



**Sant Joan de Déu**  
Institut de Recerca



UNIVERSITAT DE  
BARCELONA



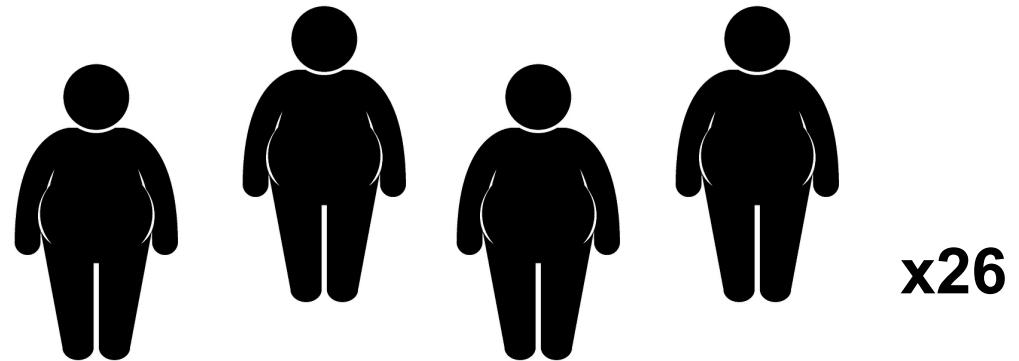
# Prediction model of response to a certain treatment in childhood obesity

**Pol Castellano Escuder**

Máster universitario en Bioinformática y Bioestadística  
Área: Estadística y Bioinformática

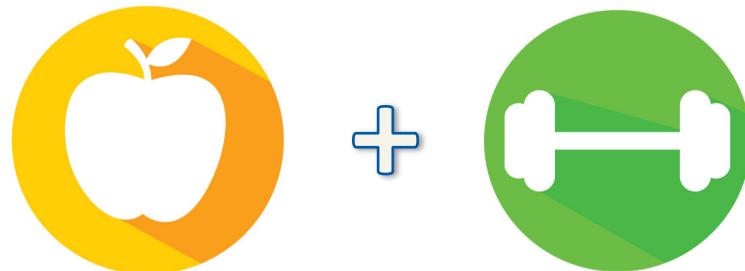
**Profesor responsable:** Esteban Vegas Lozano

24 de Mayo de 2017

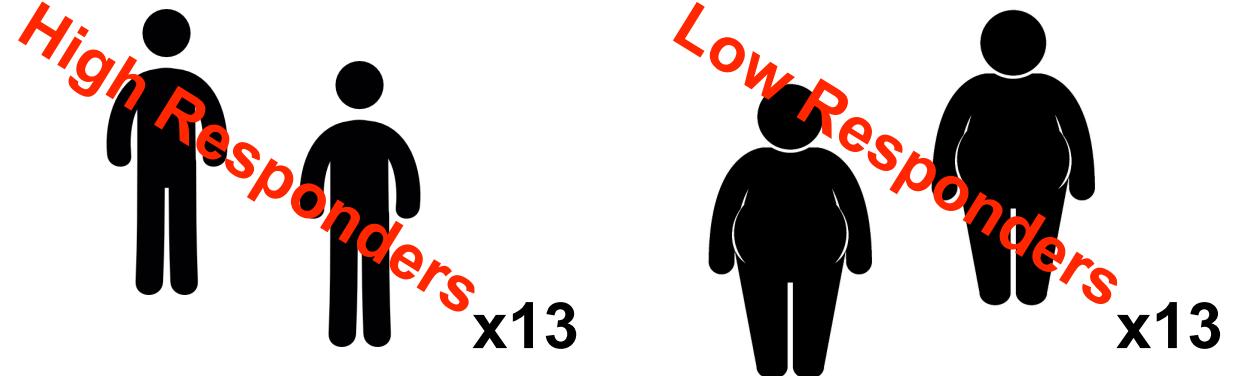


### Criterio de clasificación:

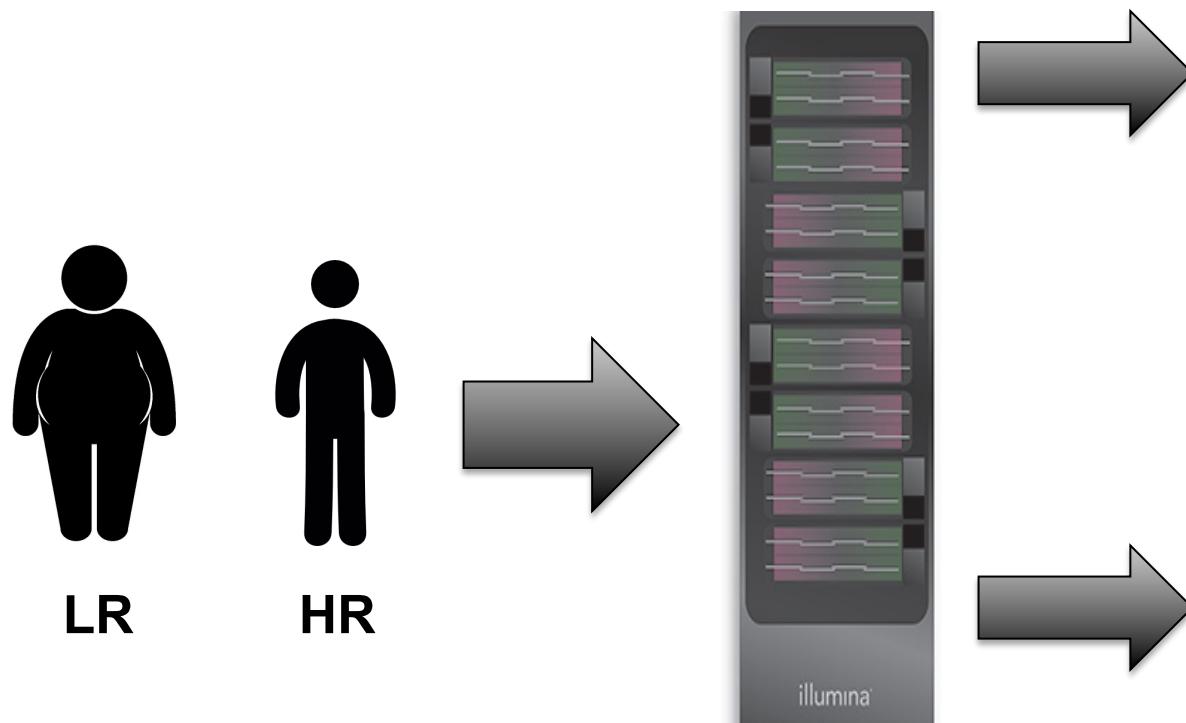
- $DzBMI < -0.3 = \text{HR}$
- $DzBMI > -0.3 = \text{LR}$



6 months



## *Output*



Extraemos muestras de  
sangre periférica

Illumina EPIC 850K  
methylation array

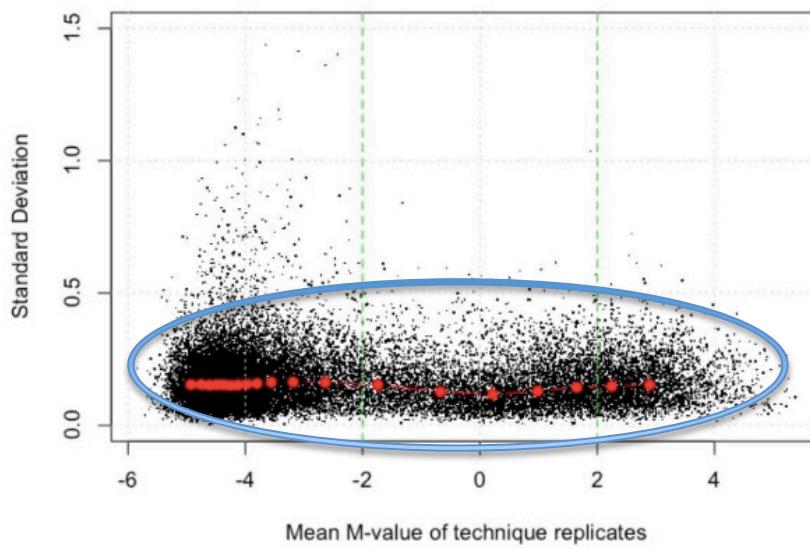
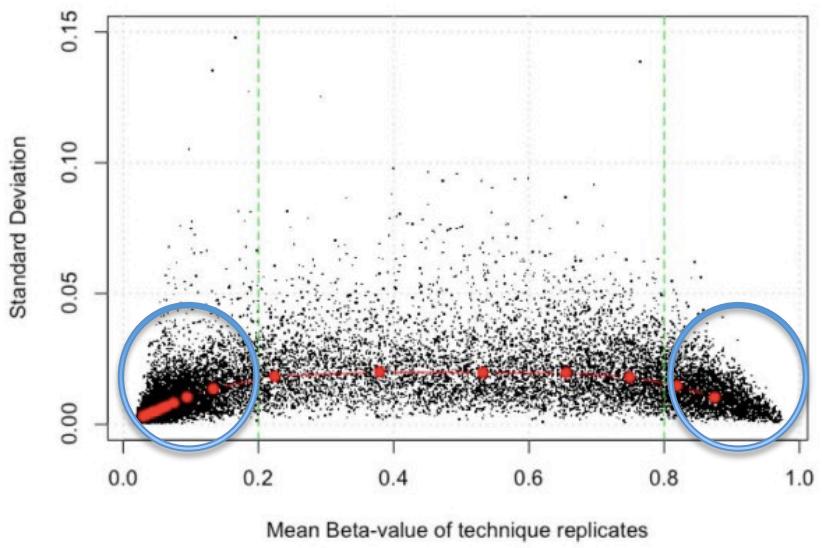
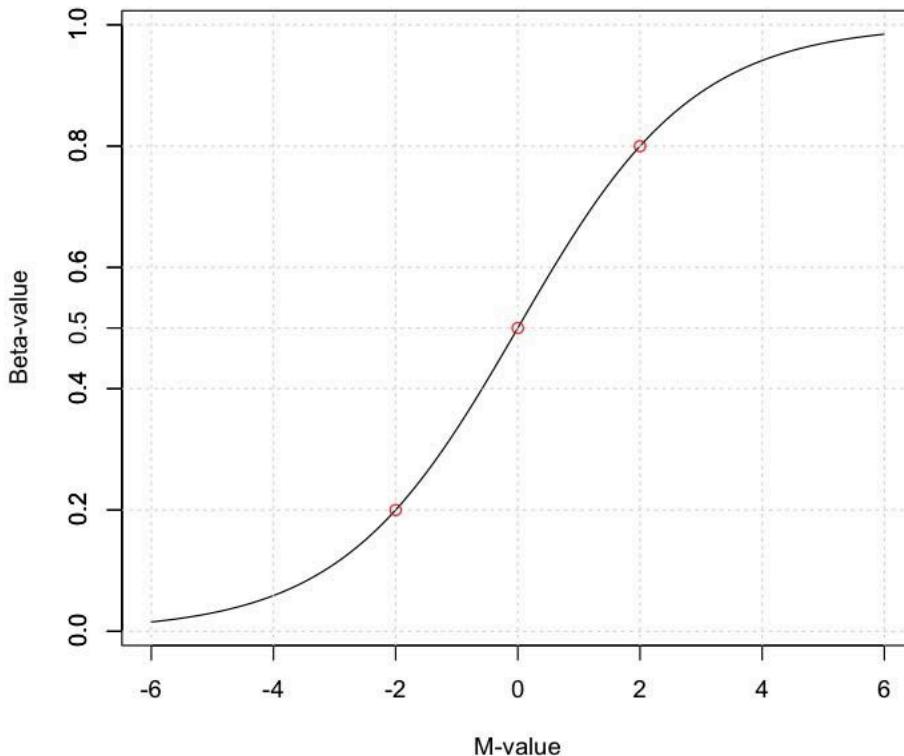
### Valor Beta:

- Valor heteroecdástico
- Rango de 0 a 1

### Valor M:

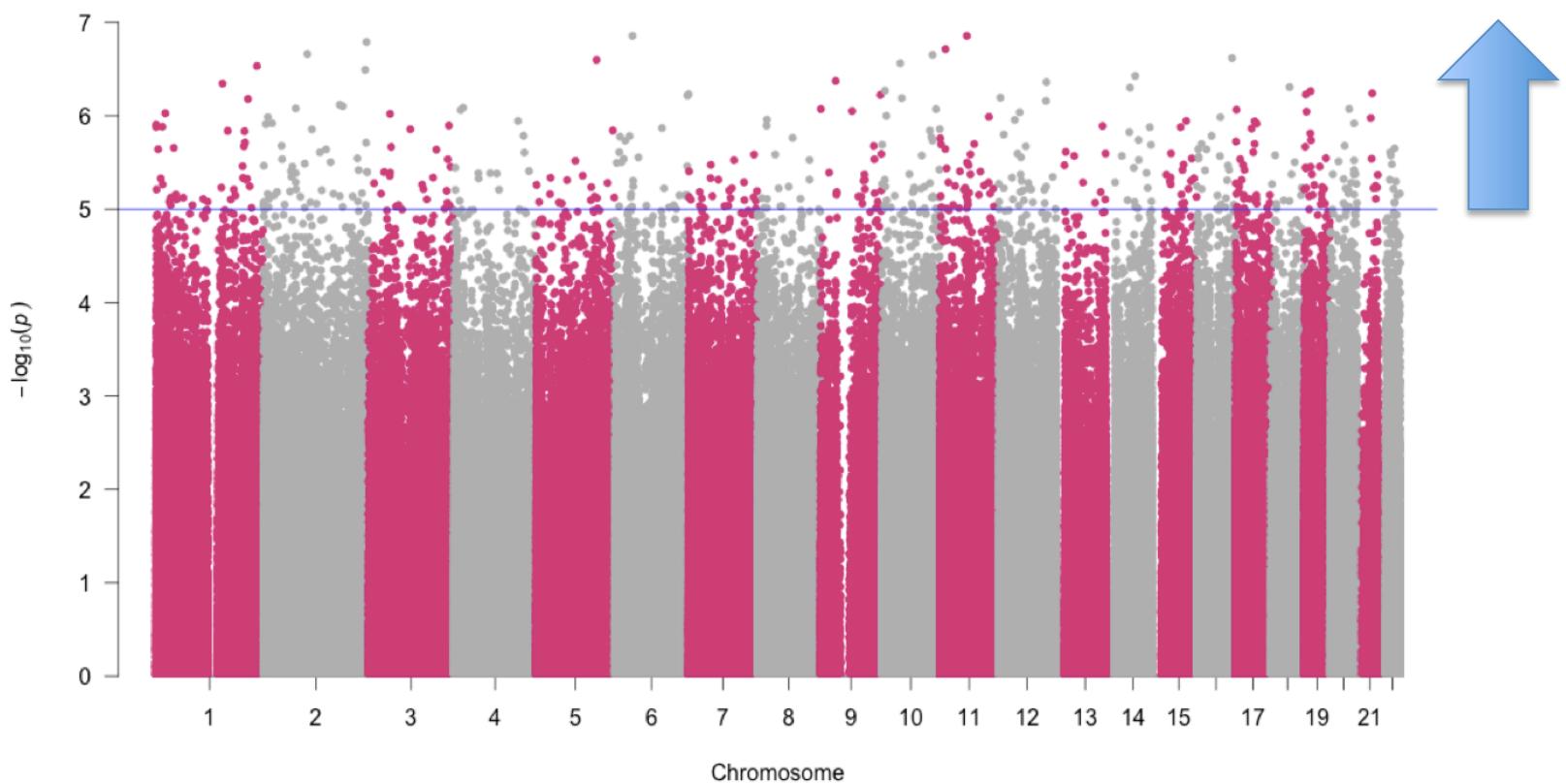
- Valor homoecedástico
- Rango de -6 a 6

Du, P., Zhang, X., Huang, C.C. et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformat. 2010;11:587.



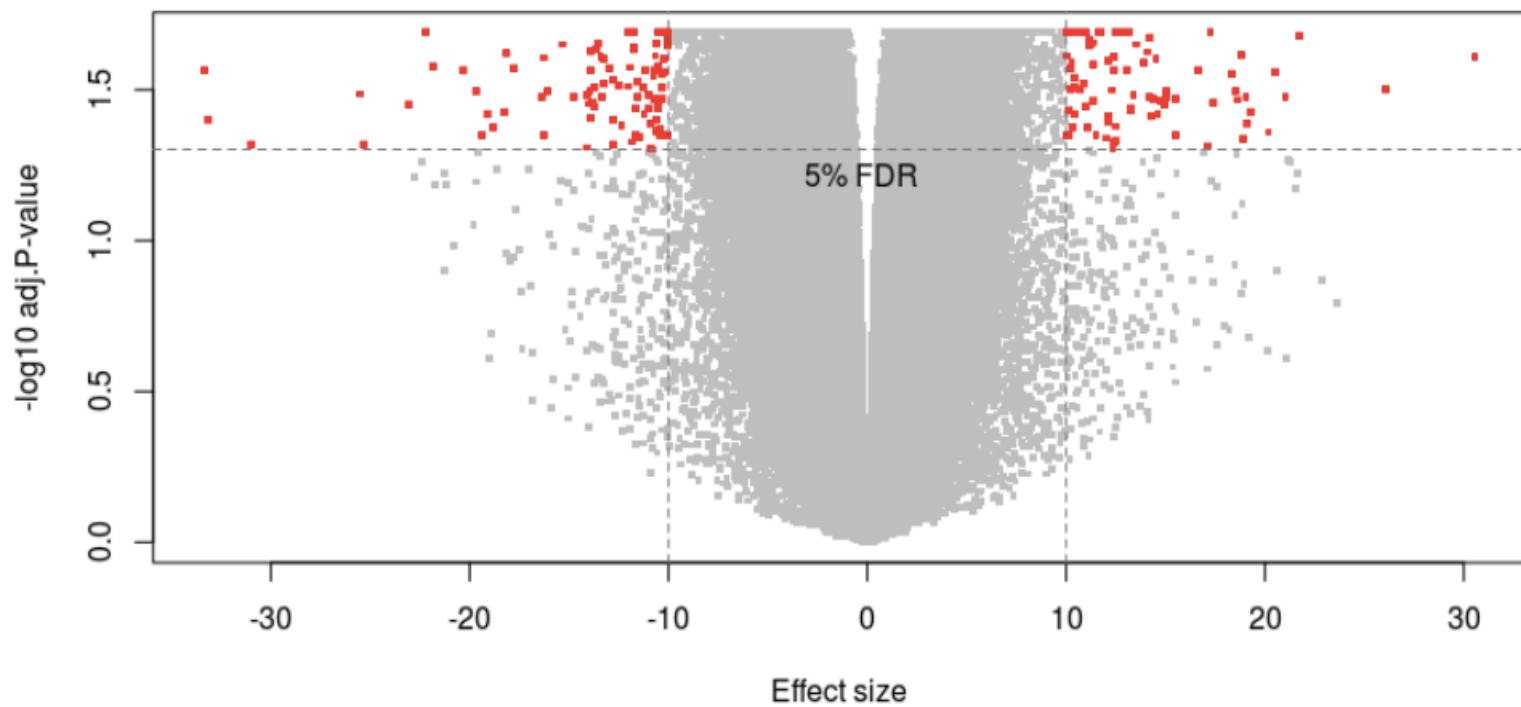
# Manhattan plot (788.373 CpG)

$P\text{-val} < 0.05$



# Volcano plot

- *Adj. P-val < 0.05*
- *Effect size 10%*



# Obtención de los CpG significativos

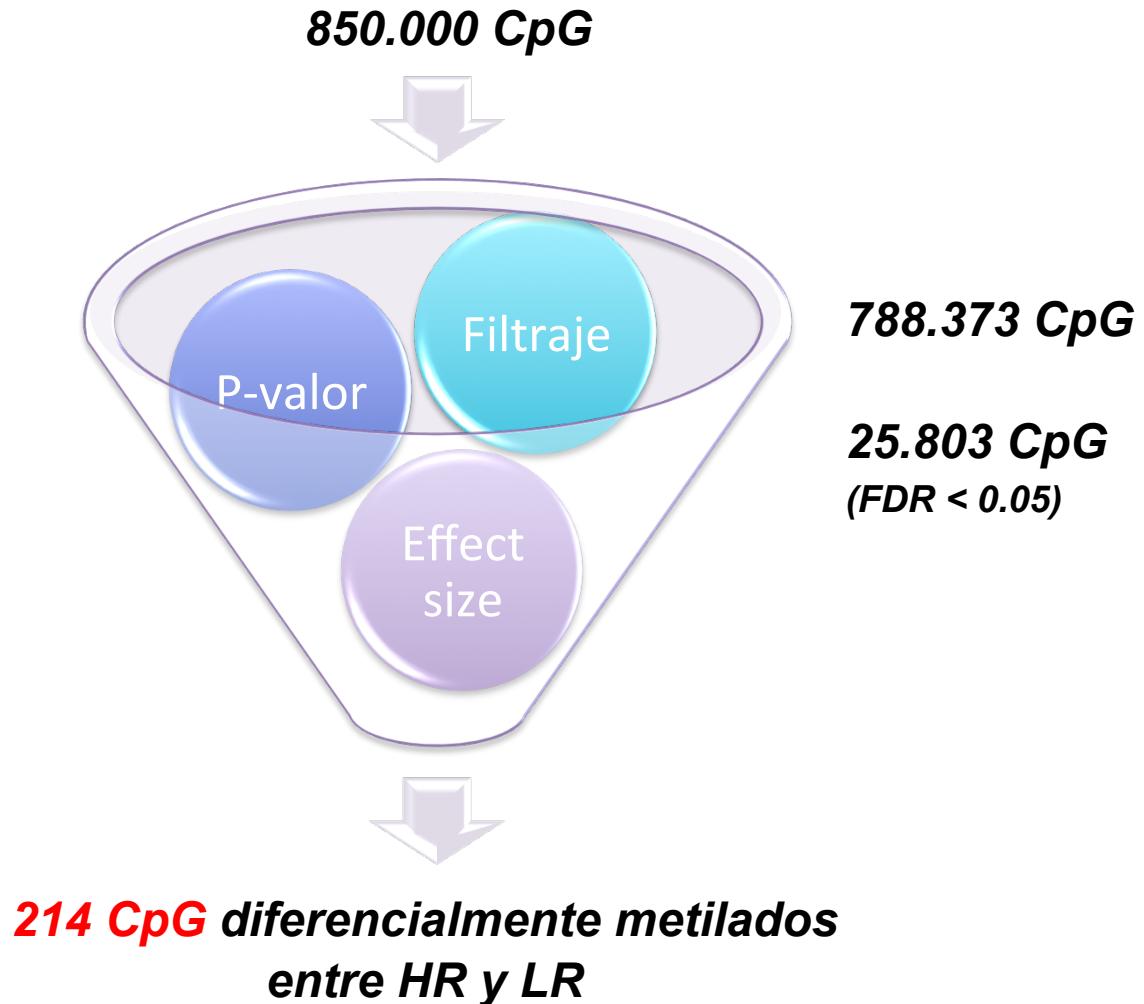
## 1) Filtraje

## 2) Regresión *limma*:

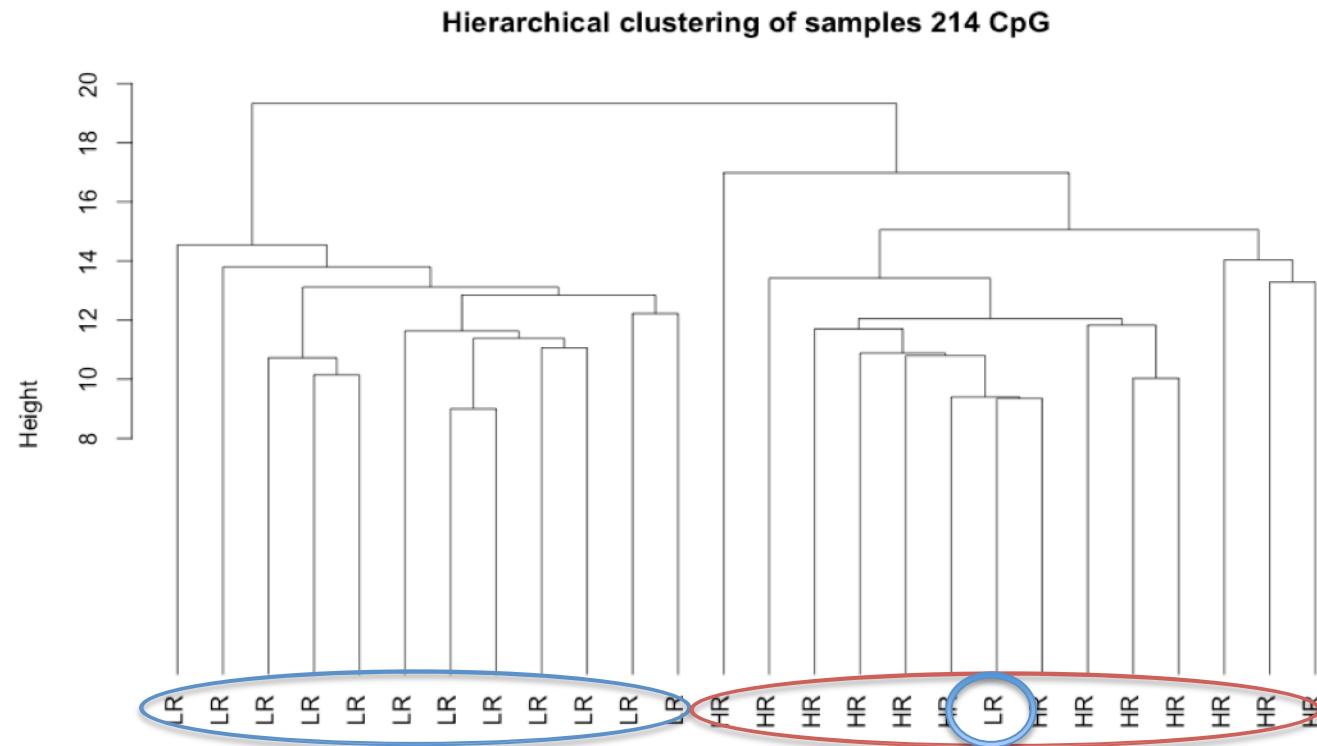
- Utilizamos los **valores de M**
- Modelo a partir del Grupo (HR o LR) y del sexo

## 3) *Effect size* (10%):

- Utilizamos los **valores de beta**



# Cluster en función de los CpG significativos



# 26 modelos de regresión con *limma*

Diseñamos una función en R que:

- Permite usar como predictor el valor de **DzBMI** o la etiqueta **HR/LR** para la regresión *limma*, según deseemos
- Permite usar los valores de **M** o de **beta** para realizar el segundo filtrado (*effect size*)
- Seleccionar el valor que deseemos para definir el *cutoff* del *effect size*

$$mval\_CpG_i \sim \beta_0 + \beta_1 DzBMI + \beta_2 Gender$$

# 26 modelos de regresión con *limma*

Usamos los valores de *DzBMI* como predictores y las cuatro condiciones siguientes:

1. Cutoff  $|\Delta\beta| \geq 0.1$  (10%)



wo_M053_0M	wo_M014_0M	wo_M048_0M	wo_M018_0M	wo_SA12_0M	wo_M013_0M	wo_M052_0M	wo_M031_0M
200	133	241	258	165	204	136	184
wo_M039_0M	wo_M027_0M	wo_M023_0M	wo_M042_0M	wo_SA10_0M	wo_M036_0M	wo_M045_0M	wo_SA32_0M
96	187	216	127	216	145	44	166
wo_M034_0M	wo_M032_0M	wo_M047_0M	wo_SA11_0M	wo_M044_0M	wo_M030_0M	wo_M028_0M	wo_M029_0M
268	201	187	262	178	141	192	151
wo_M041_0M	wo_SA37_0M			12	221		

2. Cutoff  $|\Delta\beta| \geq 0.15$  (15%)



wo_M053_0M	wo_M014_0M	wo_M048_0M	wo_M018_0M	wo_SA12_0M	wo_M013_0M	wo_M052_0M	wo_M031_0M
43	23	59	50	34	53	27	41
wo_M039_0M	wo_M027_0M	wo_M023_0M	wo_M042_0M	wo_SA10_0M	wo_M036_0M	wo_M045_0M	wo_SA32_0M
23	43	46	25	49	31	12	38
wo_M034_0M	wo_M032_0M	wo_M047_0M	wo_SA11_0M	wo_M044_0M	wo_M030_0M	wo_M028_0M	wo_M029_0M
59	53	38	70	42	33	36	32
wo_M041_0M	wo_SA37_0M			3	47		

3. Cutoff  $|\Delta\beta| \geq 0.2$  (20%)



wo_M053_0M	wo_M014_0M	wo_M048_0M	wo_M018_0M	wo_SA12_0M	wo_M013_0M	wo_M052_0M	wo_M031_0M
15	9	20	19	12	14	9	11
wo_M039_0M	wo_M027_0M	wo_M023_0M	wo_M042_0M	wo_SA10_0M	wo_M036_0M	wo_M045_0M	wo_SA32_0M
8	13	15	9	16	9	5	7
wo_M034_0M	wo_M032_0M	wo_M047_0M	wo_SA11_0M	wo_M044_0M	wo_M030_0M	wo_M028_0M	wo_M029_0M
19	24	11	23	15	10	12	10
wo_M041_0M	wo_SA37_0M			2	10		

4. Cutoff  $|\Delta M| \geq 1$



wo_M053_0M	wo_M014_0M	wo_M048_0M	wo_M018_0M	wo_SA12_0M	wo_M013_0M	wo_M052_0M	wo_M031_0M
79	49	100	106	65	90	49	74
wo_M039_0M	wo_M027_0M	wo_M023_0M	wo_M042_0M	wo_SA10_0M	wo_M036_0M	wo_M045_0M	wo_SA32_0M
37	81	89	44	93	59	16	75
wo_M034_0M	wo_M032_0M	wo_M047_0M	wo_SA11_0M	wo_M044_0M	wo_M030_0M	wo_M028_0M	wo_M029_0M
104	97	71	121	70	61	76	59
wo_M041_0M	wo_SA37_0M			6	90		

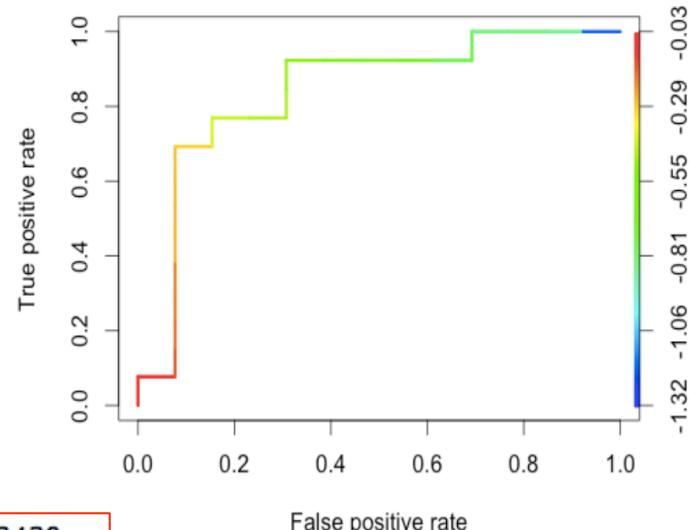
# Modelo consenso (PLS)

$$DzBMI_{excluded\_sample} \sim \beta_0 + \beta_1 CpG_1 + \beta_2 CpG_2 + \cdots + \beta_k CpG_k$$

Modelo 1. Cutoff  $|\Delta\beta| \geq 0.1$  (10%)

	CpGID	CHR	MAPINFO	UCSC_RefGene_Name	UCSC_RefGene_Group
167	cg04057818	2	67487963	LOC102800447	Body
355	cg14157435	2	206628692	NRP2;NRP2;NRP2;NRP2;NRP2	Body;Body;Body;Body;Body
744	cg05387464	2	9956256		
1612	cg00036352	8	144636448	GSDMD	5'UTR
2299	cg16007266	16	57050314	NLRP5	TSS1500
3019	cg08240913	10	117969024	GFRA1;GFRA1;GFRA1	Body;Body;Body
5589	cg23743554	11	65321226	LTBP3;LTBP3;LTBP3	Body;Body;Body
6282	cg18872420	14	78023429	SPTLC2	Body

$dzb \sim cg14157435 + cg00036352 + cg05387464 + cg08240913 + cg18872420 +$   
 $cg16007266 + cg23743554 + cg04057818$



El AUC para este modelo es del 84% y el cutoff DzBMI óptimo se encuentra en -0.426 (para este valor sensibilidad del 77% y especificidad del 85%).

# Modelo consenso (PLS)

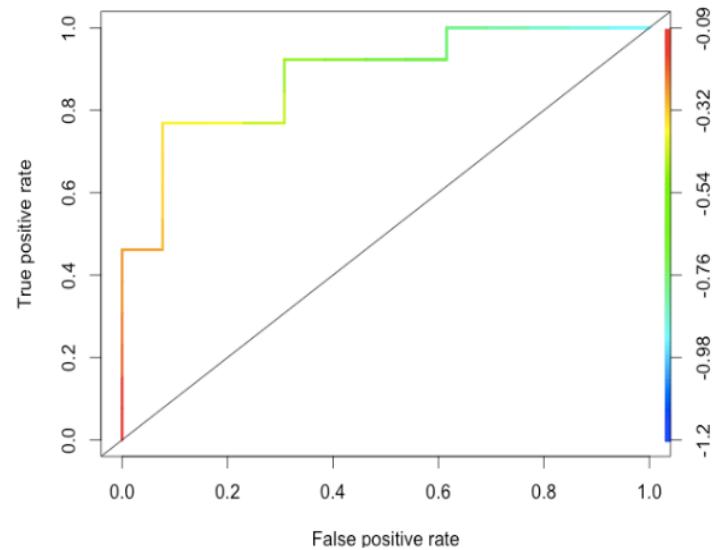
$$DzBMI_{excluded\_sample} \sim \beta_0 + \beta_1 CpG_1 + \beta_2 CpG_2 + \cdots + \beta_k CpG_k$$

Modelo 2. Cutoff  $|\Delta\beta| \geq 0.15$  (15%)

Modelo 3. Cutoff  $|\Delta\beta| \geq 0.20$  (20%)

CpGID	CHR	MAPINFO	UCSC_RefGene_Name	UCSC_RefGene_Group
355	cg14157435	2	206628692 NRP2;NRP2;NRP2;NRP2;NRP2	Body;Body;Body;Body;Body
2299	cg16007266	16	57050314 NLRC5	TSS1500

dzb ~ cg14157435 + cg16007266



El AUC del modelo es del 88% y el cutoff DzBMI óptimo se encuentra en -0.358 (para este valor sensibilidad del 77% y especificidad del 92%).

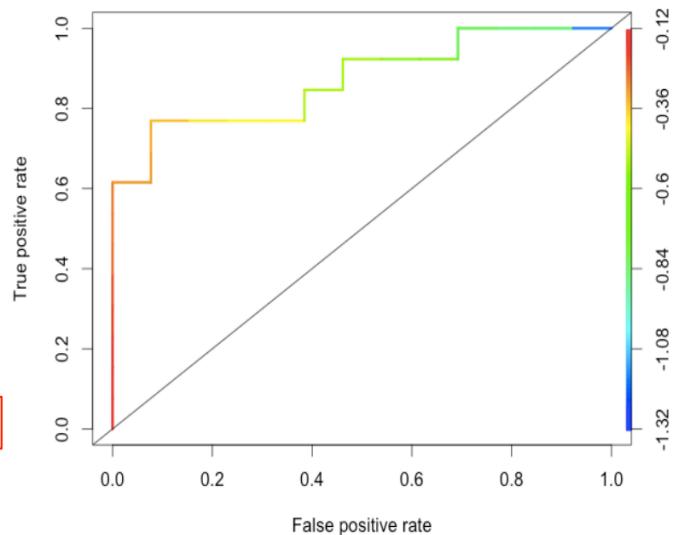
# Modelo consenso (PLS)

$$DzBMI_{excluded\_sample} \sim \beta_0 + \beta_1 CpG_1 + \beta_2 CpG_2 + \cdots + \beta_k CpG_k$$

## Modelo 4. Cutoff $|\Delta M| \geq 1$

	CpGID	CHR	MAPINFO	UCSC_RefGene_Name	UCSC_RefGene_Group
167	cg04057818	2	67487963	LOC102800447	Body
355	cg14157435	2	206628692	NRP2;NRP2;NRP2;NRP2;NRP2	Body;Body;Body;Body;Body
1612	cg00036352	8	144636448	GSDMD	5'UTR
2299	cg16007266	16	57050314	NLRC5	TSS1500

`dzb ~ cg14157435 + cg00036352 + cg16007266 + cg04057818`



El valor de AUC es del 87% y el cutoff DzBMI óptimo se encuentra en -0.354 (para este valor sensibilidad del 77% y especificidad del 92%).

# Modelo consenso final

Escogemos el modelo 1:

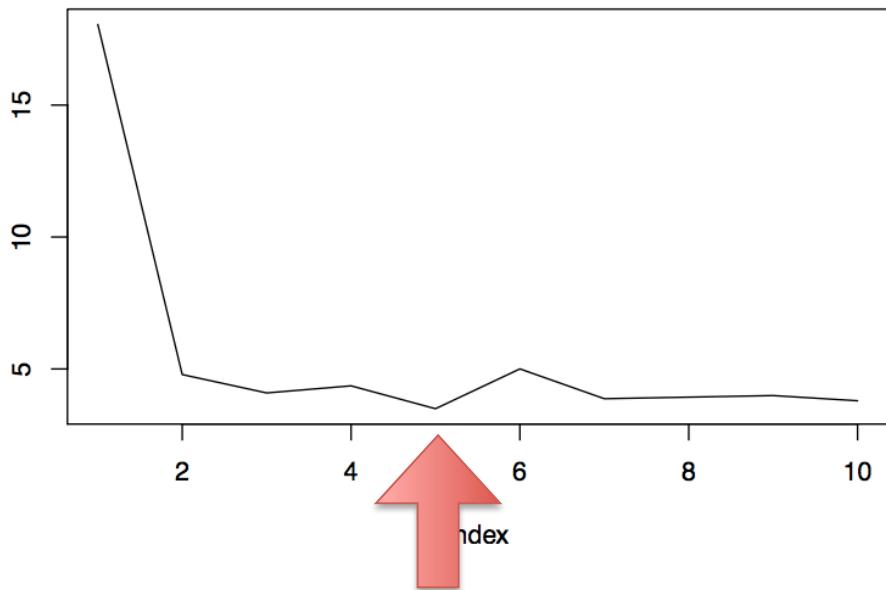
```
> modpls_selected$validation$PRESS
  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
dzb 1.939827 2.263937 2.69263 2.816453 2.815648 2.874687 2.88551 2.888362
> jack.test(modpls_selected, ncomp=1)
Response dzb (1 comps):
            Estimate Std. Error Df t value Pr(>|t|)    
cg14157435  0.0798113  0.0083020 25 -9.6135 7.062e-10 ***
cg00036352 -0.0771572  0.0178811 25 -4.3150 0.0002199 ***
cg05387464 -0.0487775  0.0066348 25 -7.3517 1.058e-07 ***
cg08240913 -0.0495441  0.0096373 25 -5.1409 2.590e-05 ***
cg18872420 -0.0676075  0.0189271 25 -3.5720 0.0014732 **
cg16007266 -0.0841298  0.0149264 25 -5.6363 7.257e-06 ***
cg23743554 -0.0583814  0.0108829 25 -5.3645 1.456e-05 ***
cg04057818  0.0817991  0.0163676 25  4.9976 3.750e-05 ***
...
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- AUC para este modelo = 84%
- Cutoff DzBMI = -0.426
- Sensibilidad del 77%
- Especificidad del 85%

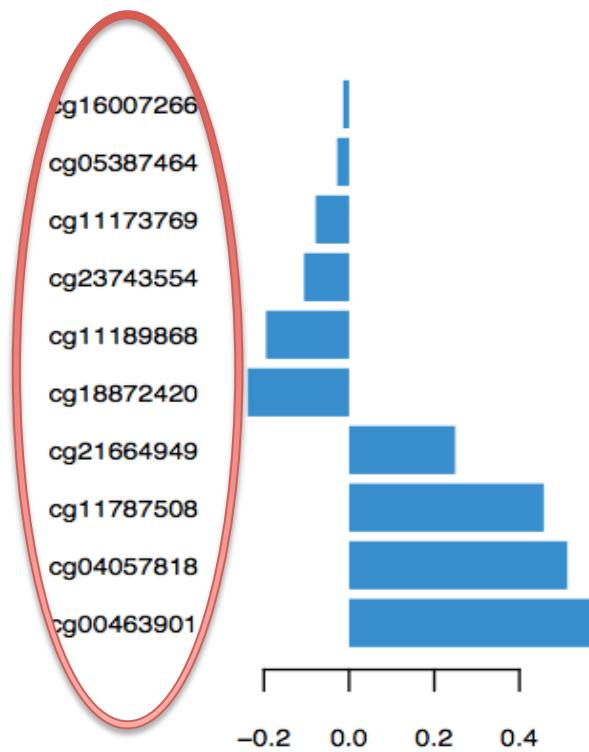
$$\begin{aligned} DzBMI = & (-0.0798113m1) + (-0.0771572m2) + \\ & (-0.0487775m3) + (-0.0495441m4) + (-0.0676075m5) + \\ & (-0.0841298m6) + (-0.0583814m7) + (0.0817991m8) \end{aligned}$$

# Modelos SPLS (214 CpG)

Usamos el paquete de R “*mixOmics*”:



Mínimo valor de PRESS  
en 5 componentes

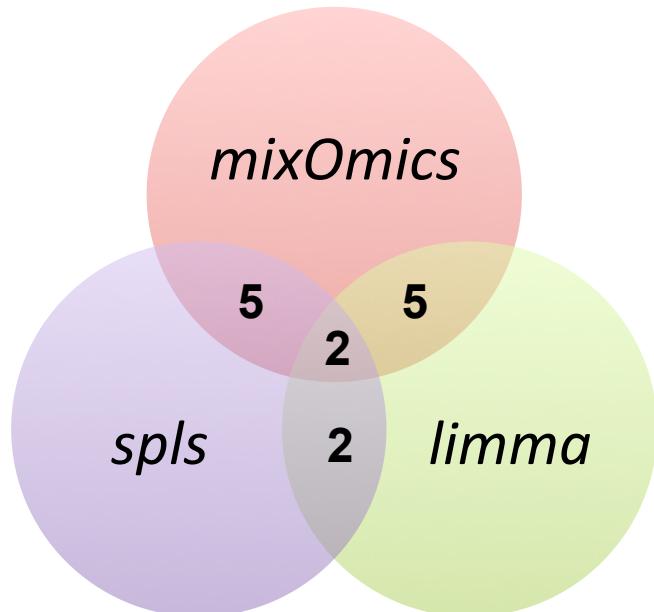


# Modelos SPLS (214 CpG)

Usamos el paquete de R “*spls*”:

- **Modelo con 5 componentes**
- **Predictores seleccionados:**

```
cg05401945 cg23118773 cg12119625 cg18872420
cg21664949 cg16613240 cg00463901 cg03047376
cg11787508 cg10062460 cg04057818 cg09357926
```



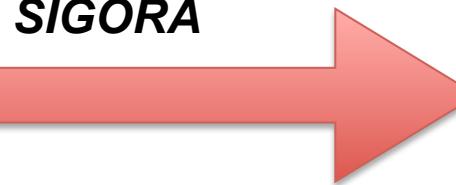
Intersección entre nuestro modelo limma y los dos otros modelos SPLS es de 2 predictores

# Enriquecimiento de genes

- Anotamos los 214 CpG
- Enriquecemos con el paquete de R “SIGORA”

IlmnID	CHR	MAPINFO	UCSC_RefGene_Name	UCSC_RefGene_Accession	UCSC_RefGene_Group
cg14157435	2	206628692	NF2;NRP2;NRP2;NRP2	NM_201279;NM_018534;N	Body;Body;Body;Body;
cg05387464	2	9956256			
cg09127314	1	152161683			
cg12372706	2	107592370			
cg03963391	16	34808479			
cg06758833	16	73575220			
cg11732055	10	9148214	KIF20B;KIF20B	NM_016195;NM_00128425	Body;Body
cg08693938	14	7343928	ZFYVE1;ZFYVE1;ZFYVE1	NM_178441;NM_00128173	Body;Body;Body;Body
cg22166084	17	360203			
cg09764652	14	55712439			
cg16081325	12	10651744	NUAK1	NM_014840	Body
cg20163335	20	6130954			
cg06538684	12	1251123	LOH12CR2;LOH12CR1	R_024061;R_032289	
cg07684019	5	13917459	PSD2	M_032289	
cg121664949	10	11801562	GFRA1;GFRA1;GFRA1	M_145793;NM_005264;N	Body;Body;Body
cg12271079	10	11799774	GFRA1;GFRA1;GFRA1	M_001145453;NM_00526	Body;Body;Body
cg12426470	17	1104627			
cg04109556	13	1918590			
cg15022308	10	4344707			
cg07114310	7	5153892			
cg18056749	20	55836268	BMP7	NM_001719	Body
cg04057818	2	67487963	LOC102800447	NR_110564	Body
cg27336068	10	1336103	ADAR2	NM_018702	Body
cg26173906	19	14376389			
cg06565913	16	1584452	IT140;TMEM204	NM_014714;NM_024600	Body;1stExon
cg00846098	21	43198791			
cg14920334	12	4829827	GALNT8;GALNT8	NM_017417;NM_017417	1stExon;5'UTR

SIGORA



# Enriquecimiento de genes

Solo encontramos un *pathway* sobrerepresentado en la base de datos REACTOME

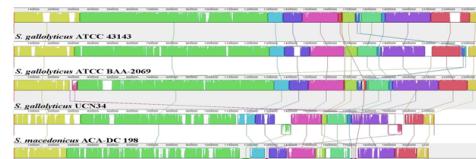
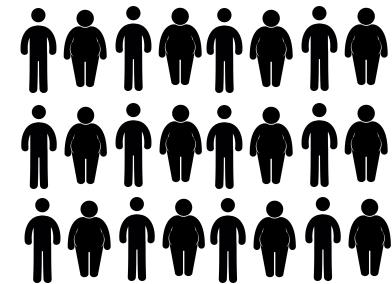


	description	pvalues	Bonferroni	successes
	Basal cell carcinoma	0.0002388	0.06686	3.00
	Natural killer cell mediated cytotoxicity	0.0031480	0.88140	3.72
	Cell adhesion molecules (CAMs)	0.0090980	1.00000	3.62
	cGMP-PKG signaling pathway	0.0217200	1.00000	2.17

	description	pvalues	Bonferroni	successes
	Endosomal/Vacuolar pathway	1.785e-27	1.571e-24	15.00
	RMTs methylate histone arginines	8.958e-04	7.883e-01	2.00
	Defective B4GALT7 causes EDS, progeroid type	3.636e-03	1.000e+00	1.00

# Líneas de trabajo futuro

1. Validación de nuestro modelo en la misma cohorte extendida y en cohortes externas
2. Análisis de enriquecimiento con más genes mediante una bajada de restricción en el segundo filtro (*effect size*)
3. Posible validación de los resultados del punto anterior en el laboratorio (mediante PCR, Western Blot...)



# *Muchas gracias por su atención*

