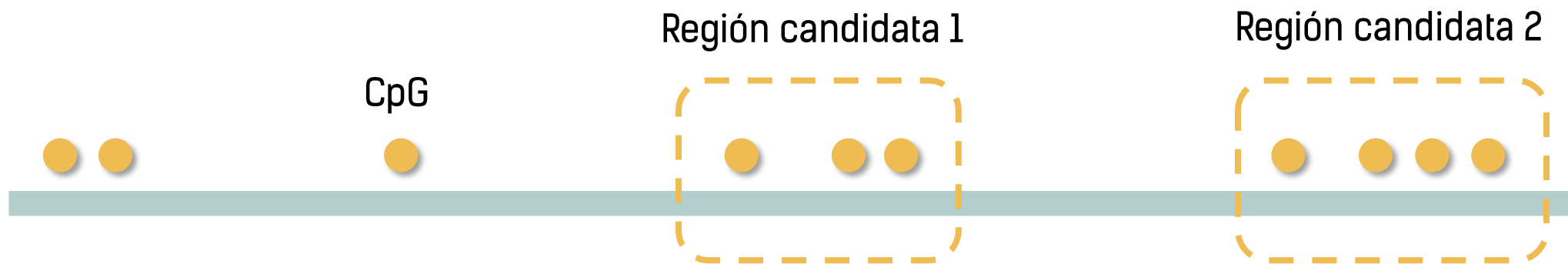
# MÉTODOS DE DETECCIÓN

BUMPHUNTER + OUTLIER TEST

01

## IDENTIFICACIÓN DE REGIONES CANDIDATAS

Regiones con al menos 3 CpGs separados por menos de 1 kb



~ 450,000 CpGs → ~ 40,000 regiones candidatas

La Marató 3

brge

ISGlobal

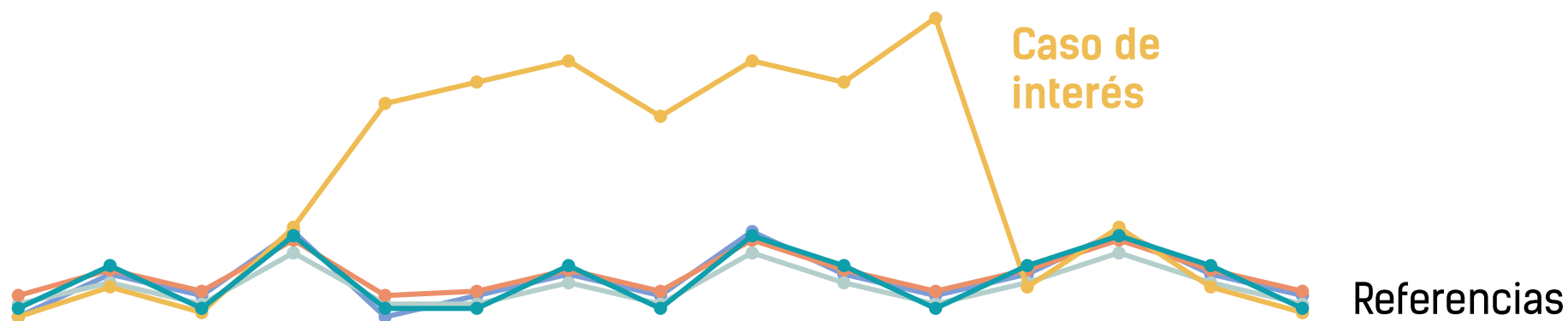
# MÉTODOS DE DETECCIÓN

BUMPHUNTER + OUTLIER TEST

02

## IDENTIFICACIÓN DE REGIONES DIFERENCIALMENTE METILADAS

Compara la muestra de interés con el panel de referencia e identifica todas aquellas regiones que están diferencialmente metiladas en la muestra.



# MÉTODOS DE DETECCIÓN

BUMPHUNTER + OUTLIER TEST

03

## OUTLIER TEST

Usando uno de los 4 métodos posibles, se evalúa cada región por separado y se identifican aquellas que son significativas.

### MANOVA

Análisis multivariante  
de la varianza

### MLM

Modelo lineal  
multivariante

### iForest

*Isolation forest*

### MAH-DIST

Distancia de  
Mahalanobis

# MANOVA Y MODELO LINEAL MULTIVARIANTE

**ANOVA:**

$$\text{cp}g_i = \alpha + \beta x_i$$

$$i = 1 \dots n$$

$x$  = caso/control

**MANOVA:**

$$c(\text{cp}g_i, \text{cp}g_{i+1}, \text{cp}g_{i+2}, \dots, \text{cp}g_{i+k}) = \alpha + \beta x$$

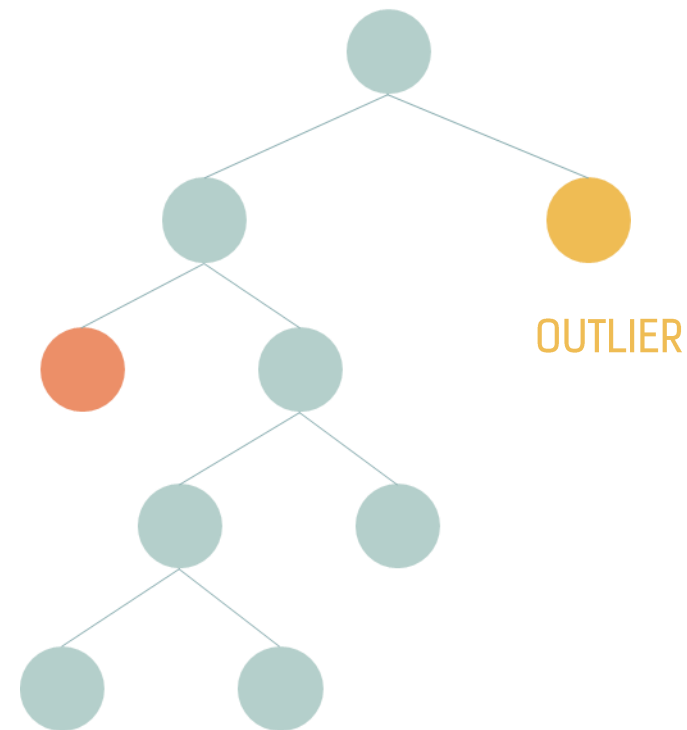
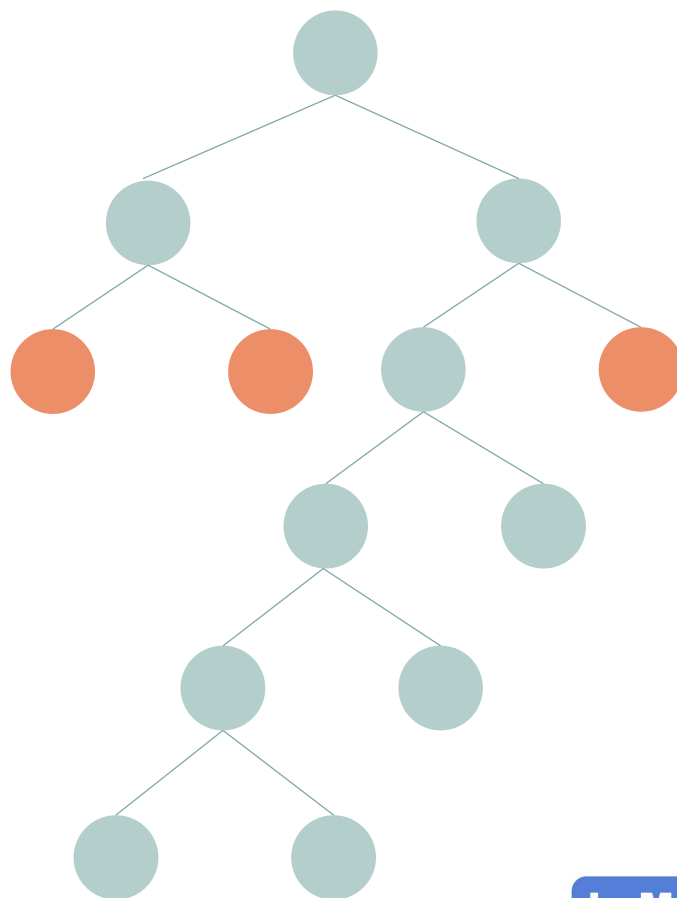
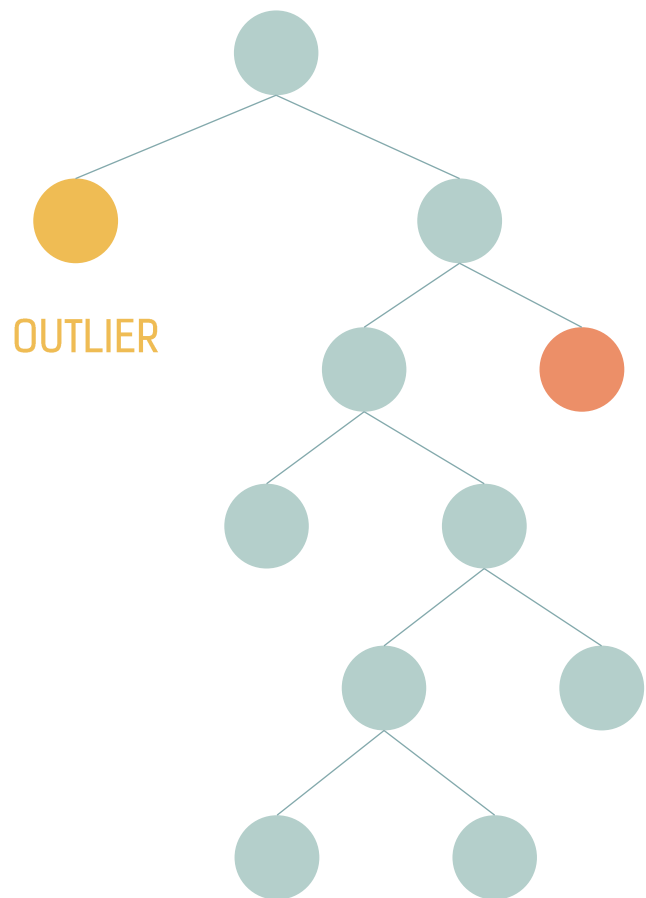


Asume distribución normal multivariada

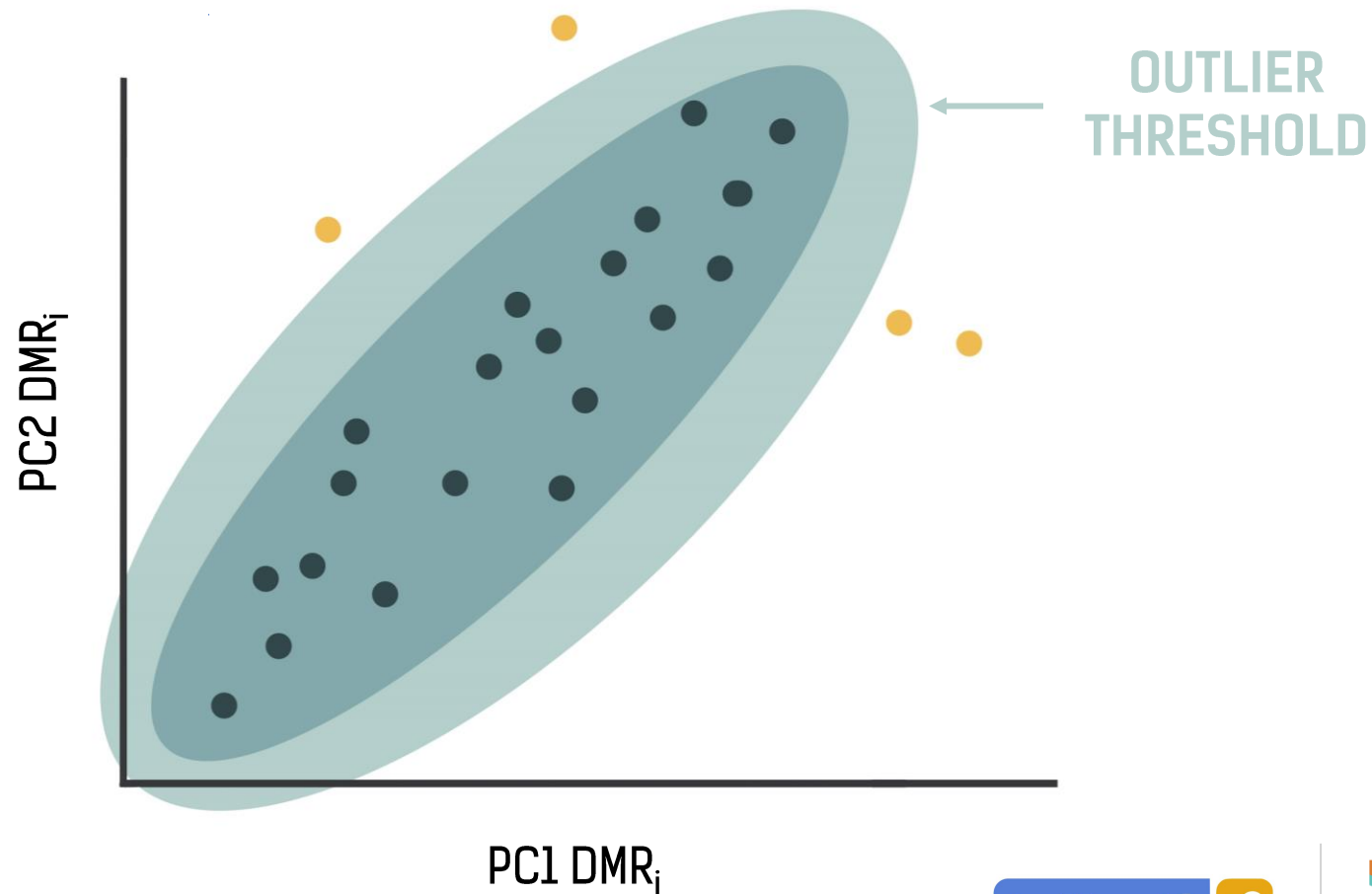


**MLM:** Método robusto y no paramétrico

# ***ISOLATION FOREST***



# DISTANCIA DE MAHALANOBIS





# MÉTODOS DE DETECCIÓN

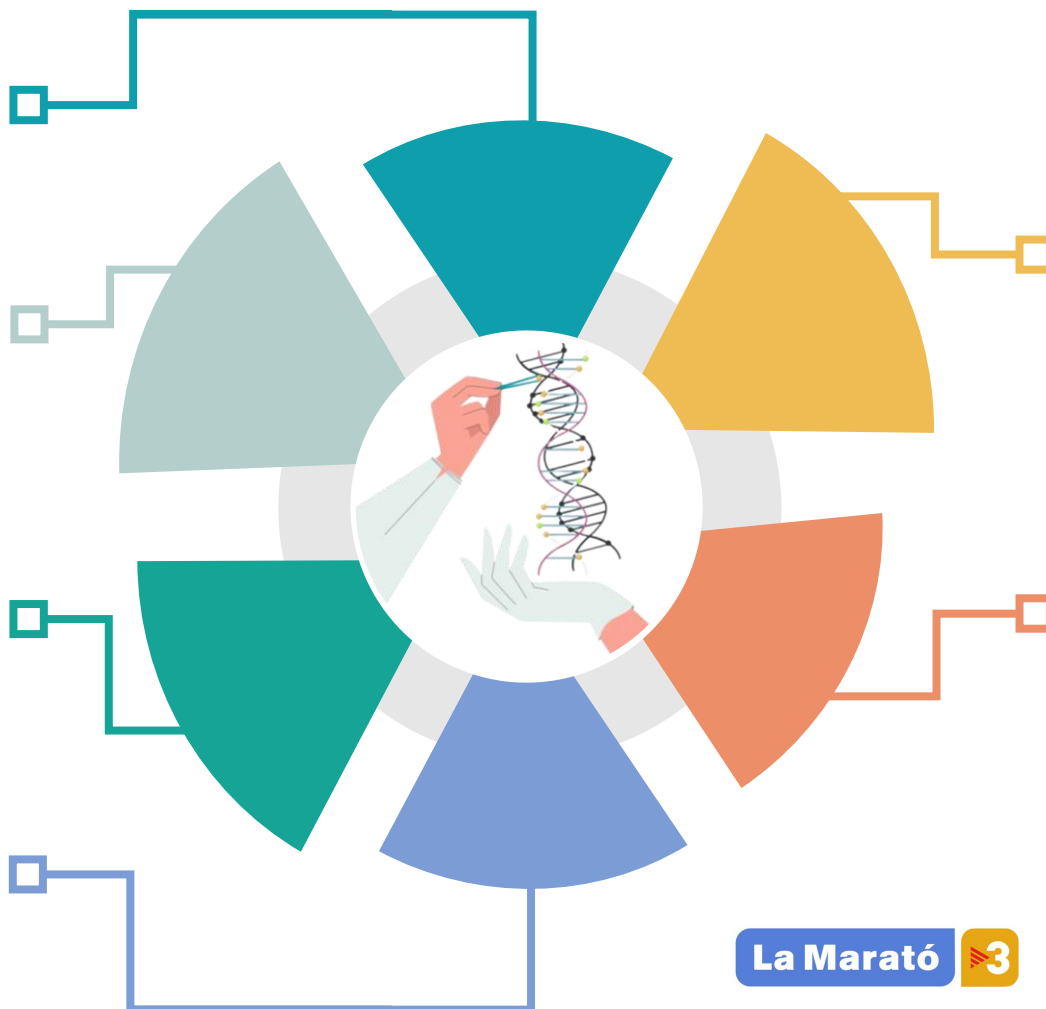
BUMPHUNTER + OUTLIER TEST

★ **MANOVA**  
Análisis multivariante  
de la varianza

**MLM**  
Modelo lineal  
multivariante

**iForest**  
*Isolation forest*

**MAH-DIST**  
Distancia de  
Mahalanobis



**QUANTILE** ★  
Distribución de  
cuantiles

**BETA**  
Distribución de  
valores betas

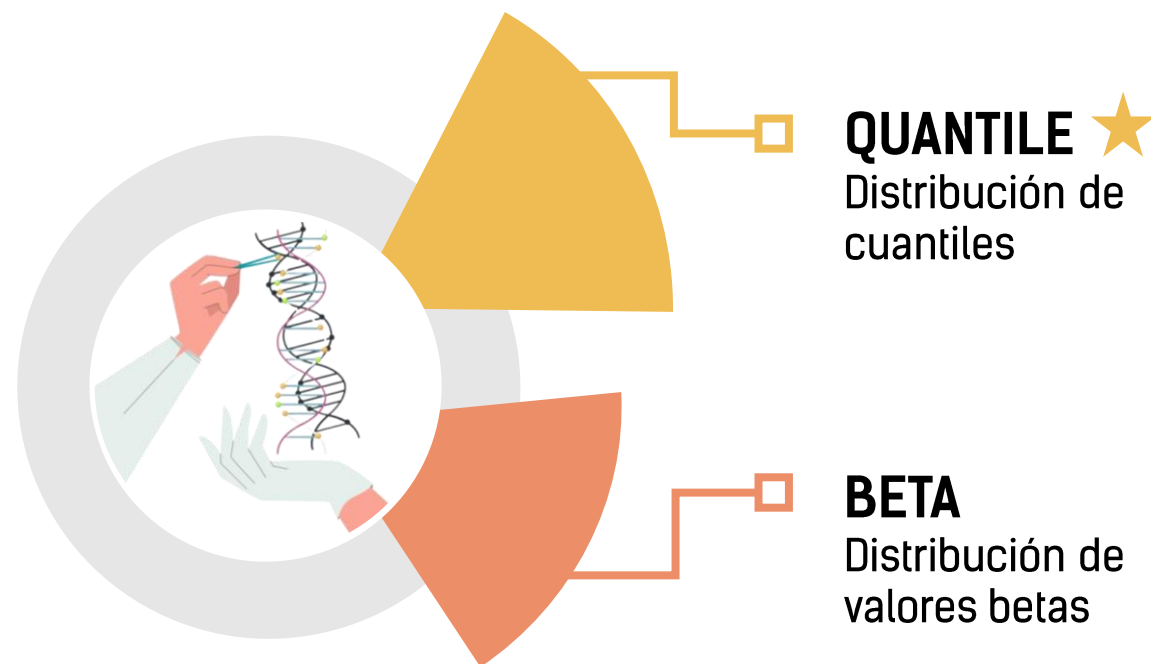
OUTLIERS + CLUSTERING

La Marató 3

brge

ISGlobal

# MÉTODOS DE DETECCIÓN



OUTLIERS + CLUSTERING

# MÉTODOS DE DETECCIÓN

## OUTLIERS + CLUSTERING

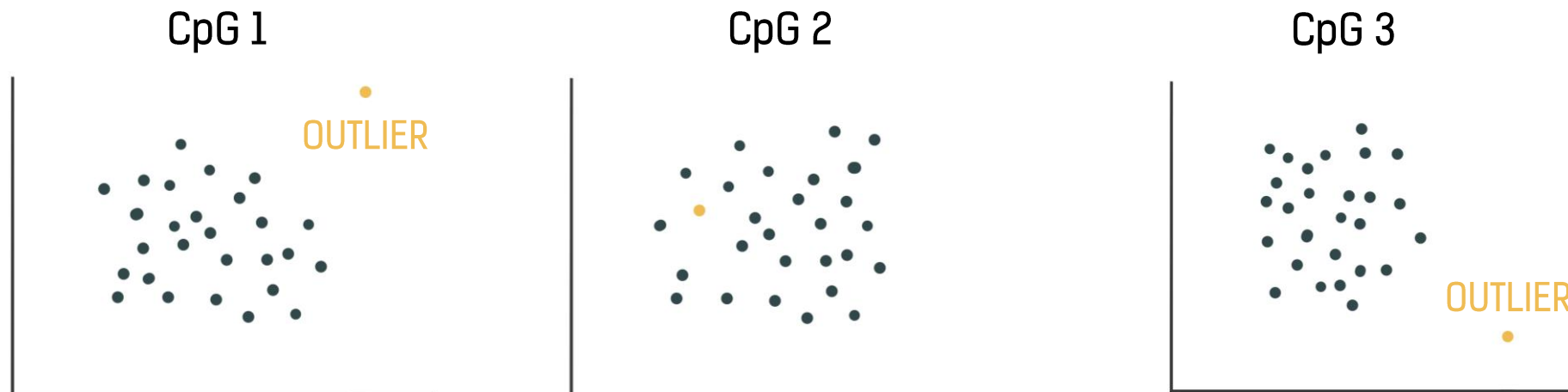
01

### IDENTIFICACIÓN DE OUTLIERS PARA CADA CpG

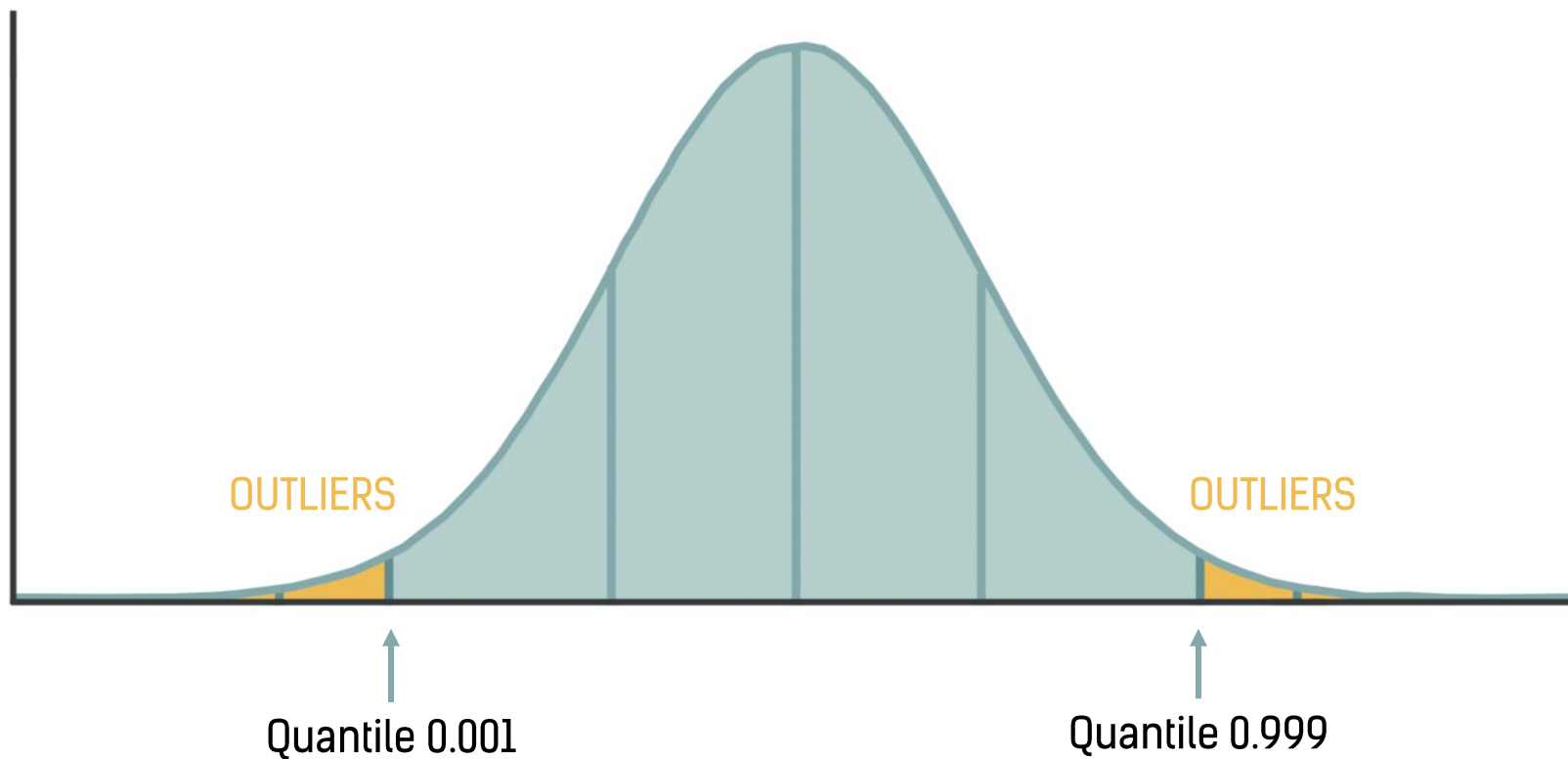
Para cada CpG, se analiza si la muestra tiene una metilación anormal comparado con el panel de referencia.

**QUANTILE**  
Distribución de  
cuantiles

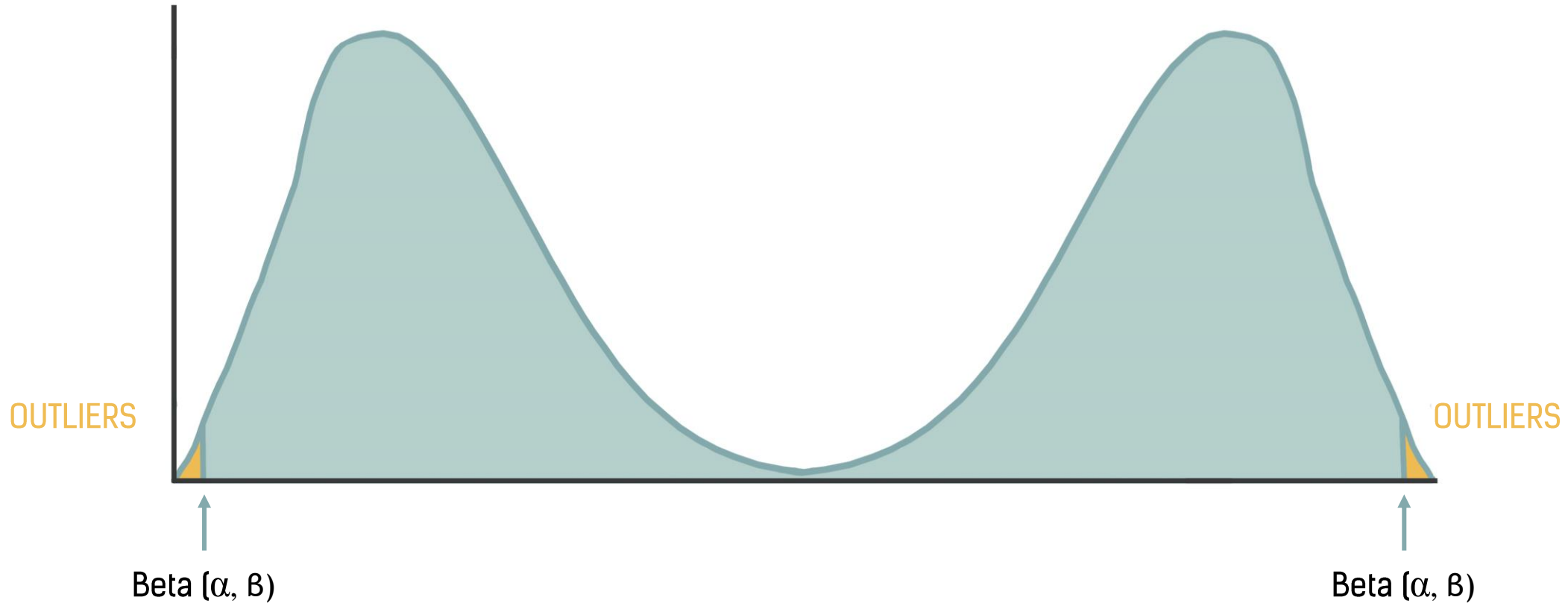
**BETA**  
Distribución de  
valores betas



# DISTRIBUCIÓN DE CUANTILES



# DISTRIBUCIÓN DE VALORES BETAS





# MÉTODOS DE DETECCIÓN

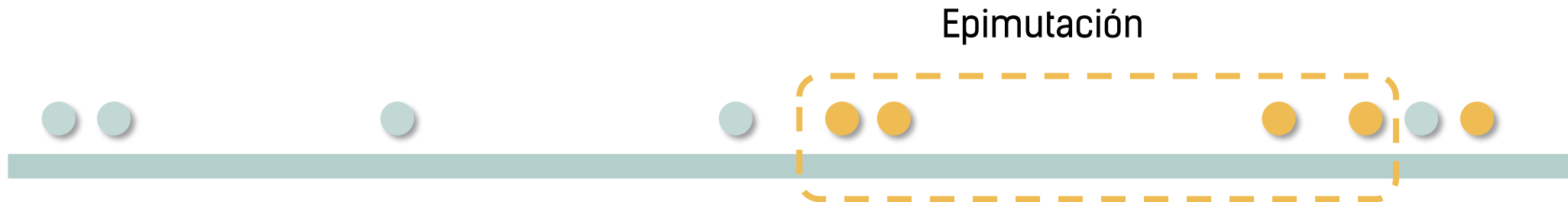
OUTLIERS + CLUSTERING

02

## IDENTIFICACIÓN DE EPIMUTACIONES

Cuando encontramos 3 o más CpGs *outliers* consecutivos con menos de 1kb entre ellos, se considera una epimutación.

- OUTLIER
- METILACIÓN NORMAL

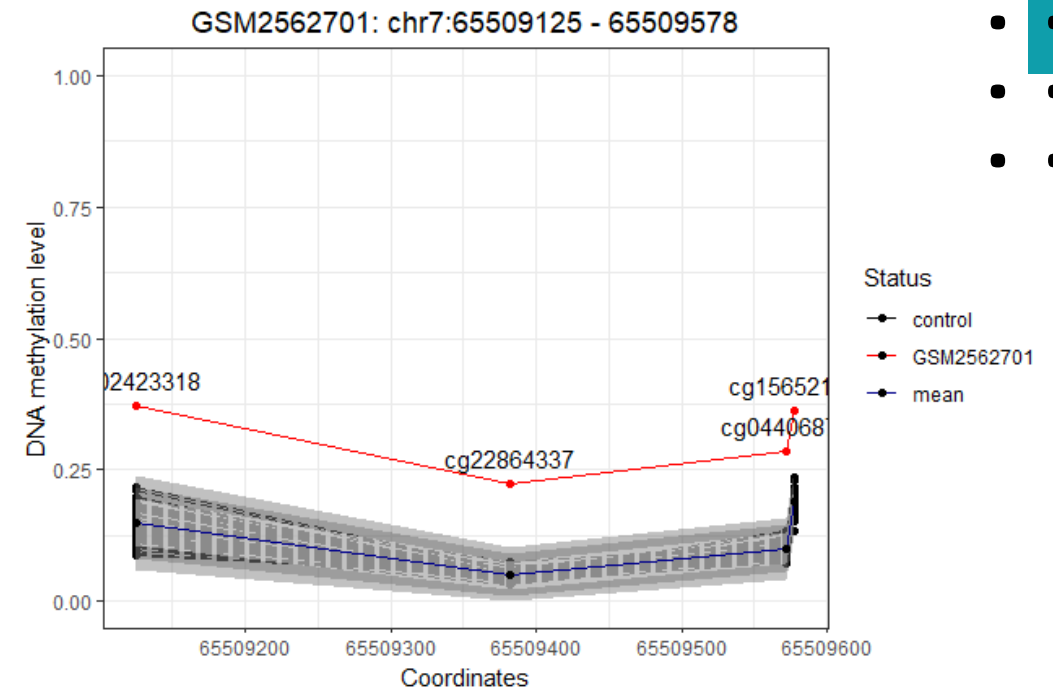


# PAQUETE *EPIMUTACIONES*

## OUTPUT

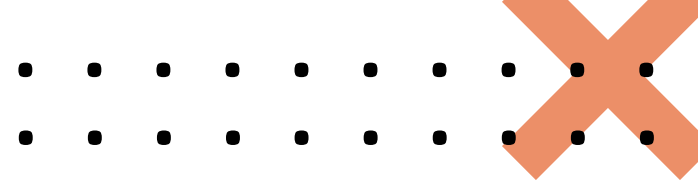
| Sample     | Chr   | Start    | End      | Size | Num. CpGs |
|------------|-------|----------|----------|------|-----------|
| GSM256269  | chr19 | 12777736 | 12777903 | 167  | 4         |
| CLT1236766 | chr8  | 89554    | 89634    | 80   | 3         |

| CpGs  | Direction        | P value               | P value adjusted      | GENCODE                         |
|---|------------------|-----------------------|-----------------------|---------------------------------|
| cg20791841,cg25267526,<br>cg03641858,cg25441478 | hypermethylation | $3.44 \cdot 10^{-9}$  | $2.17 \cdot 10^{-6}$  | <i>MSRA, TNKS, RP1L1, RP1L1</i> |
| cg65267526,cg13641858,<br>cg890374639           | Hypomethylation  | $2.03 \cdot 10^{-13}$ | $6.11 \cdot 10^{-11}$ | <i>IL2, ADAD1, IL2</i>          |

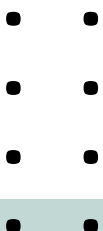




# CONSIDERACIONES



- El método que da mejores resultados es **Quantile**.
- Los **controles** deberían estar hechos con el **mismo array** y ser del **mismo sexo** que la muestra de interés.
- En la medida de lo posible, usar el método **caso-control** y no One-leave-Out.
- Para priorizar epimutaciones, es recomendable comprobar si la epimutación afecta la **expresión** de algún gen cercano.







**¡MANOS A  
LA OBRA!**

La Marató 

 **brge**

**ISGlobal**



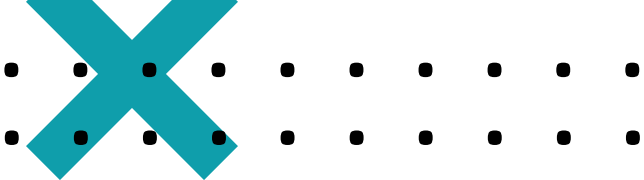
**01 Copiar el siguiente link en vuestro PC:**

<https://epimutaciones.isglobal.org/>

**02 Entrar con el usuario y contraseña:**

User: idX (X: a cada uno le asignaremos un número)

Password: 123



epimutationsShiny

Home

Data selection

Epimutations

Acknowledgements

Available datasets

- 4 case samples IDAT files (GEO: GSE131350) + `reference_panel` : a `RGChannelSet` class object containing 22 healthy individuals (GEO: GSE127824). This dataset is to test the pre-process functionality.
- `reference_panel` : a `GenomicRatioSet` class object containing 22 healthy individuals (GEO: GSE127824). This dataset is to test the one leave out epimutations functionality.
- `methy` : a `GenomicRatioSet` object which includes 49 controls (GEO: GSE104812) and 3 cases (GEO: GSE97362).

Select dataset

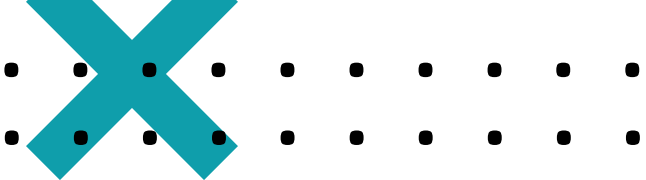
Datasets

IDAT files

Load dataset

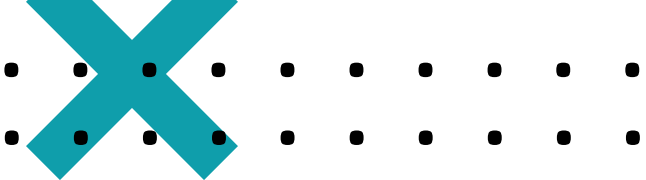
**03** Copia el siguiente link para la Shiny:

[epishiny.isglobal.org](http://epishiny.isglobal.org)



## EJERCICIO (con R o Shiny)

- 01 ¿Cuáles son las 3 epimutaciones más significativas con el **método mlm**?
- 02 ¿Cuántos CpGs tiene cada epimutación?
- 03 Crea un **plot** para la epimutación más significativa.



## EJERCICIO (con R o Shiny)

```
epi_mvo <- epimutations (case_samples,  
                          control_samples,  
                          method = "mlm")
```



# EJERCICIO (con R o Shiny)

**01** ¿Cuáles son las 3 epimutaciones más significativas con el **método mlm**?



|   | epi_id     | sample     | chromosome | start    | end      | sz  | cpg_n | cpg_ids                                | outlier_direction | pvalue       | adj_pvalue   |
|---|------------|------------|------------|----------|----------|-----|-------|--|-------------------|--------------|--------------|
| 1 | epi_mlm_2  | GSM2562699 | chr19      | 12777736 | 12777903 | 167 | 7     | cg20791841,cg25267526,cg03641858,cg254 | hypermethylation  | 1.000000e-14 | 6.600000e-13 |
| 2 | epi_mlm_7  | GSM2562699 | chr7       | 1080558  | 1081068  | 510 | 4     | cg03916490,cg12581298,cg22785556,cg243 | hypomethylation   | 1.021405e-14 | 6.639134e-13 |
| 3 | epi_mlm_46 | GSM2562701 | chr7       | 65509125 | 65509578 | 453 | 4     | cg02423318,cg22864337,cg04406873,cg156 | hypermethylation  | 1.000000e-14 | 8.600000e-13 |



# EJERCICIO (con R o Shiny)

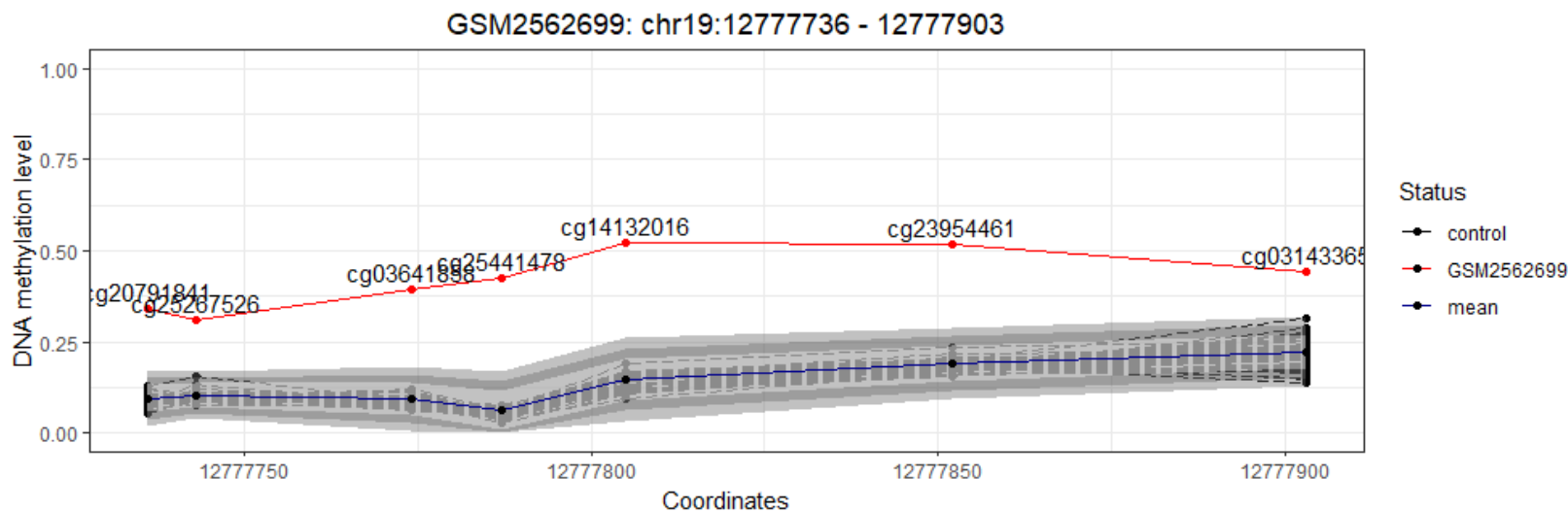
## 02 ¿Cuántos CpGs tiene cada epimutación?



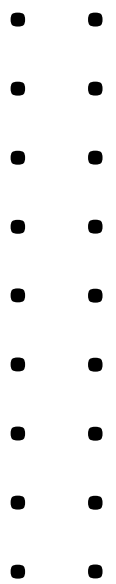
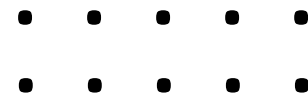
|   | epi_id     | sample     | chromosome | start    | end      | sz  | cpg_n | cpg_ids                                | outlier_direction | pvalue       | adj_pvalue   |
|---|------------|------------|------------|----------|----------|-----|-------|--|-------------------|--------------|--------------|
| 1 | epi_mlm_2  | GSM2562699 | chr19      | 12777736 | 12777903 | 167 | 7     | cg20791841,cg25267526,cg03641858,cg254 | hypermethylation  | 1.000000e-14 | 6.600000e-13 |
| 2 | epi_mlm_7  | GSM2562699 | chr7       | 1080558  | 1081068  | 510 | 4     | cg03916490,cg12581298,cg22785556,cg243 | hypomethylation   | 1.021405e-14 | 6.639134e-13 |
| 3 | epi_mlm_46 | GSM2562701 | chr7       | 65509125 | 65509578 | 453 | 4     | cg02423318,cg22864337,cg04406873,cg156 | hypermethylation  | 1.000000e-14 | 8.600000e-13 |

# EJERCICIO (con R o Shiny)

03 Crea un **plot** para la epimutación más significativa.







# KAHOOT

<https://kahoot.it/>





# MATERIAL CURSO



[https://isglobal-brge.github.io/course\\_epimutations/](https://isglobal-brge.github.io/course_epimutations/)