Methylation analysis

Lab meeting

Friday 3rd October 2023

Experimental design

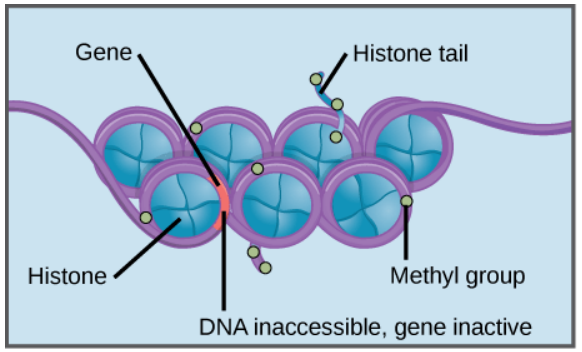
Sample	Age	Condition	Sex	Batch
30 mouse utricles (18 after filtering)	P2, P8, P30 (no repeated sampling)	Gentamycin +/-	M/F	5 batches of 6 mice

How does methylation level change:

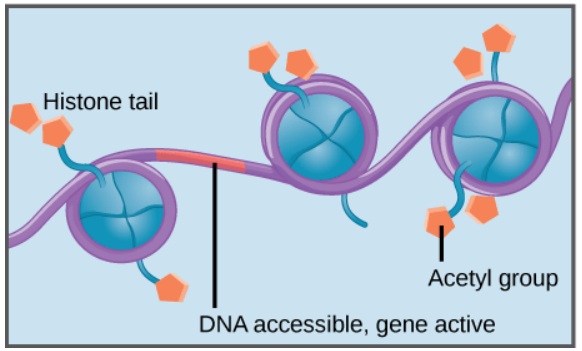
- in controls over time?
- in antibiotic-treated utricle over time?

How does methylation level compare between antibiotic-treated utricle vs control at each stage?

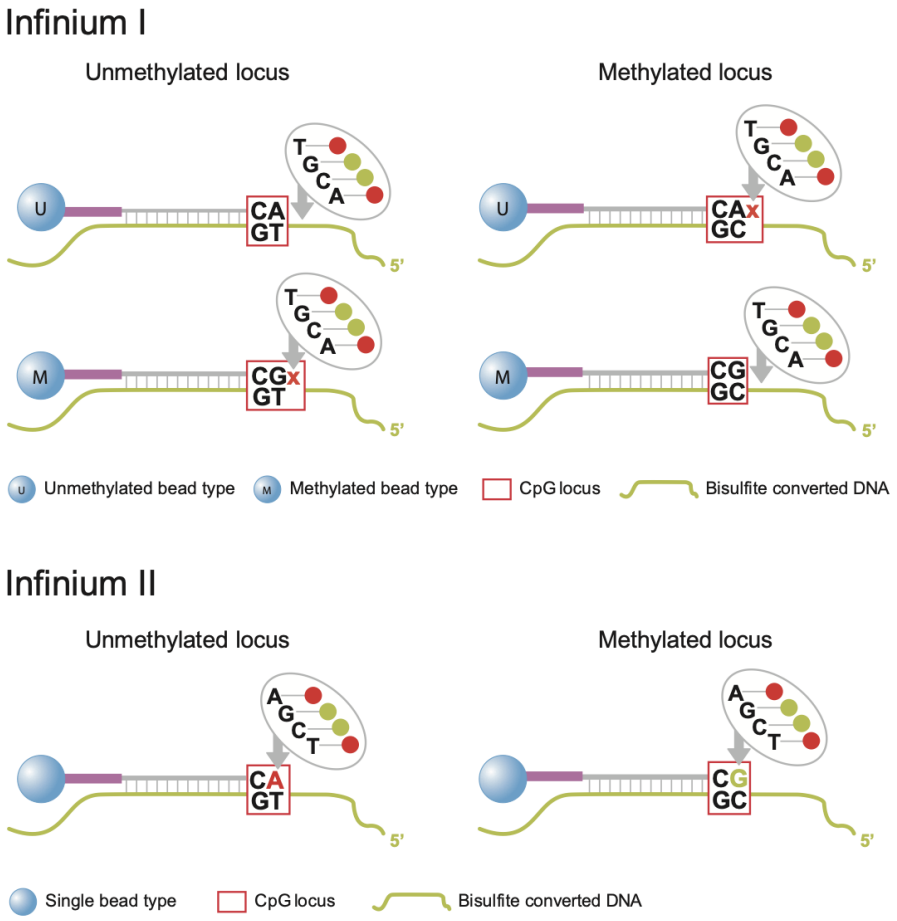
NB: Not really interested in sex or batch but they may have an effect.



Methylation of DNA and histones causes nucleosomes to pack tightly together. Transcription factors cannot bind the DNA, and genes are not expressed.

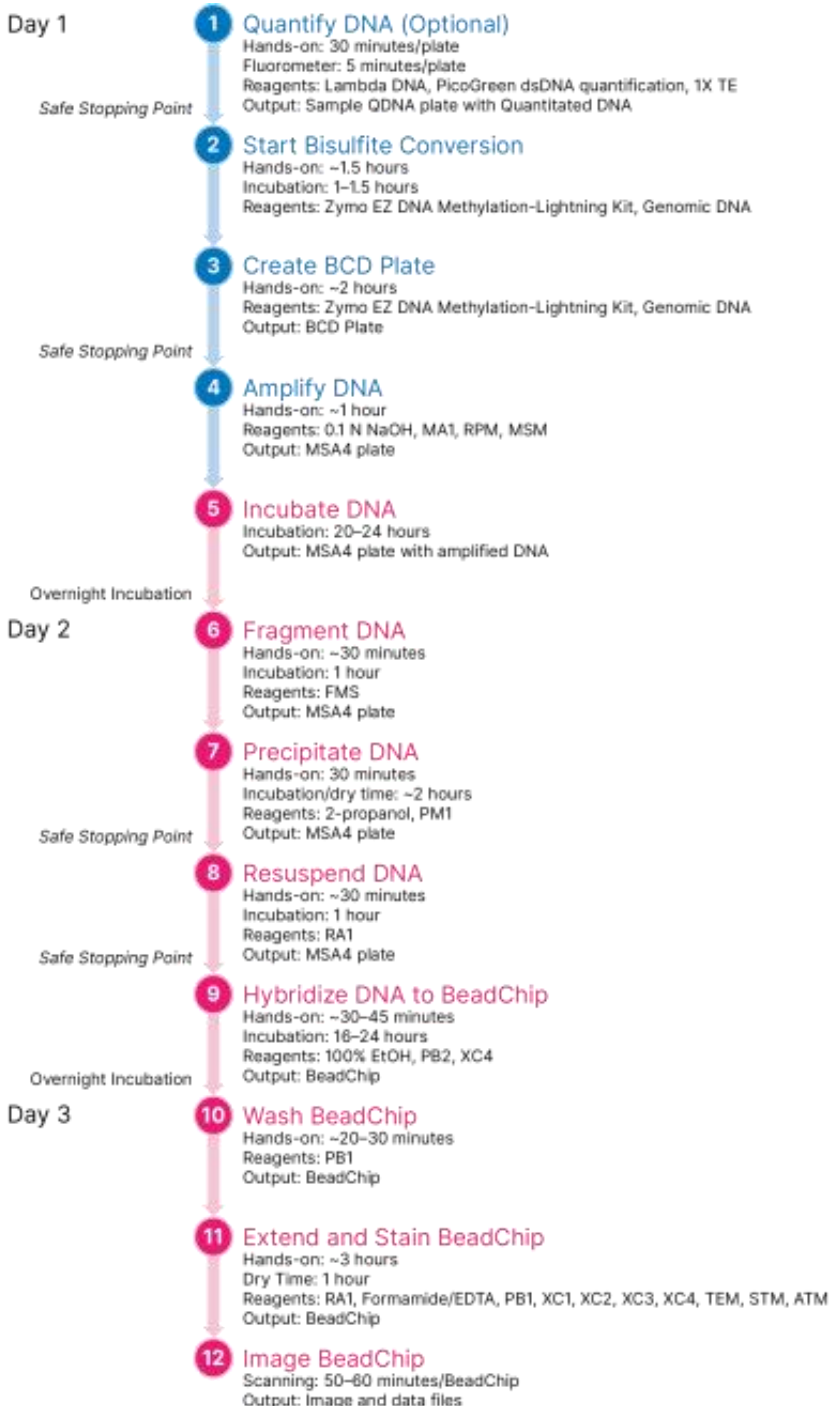


Histone acetylation results in loose packing of nucleo-somes. Transcription factors can bind the DNA and genes are expressed.



Method – Illumina Infinium BeadChip

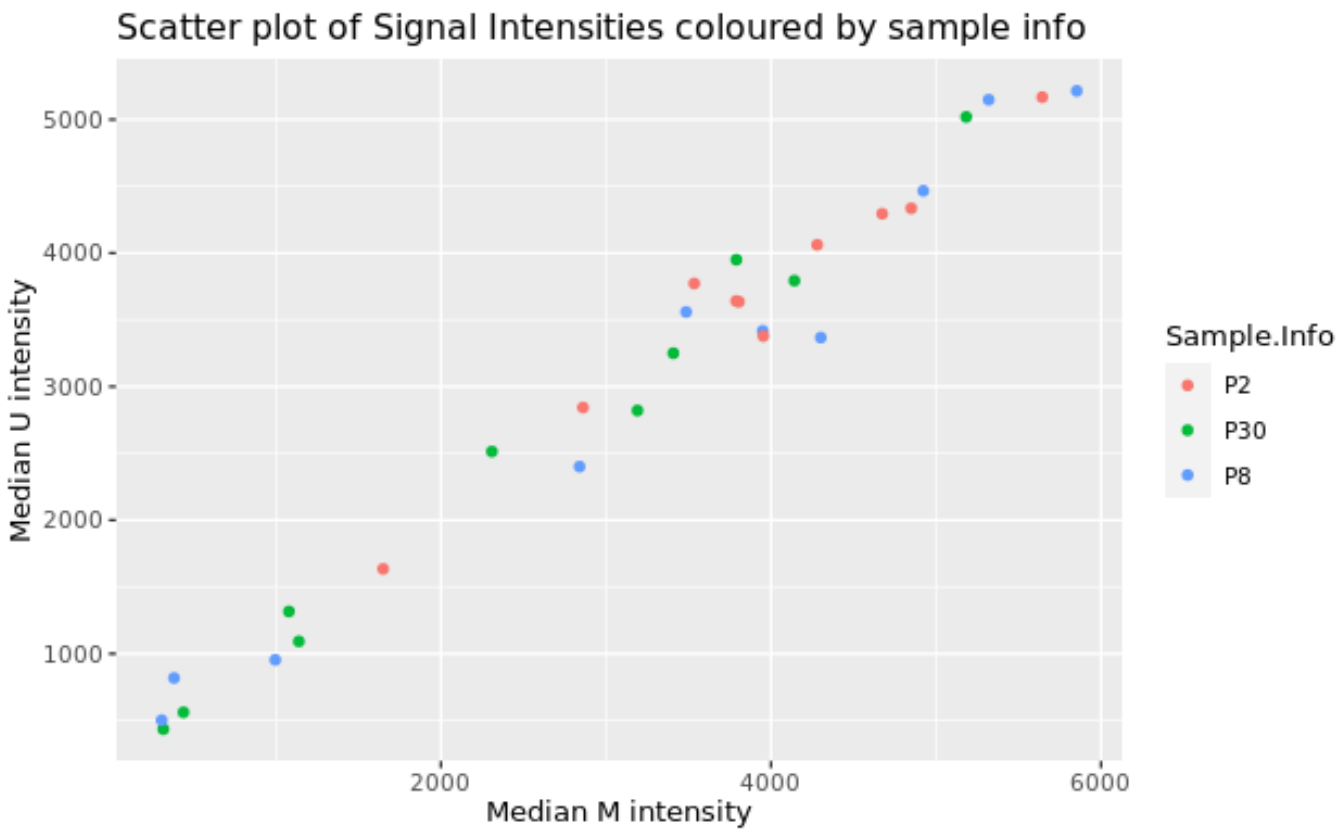
- Bisulphite conversion of genomic DNA allows detection of methylated loci
- Unmethylated C → U → T
- Two probe types are used: Infinium I and Infinium II
- Over 280k probes in total across the genome
- Probe signal intensity is then measured and quantified



Quality control – overview

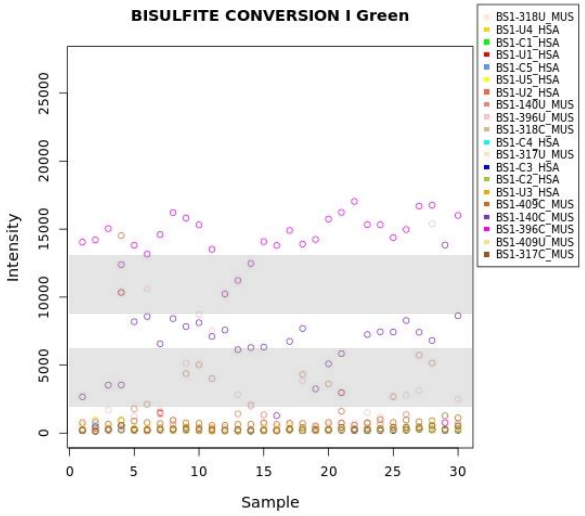
- Bisulphite conversion
- Intensity check
- Probe detection
- Probe-type normalisation

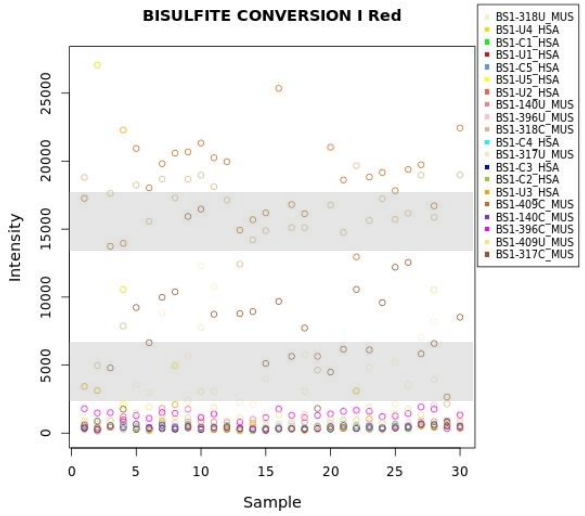
Quality control – intensity check

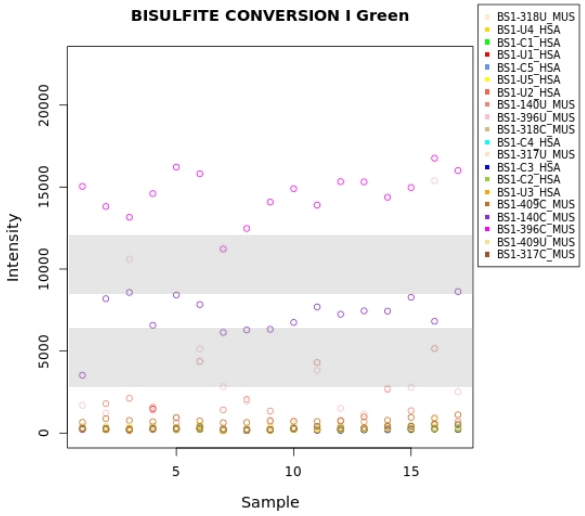


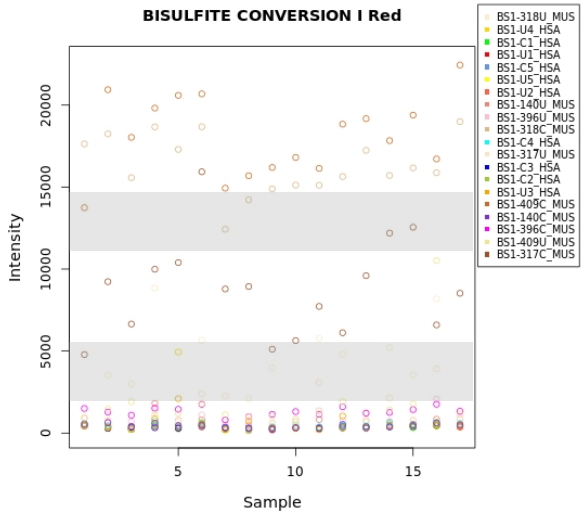
8 samples fail at a threshold of 2000 for median M and U intensity.

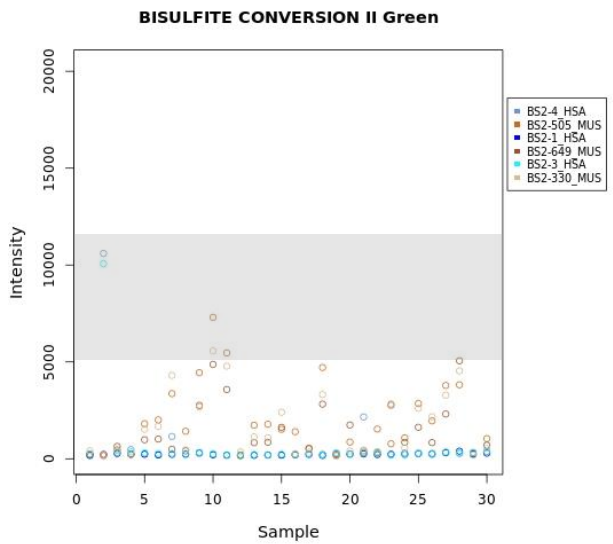
Quality control – bisulphite conversion

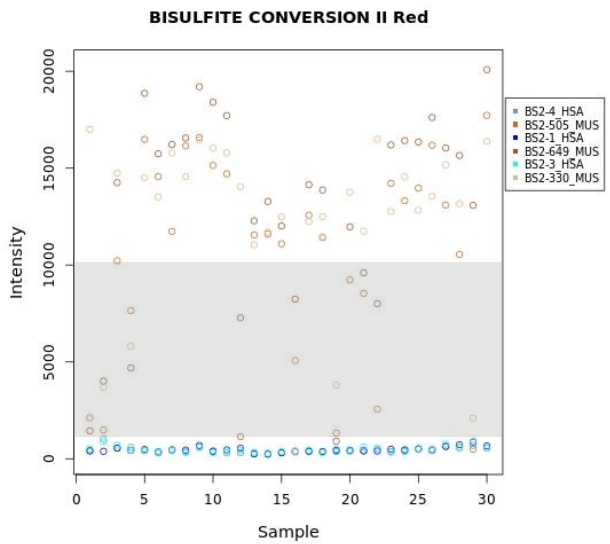


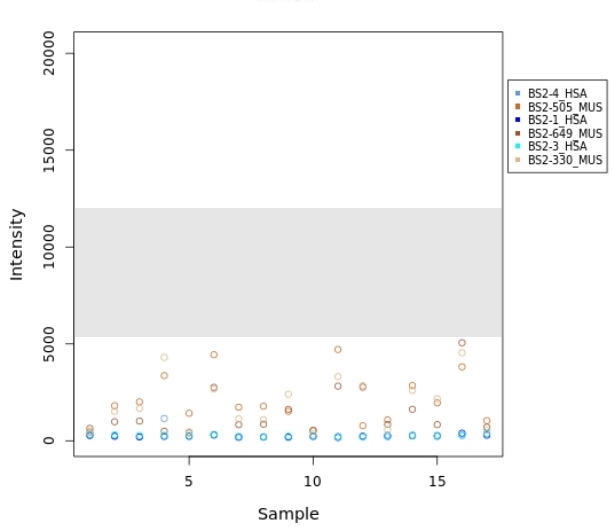


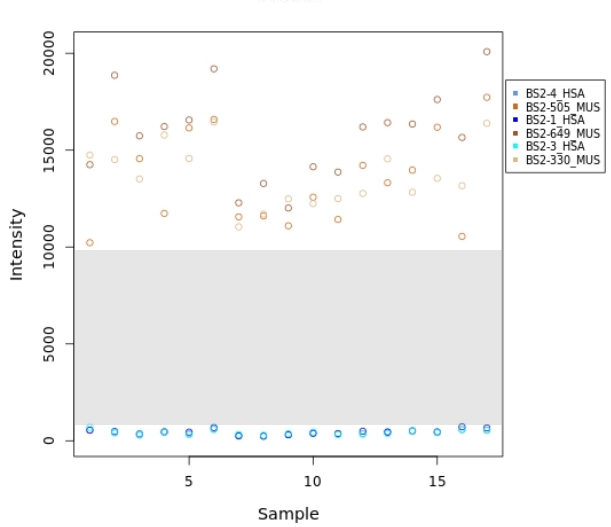




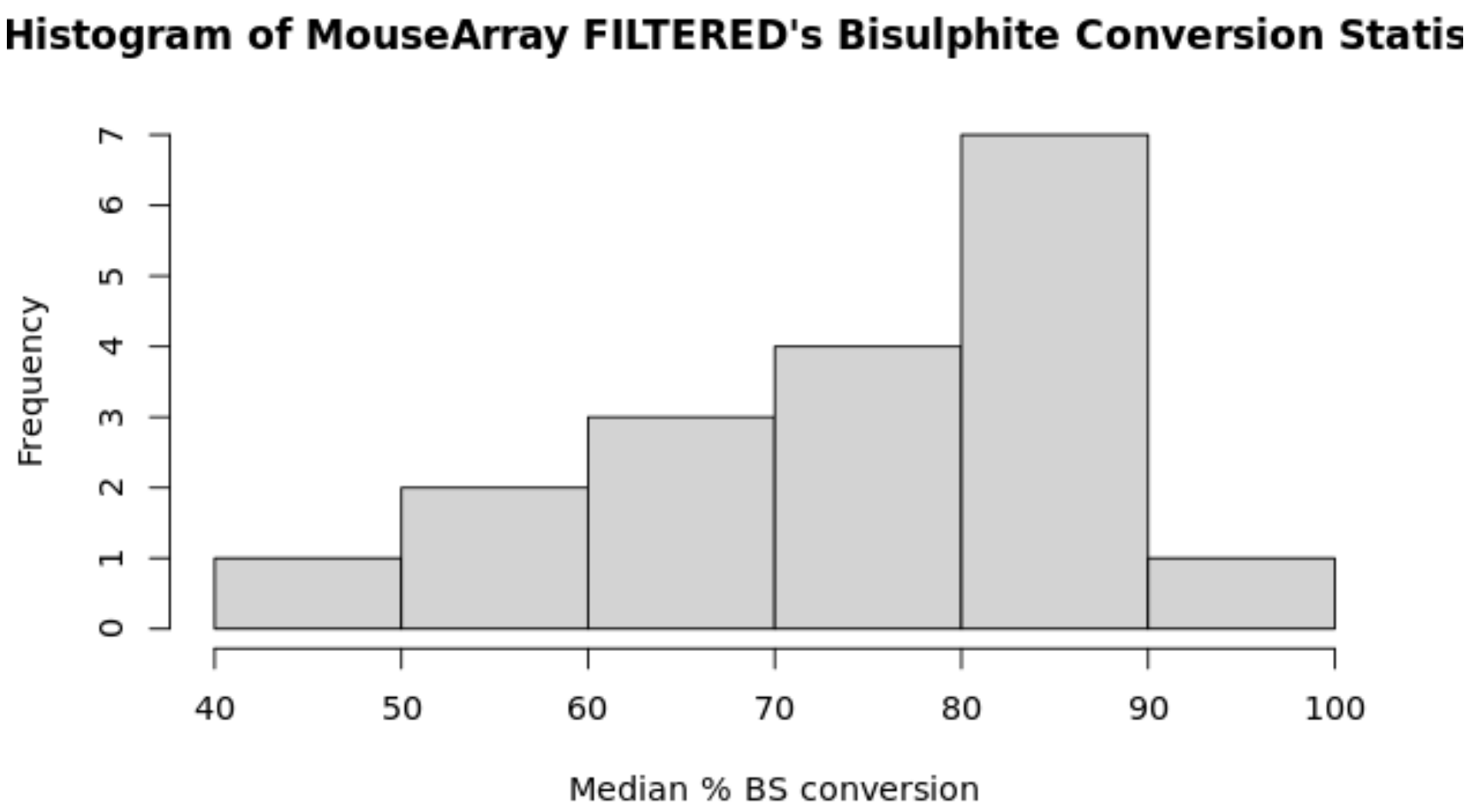








Quality control – bisulphite conversion



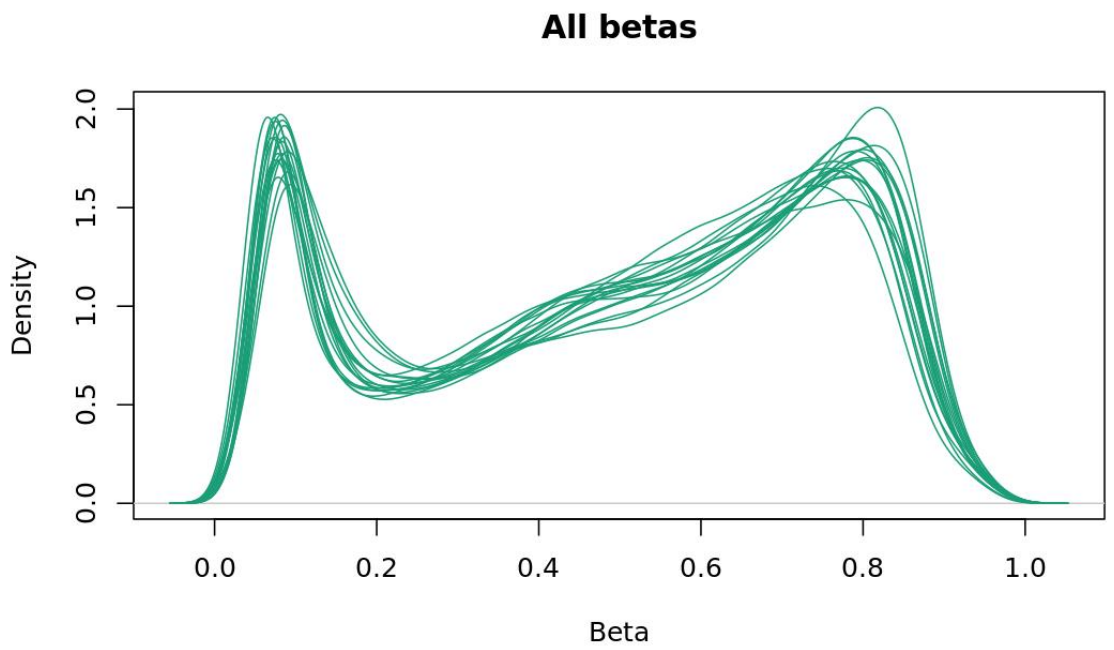
Quality control – probe detection

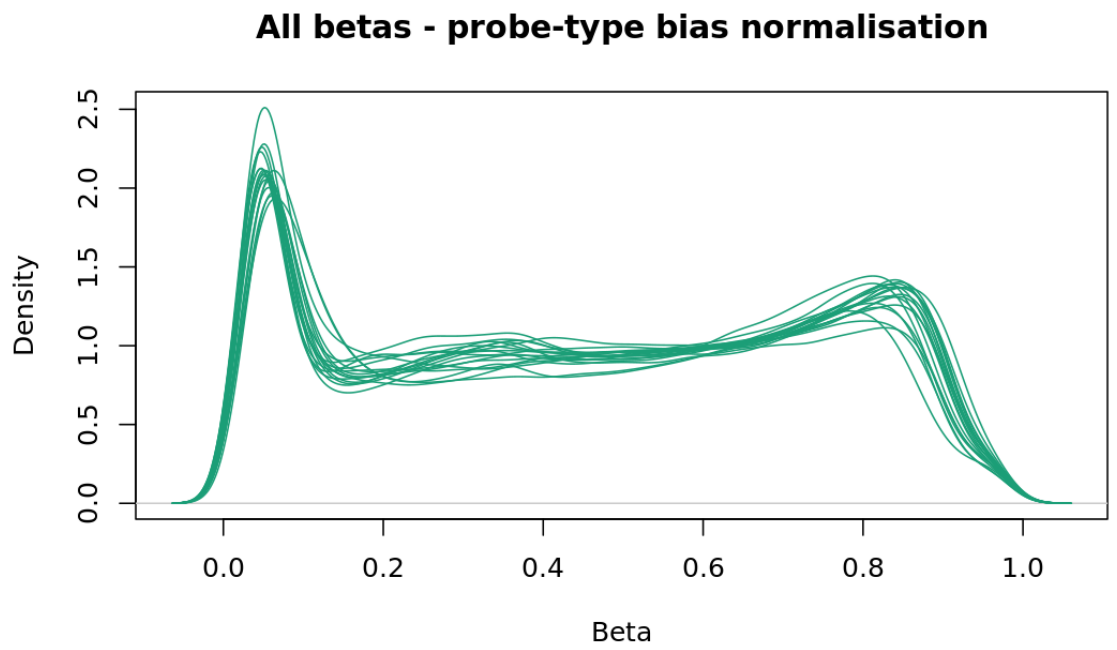
- Negative internal control (NIC) probes indicate background signal as they don’t bind DNA
- Calculated detection P values based on 600 NIC probes
- 5144 experimental probes have a detection P value of > 1% and therefore fail this step

Negative Controls

Negative control probes are randomly permutated sequences that should not hybridize to the DNA template. Negative controls are particularly important for methylation studies because of a decrease in sequence complexity after bisulfite treatment. The mean signal of these probes defines the system background. It is a comprehensive measurement of background, including signal resulting from cross-hybridization and nonspecific extension and imaging system background. The GenomeStudio platform uses the Average signal and standard deviation of 600 negative controls to establish detection limits for the methylation probes. Performance of negative controls should be monitored in both green and red channels.

Quality control – normalisation





Background & R/G colour corrected → Inter-array normalisation → Probe-type bias normalisation

Sample phenotypes after QC

Age	Condition	N	F	M
P2	Control	4	2	2
P2	Treated	3	2	1
P8	Control	3	1	2
P8	Treated	3	1	2
P30	Control	3	1	2
P30	**Treated**	**2**	**1**	**1**

All samples have a unique combination of age, condition, sex, and batch.

How is methylation quantified?

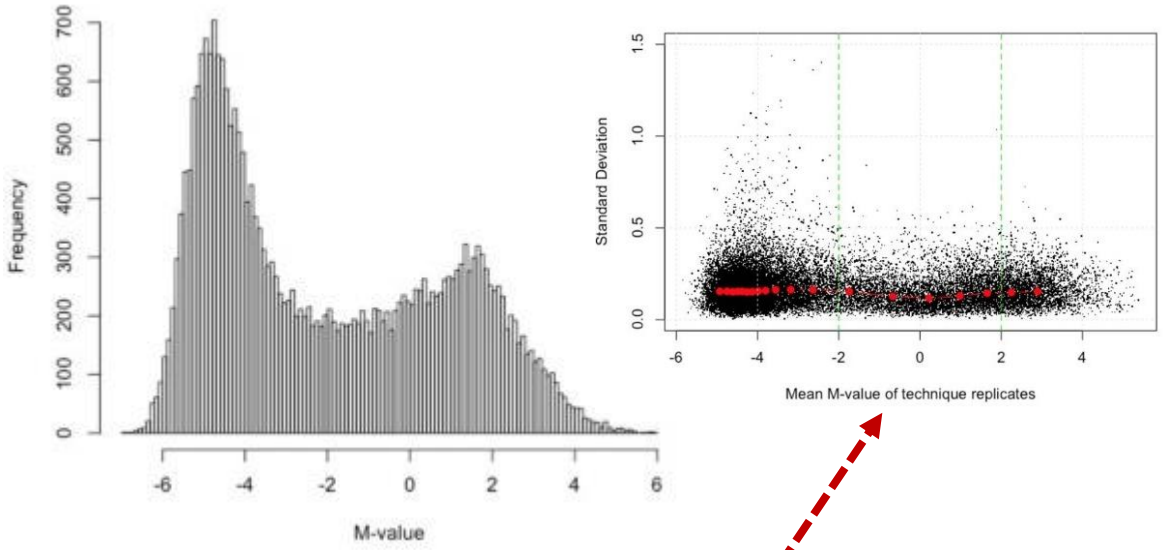
Data structure

Probes	Sample 1	Sample 2	Sample 3	…	Sample N
Probe 1	0.xx	0.xx	0.xx	0.xx	0.xx
Probe 2	0.xx	0.xx	0.xx	0.xx	0.xx
Probe 3	0.xx	0.xx	0.xx	0.xx	0.xx
…	0.xx	0.xx	0.xx	0.xx	0.xx
Probe N	0.xx	0.xx	0.xx	0.xx	0.xx
	phenotype	phenotype	phenotype	phenotypes	phenotype

Methylation can be quantified with M values or Beta values.

M: Log2 ratio of the intensities between methylated and unmethylated probe

Beta: % methylation at probe site



M values

- Homoscedastic
  - Great for stats! ☺
- Change in M is hard to interpret biologically
- **Use for differential methylation analysis ☺**

Du et al., BMC Bioinformatics, 2010

Data analysis

Probes	Sample 1	Sample 2	Sample 3	…	Sample N
Probe 1	0.xx	0.xx	0.xx	0.xx	0.xx
Probe 2	0.xx	0.xx	0.xx	0.xx	0.xx
Probe 3	0.xx	0.xx	0.xx	0.xx	0.xx
…	0.xx	0.xx	0.xx	0.xx	0.xx
Probe N	0.xx	0.xx	0.xx	0.xx	0.xx
	phenotype	phenotype	phenotype	phenotypes	phenotype

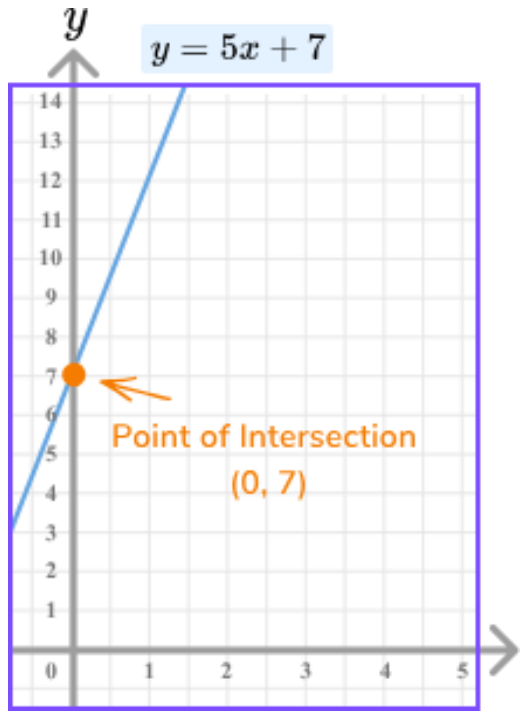
There is a complex relationship between methylation values and the phenotype of each sample.

This relationship can be described with a linear model.

What is a linear model?

y ~ m * x + c

- A familiar linear model!
- Allows you to:
  - calculate y for a given change in x
  - calculate the x-intercept of the line
- **~** means “is a function of”
- ***** means “multiplied by”



Methylation ~ age * condition

Interaction

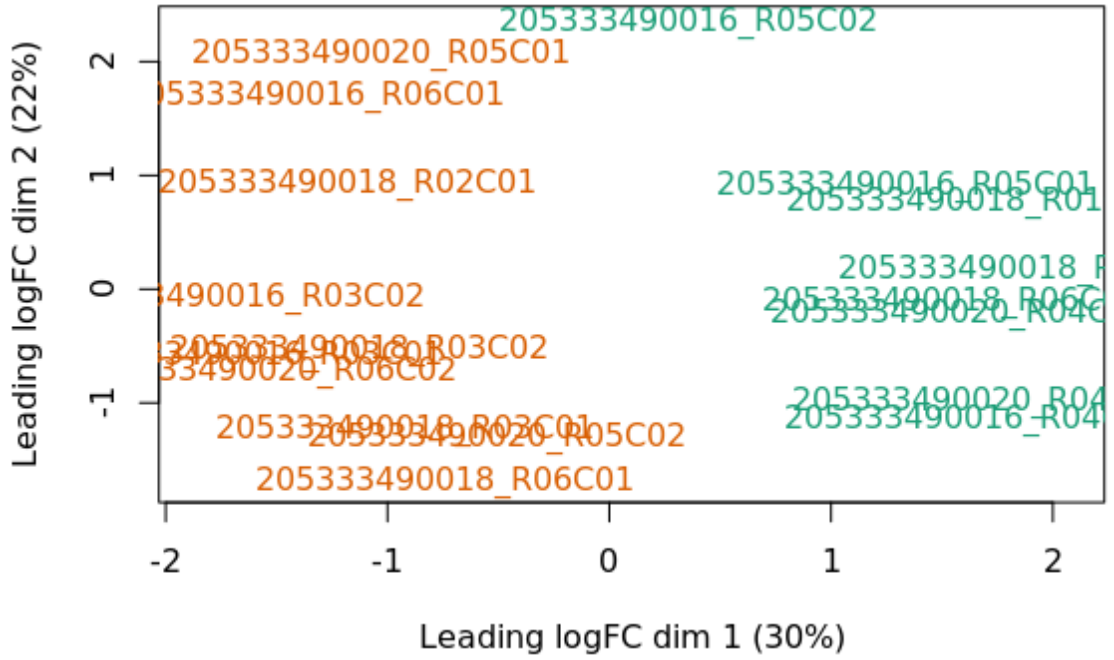
Increased methylation = reduced chromatin accessibility = low TF binding

- Linear model with interaction
- Age: Methylation in P2 vs P8 vs P30
- Condition: Methylation in gentamycin treated/untreated mice
- *: Methylation in P2- vs P2+, P8- vs P8+, P30- vs P30+
- *: P2-vsP8-vsP30- vs P2+vsP8+vsP30+ and separately!

Methylation ~ age * condition + sex

No interaction between factors

- Linear model with sex as a blocking factor
- This allows us to control for the effects of sex differences in the model
- MDS plot reveals sex to be the largest discriminant between the samples



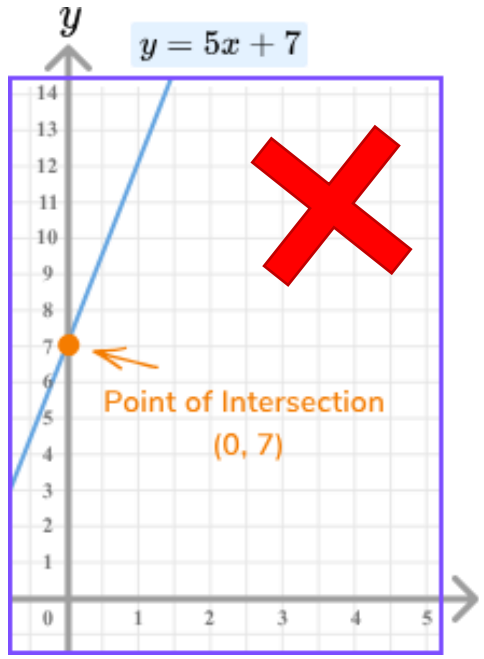
Methylation ~ age * condition + sex **+ (1|batch)**

Random effect

- Linear mixed effects model (lmer, a.k.a. random effects)
- Revealed that the variance of batch effect is close to zero
- i.e. Batch is correlated to another variable and not useful in the model
- **We can exclude it from the model!**

Methylation ~ 0 + age * condition + sex

- Age, condition, and sex are all factors (i.e. not numerical)
- We're not interested in the "intercept" x=0
- We can replace it with P2 controls by adding zero



Methylation ~ 0 + age * condition + sex

```
lmFit(mvals ~ 0 + age_condition + sex)
```

lmFit is a function from the limma package.

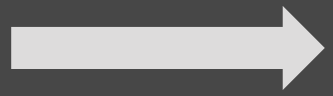
mvals = M values for each probe in each sample.

Age, condition, and sex phenotype for each sample taken from a summary table.

Identifying differentially methylated probes (DMPs)

Model contrasts

How does methylation change in controls over time?



```
"Control_P08-Control_P02",
"Control_P30-Control_P08",
"Control_P30-Control_P02",
```

How does methylation change in gentomycin treated utricle over time?

```
"Treated_P08-Treated_P02",
"Treated_P30-Treated_P08",
"Treated_P30-Treated_P02",
```

How does methylation in gentamycin treated utricle compare to control at each stage?

```
"Treated_P08-Treated_P02",
"Treated_P30-Treated_P08",
"Treated_P30-Treated_P02",
```

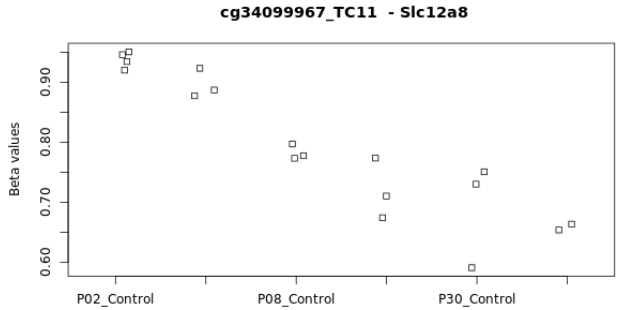
What is the interaction between treatment & age?

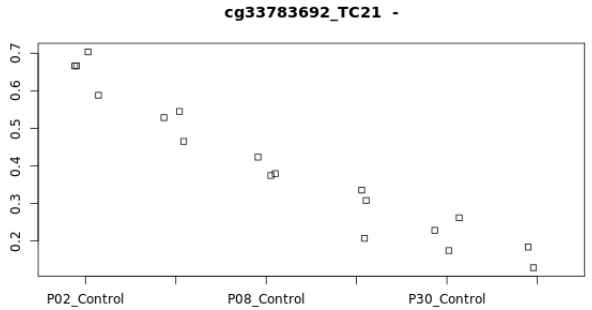
```
(Treated_P08-Treated_P02)-(Control_P08-Control_P02),
(Treated_P30-Treated_P08)-(Control_P30-Control_P08),
(Treated_P30-Treated_P02)-(Control_P30-Control_P02),
```

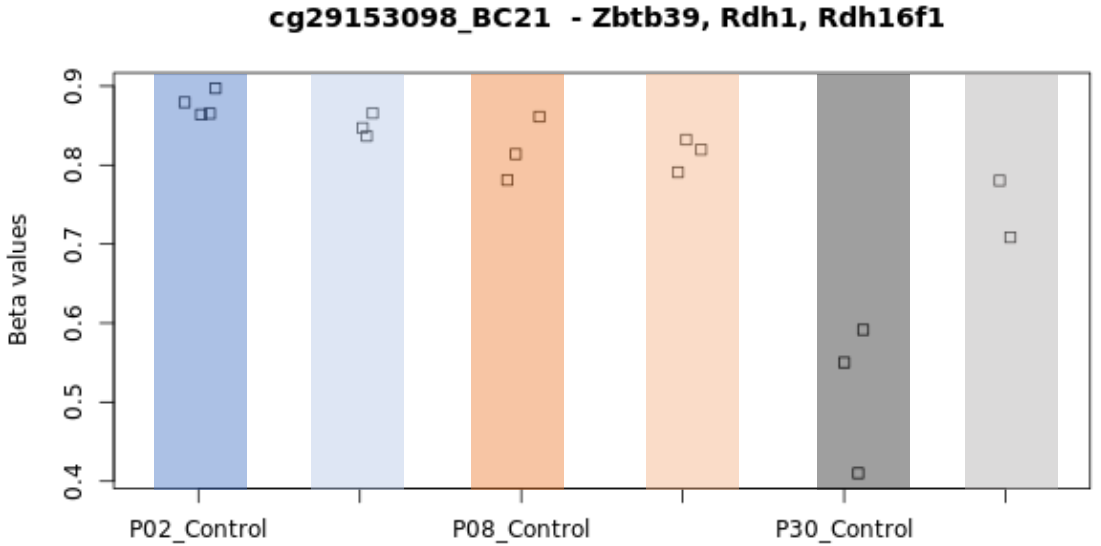
P2 vs P30 controls - table

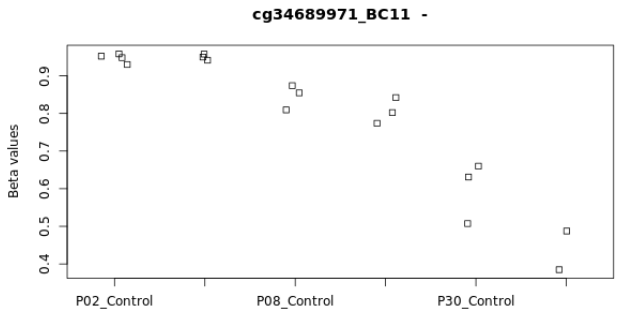
Name	adj.P.Val	P8-P2	P30 - P8	P30 - P2	AveExpr	chr	gene_name	gene_type
cg34689971_BC11	3.62E-04	-1.73	-1.89	-3.61	2.45	17		
cg40318641_TC21	3.62E-04	-1.80	-1.44	-3.24	1.03	3	Elovl6	protein_coding
cg33783692_TC21	3.62E-04	-1.50	-1.20	-2.70	-0.69	15		
cg36142168_TC11	3.62E-04	-1.38	-2.80	-4.17	0.73	19	Rab3il1, Fads3	protein_coding, protein_coding
cg42136853_TC21	3.62E-04	-1.36	-2.01	-3.38	1.12	5		
cg29153098_BC21	1.19E-03	-0.59	-2.10	-2.69	1.96	10	Zbtb39, Rdh1, Rdh16f1	protein_coding, protein_coding, protein_coding
cg47157417_BC11	1.19E-03	-1.32	-1.07	-2.39	1.36	9		
cg35622047_BC21	1.86E-03	-1.86	-1.22	-3.08	0.66	18	Pcdh1	protein_coding
cg34099967_TC11	2.21E-03	-2.10	-0.66	-2.76	2.24	16	Slc12a8	protein_coding
cg40060908_BC21	2.21E-03	-2.00	-0.43	-2.43	0.03	3	Bola1, Gm15444, H2ac20	protein_coding, antisense, protein_coding

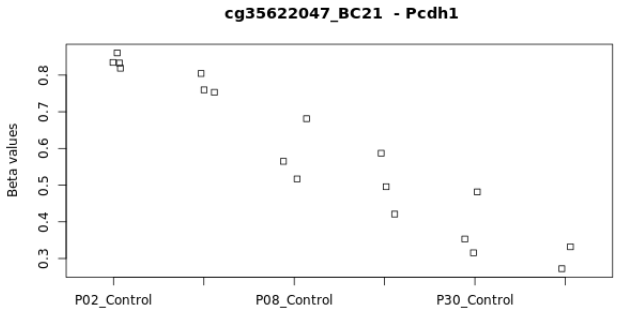
P2 vs P30 controls - plots

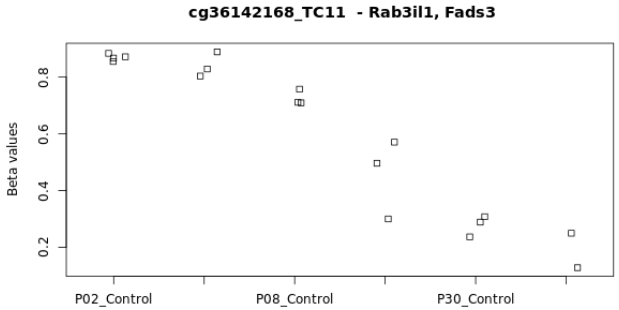


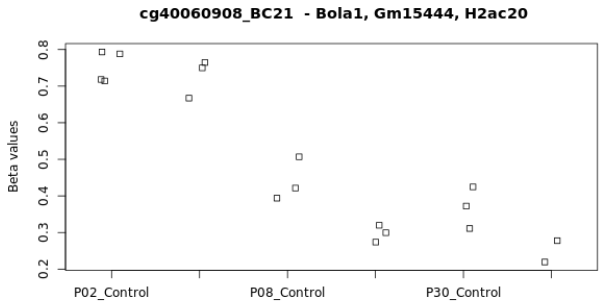


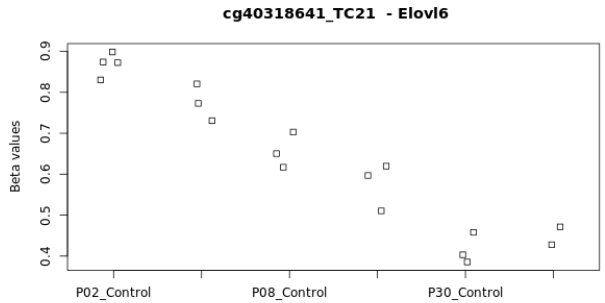


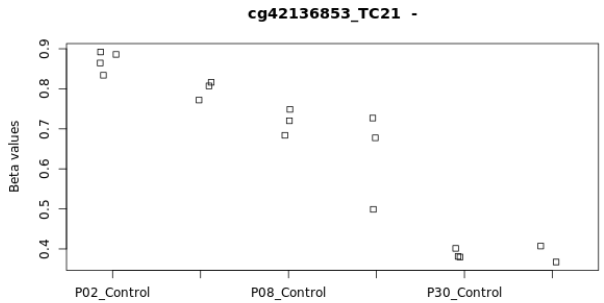


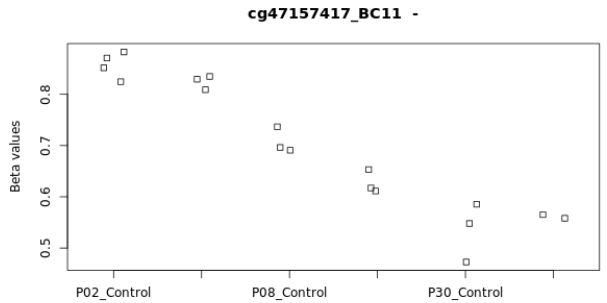










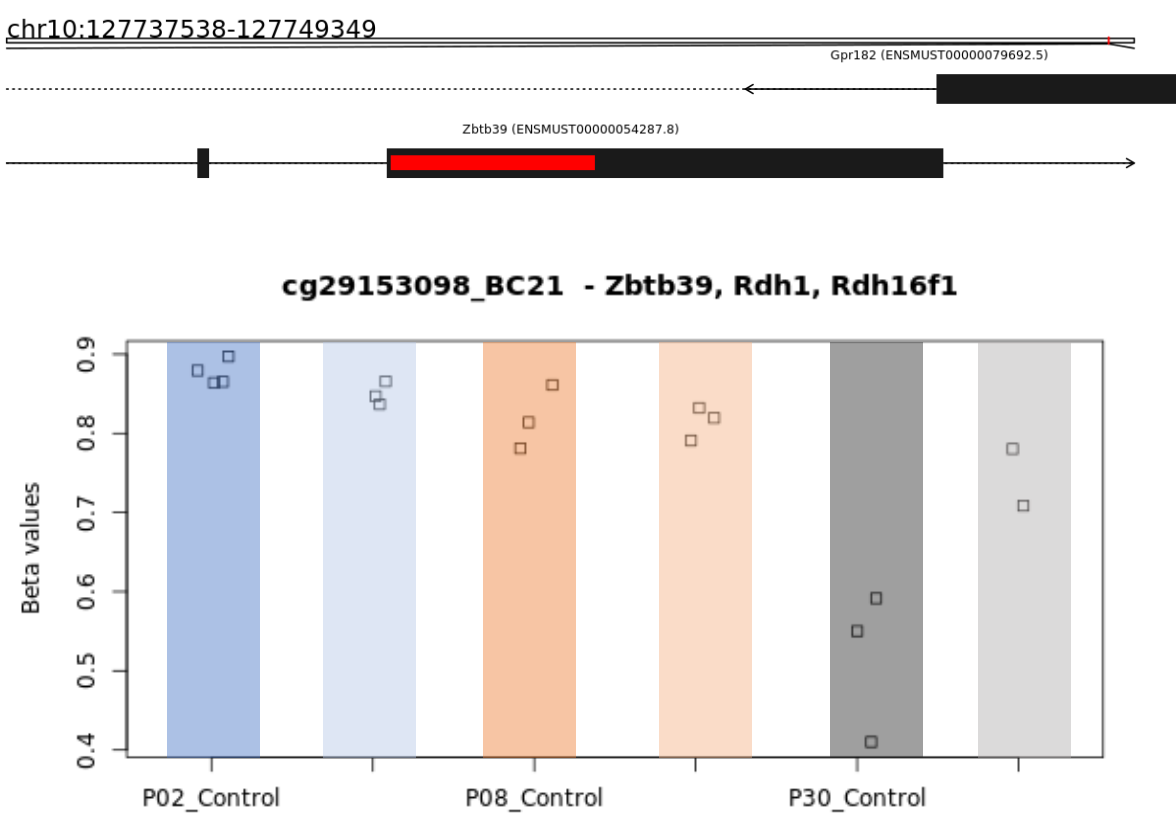


P2 vs P30 genes of interest

Gene	Function	Protein Class
Elovl6	Elongation of very long chain fatty acids protein 6	Metabolite interconversion enzyme
Rab3il1	Guanine nucleotide exchange factor for Rab-3A	Protein-binding modulator
Fads3	Fatty acid desaturase 3	Metabolite interconversion enzyme
Zbtb39	Zinc finger and BTB domain-containing 39	Gene specific transcriptional regulator
Rdh1 (Rdh16f1)	Retinol dehydrogenase 1 (All trans)	Metabolite interconversion enzyme
Pcdh1	Protocadherin 1	Cadherin
Slc12a8	Solute carrier family 12 members 8 and 9	Transporter
Bola1	BolA-like protein 1	Cell adhesion molecule
Gm15444	Unmapped (antisense)	-
H2ac20	Histone H2A type 2-C	Chromatin binding

Zbtb39

- Probe cg29153098_BC21
- Methylation decreases after P8
- Predicted to be involved in regulation of transcription by RNA polymerase II
- May be associated with Duane-Radial Ray Syndrome a.k.a. Okihiro syndrome
  - Autosomal dominant disorder characterised by upper limb, ocular, and renal anomalies
  - Can also be accompanied by **sensorineural deafness** and gastrointestinal anomalies
  - Onset at birth… probably not so relevant



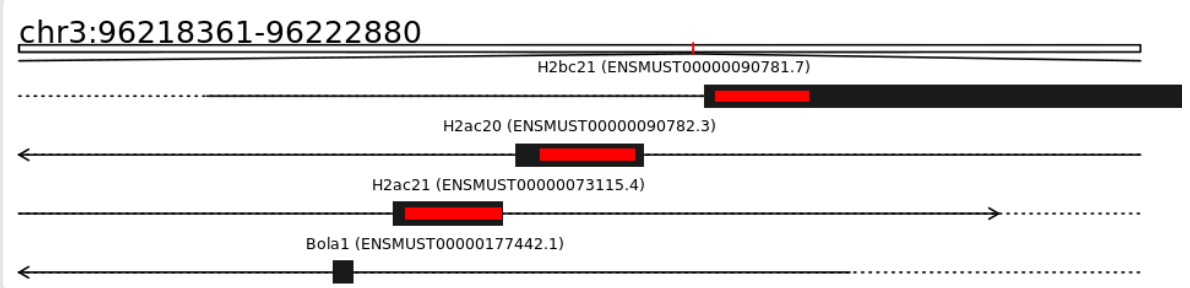
Molecular Mechanisms of Cereblon-Interacting Small Molecules in Multiple Myeloma Therapy.
Costacurta M, He J, Thompson PE, (and 1 more) ; J Pers Med (2021); PMID: 34834536
[View text]

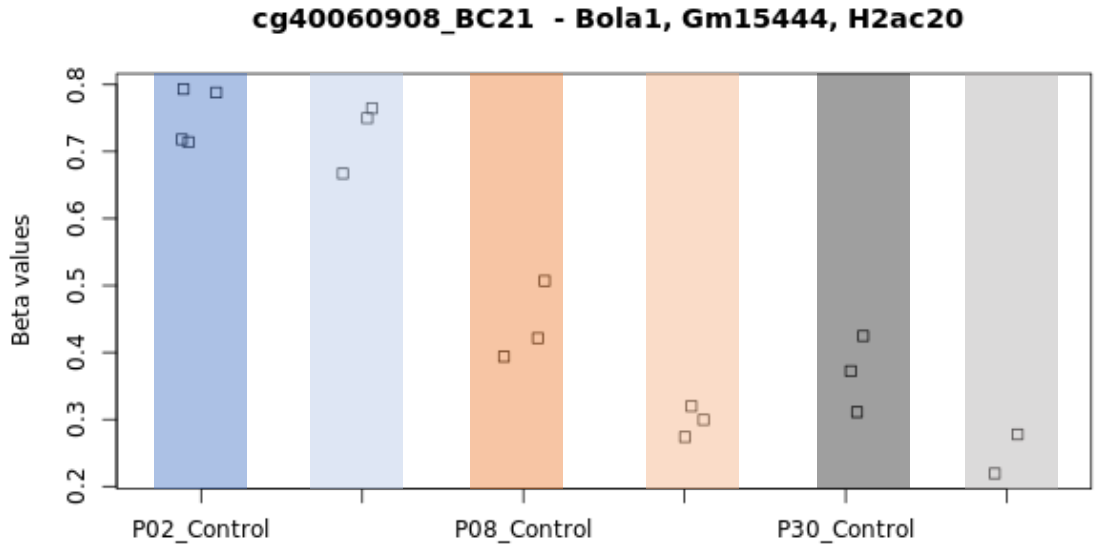
Thalidomide and its metabolite 5-hydroxythalidomide induce teratogenicity via the cereblon neosubstrate PLZF.
Yamanaka S, Murai H, Saito D, (and 9 more) ; EMBO J (2021); PMID: 33470442
[View text]

Thalidomide promotes degradation of SALL4, a transcription factor implicated in Duane Radial Ray syndrome.
Donovan KA, An J, Nowak RP, (and 5 more) ; Elife (2018); PMID: 30067223
[View text]

H2ac20

- Probe cg40060908_BC21
- Methylation decreases after P2
- H2A clustered histone 20
- “This gene is intronless and encodes a replication-dependent histone that is a member of the histone H2A family.”



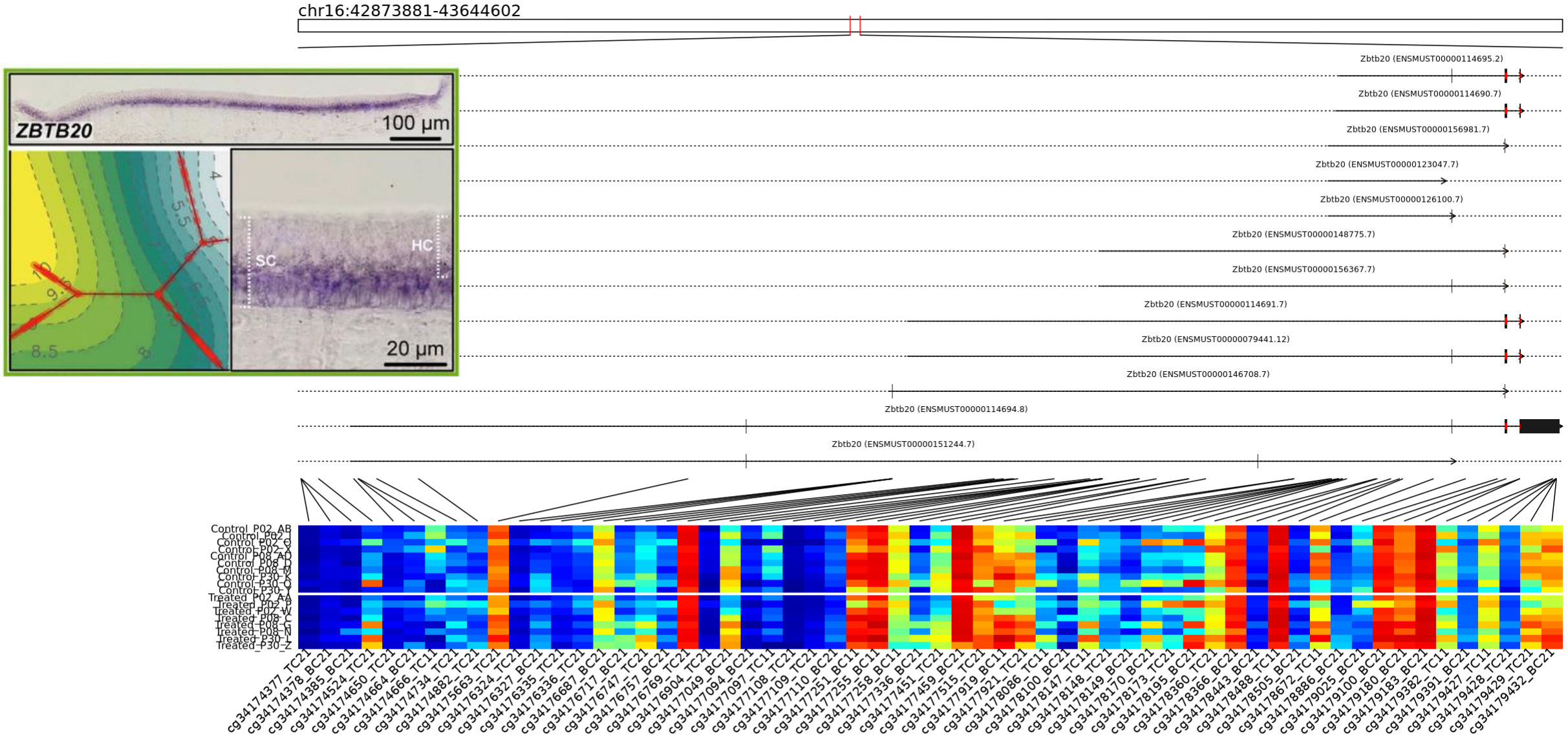


J Otol. 2017 Jun; 12(2): 47–54. PMCID: PMC5963466
Published online 2017 Apr 28. doi: 10.1016/j.joto.2017.04.002 PMID: 29937837

Histone deacetylases in hearing loss: Current perspectives for therapy

Daishi Chen,[a] Ming Xu,[a,b] Beibei Wu,[c] and Lei Chen[a,*]

It gets way more complicated…



Next steps

- Look at our favourite genes!
- Look for subtle changes in methylation
- Look at other contrasts in the dataset (ages, treatments, sex)

Is this methylation pipeline “worth it”?

Yes!

We’re getting interesting results already from a small sample with low statistical power.

It’s worth repeating with more samples:

- Do we want to include more ages?
- Do we need antibiotic treatment comparison?