Validation of Passage Effect

Email from Komal August 16 with instructions to download etc.

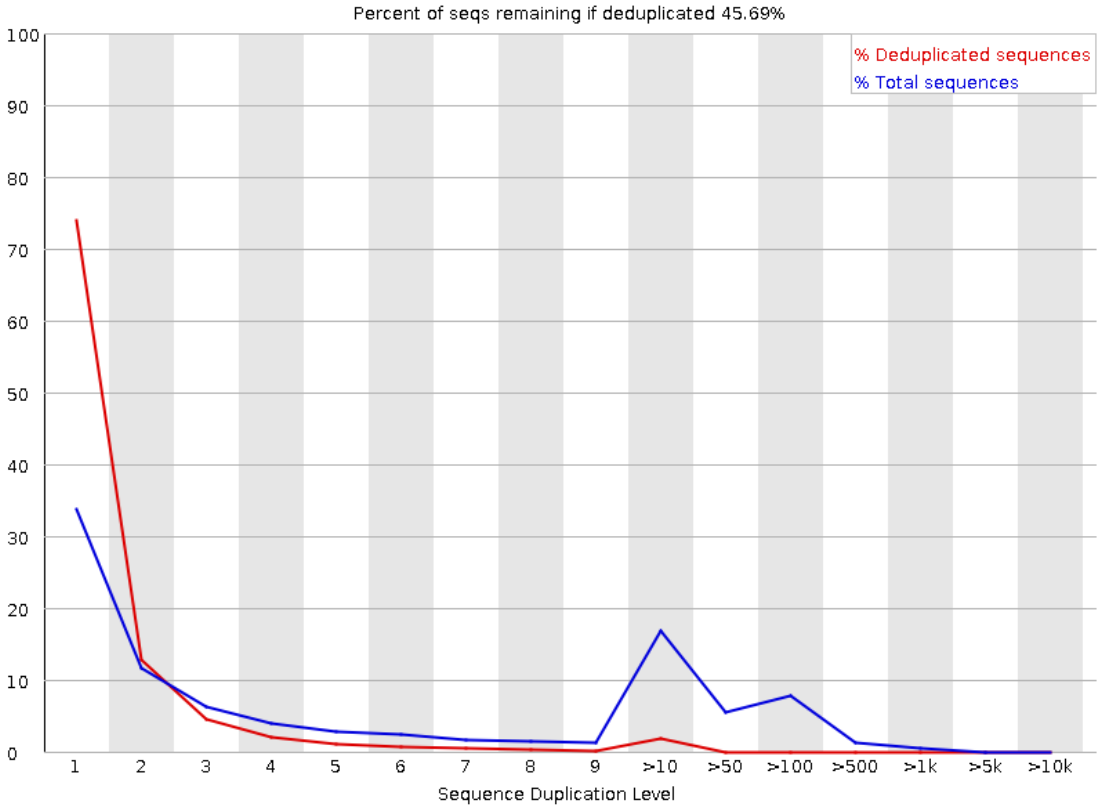
Quality of RNA-seq data

Looking at multi QC of run 1 and run 2 both pass all measures but "Per Base Sequence Content" "Sequence Duplication Levels"

The Per Base Sequence Content failure can be explained in fastQC help

"*Biased fragmentation: Any library which is generated based on the ligation of random hexamers or through tagmentation should theoretically have good diversity through the sequence, but experience has shown that these libraries always have a selection bias in around the first 12bp of each run. This is due to a biased selection of random primers, but doesn't represent any individually biased sequences. Nearly all RNA-Seq libraries will fail this module because of this bias, but this is not a problem which can be fixed by processing, and it doesn't seem to adversely affect the ablity to measure expression.*"

The Sequence Duplication Levels seem of for RNA-seq too. Possibly this is normal for RNA-seq libraries. Seems like yes this is ok for RNA-seq and represents highly expressed transcripts and an assay at saturation. Here is an example library from run1



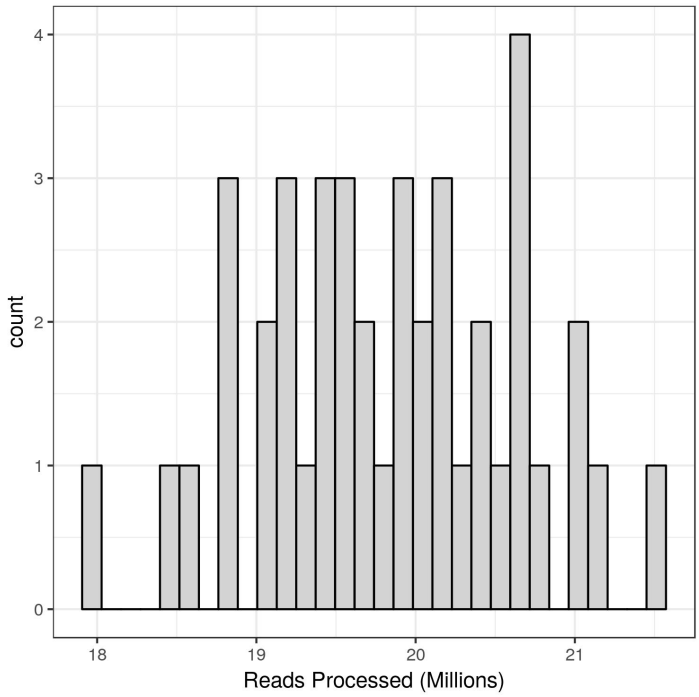
Did not trim or deduplicate the data. Went right to aligning and kallisto quantification b=100 to GRCh38.

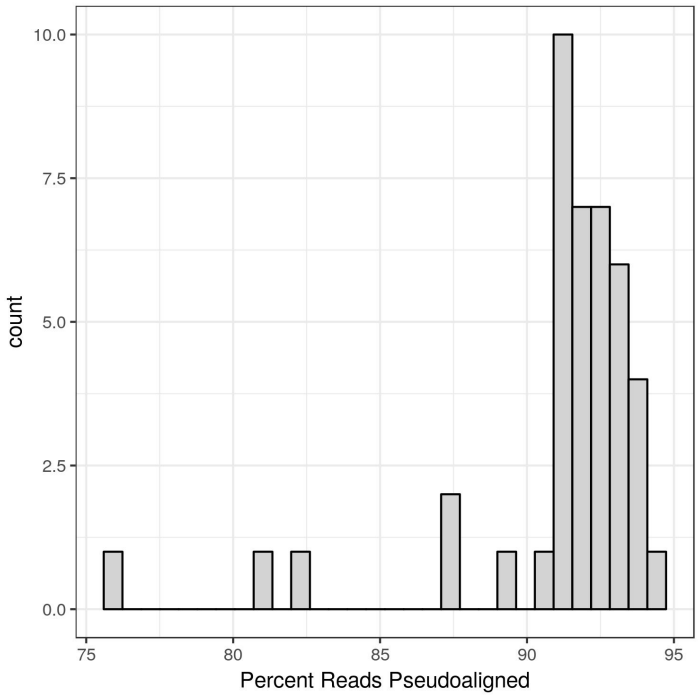
The data is "single-end sequencing to generate 10 - 15 million reads per sample for differential gene expression analysis only" and each sample is included in two runs. These files will be merged and kallisto run single-ended. Sequencing was done on NextSeq 75 cycle high output run (ie 75bp read length) and the kit was Truseq mRNA.

The kallisto needs a read length, which is 76 according to fastQC, but also a standard deviation of that read length.

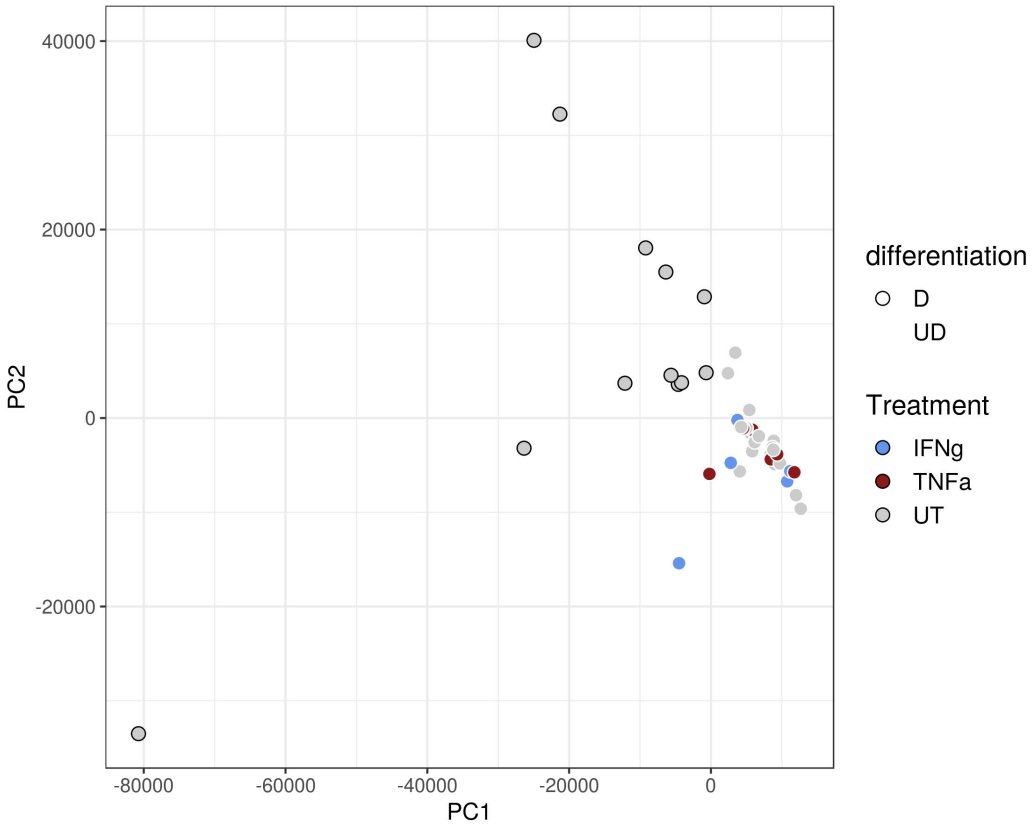
I was able to calculate these from the multQC output (by exporting the data behind the fragment length distribution plot). For both runs the mean read length (should be fragment length but that you can't get from the sequencing data) is 75.49 and the sd is 1.48. I will use 75 as it is NextSeq 75 cycle and sd of 1.5 for simplicity.

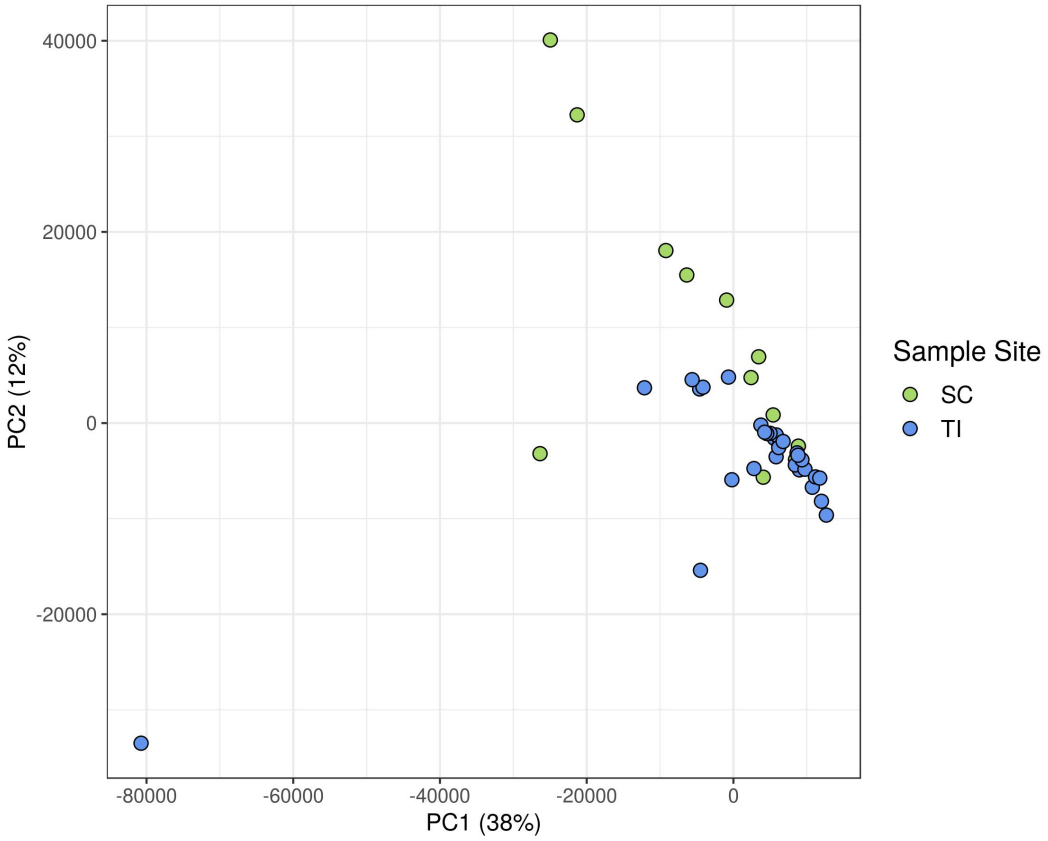
On the merged files the alignment percent from kallisto is very high.

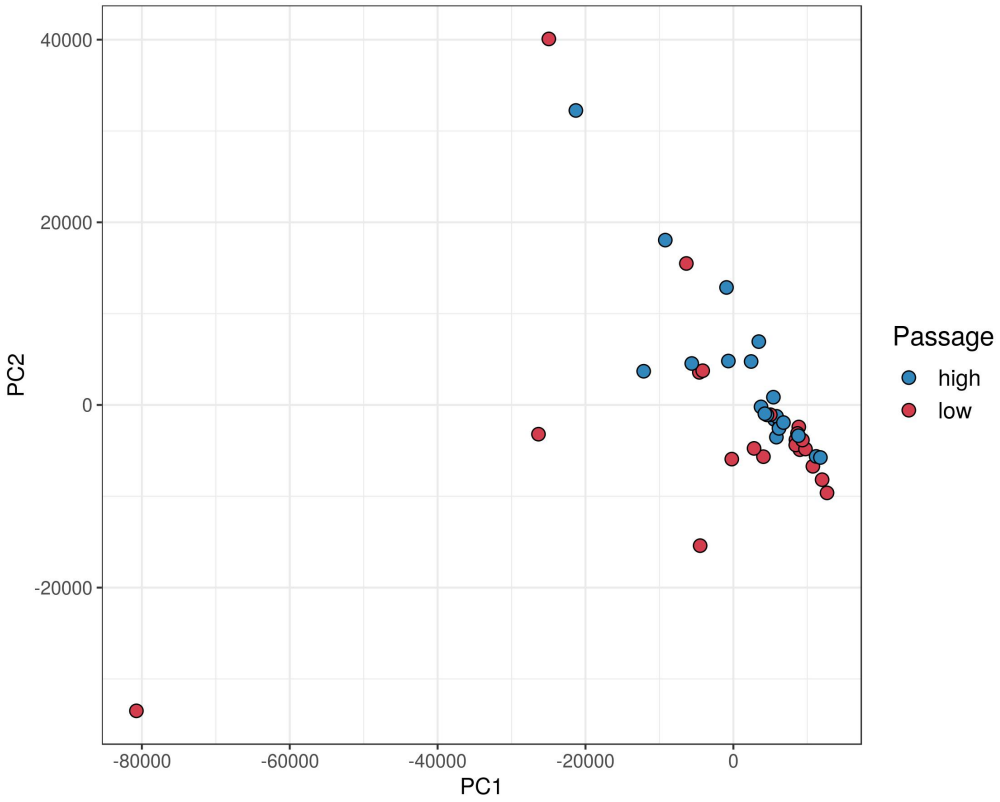


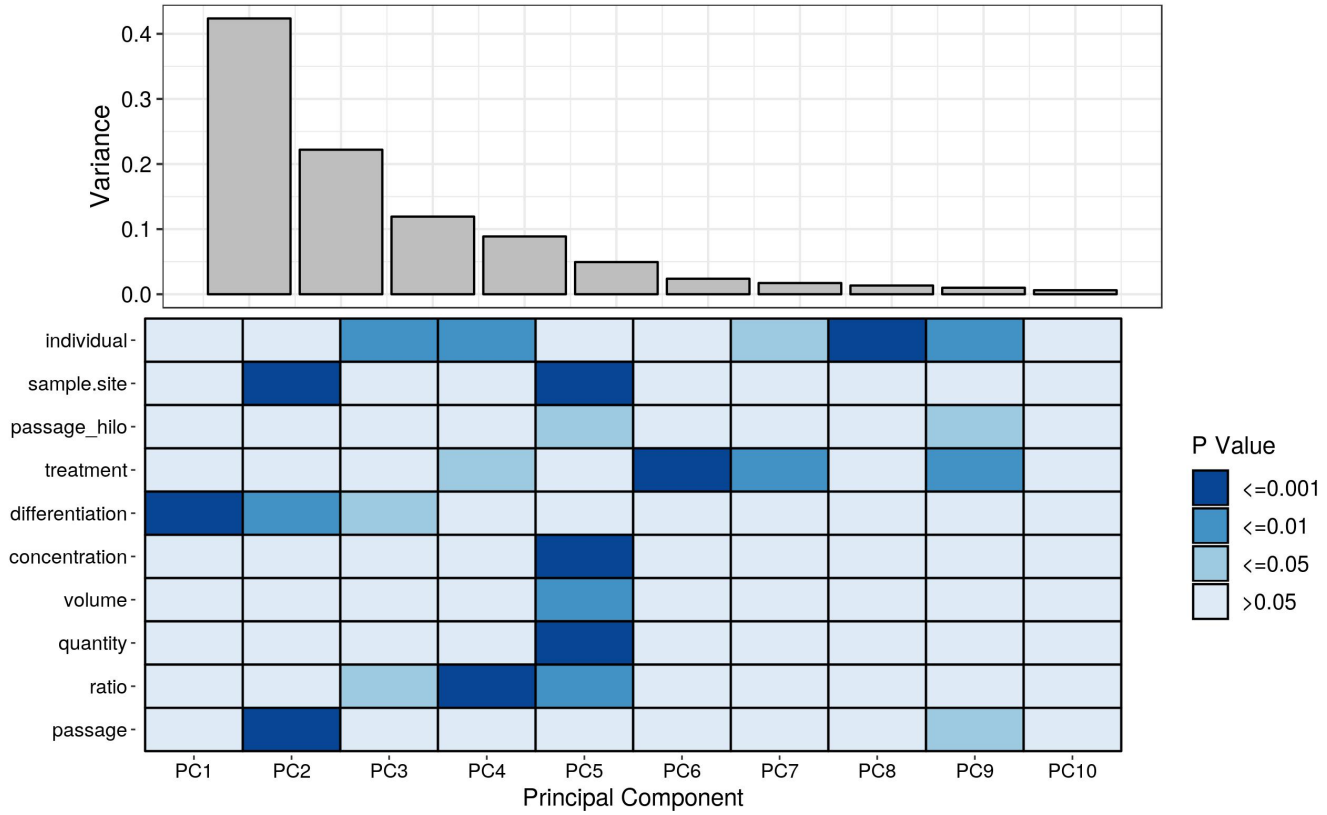


RNAseq PCA

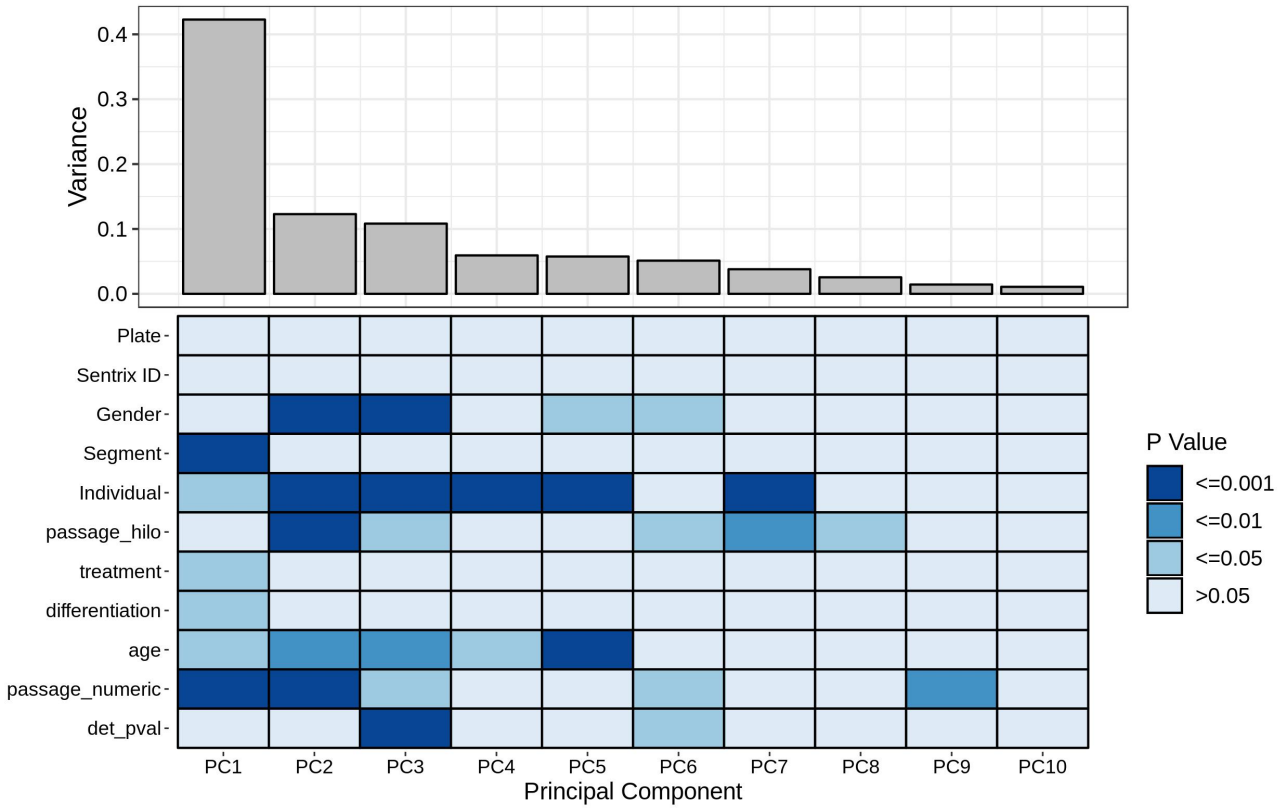

40000
20000
0
-20000
PC2
-80000
-60000
-40000
-20000
0
PC1
differentiation
D
UD
Treatment
IFNg
TNFa
UT

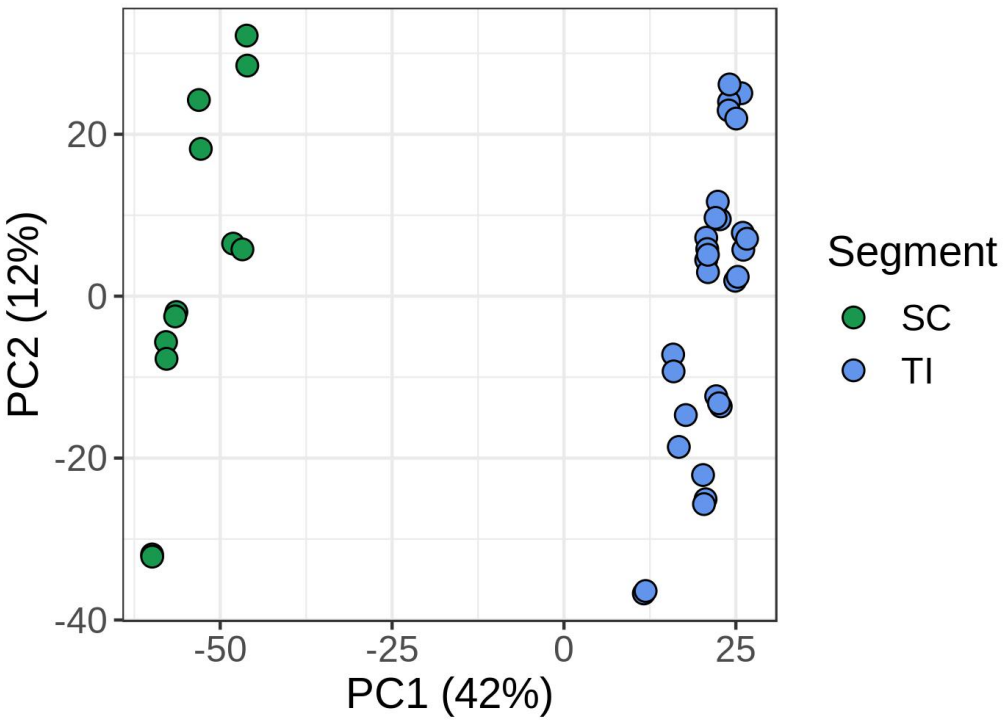

40000
20000
0
-20000
PC2 (12%)
-80000
-60000
-40000
-20000
0
PC1 (38%)
Sample Site
SC
TI

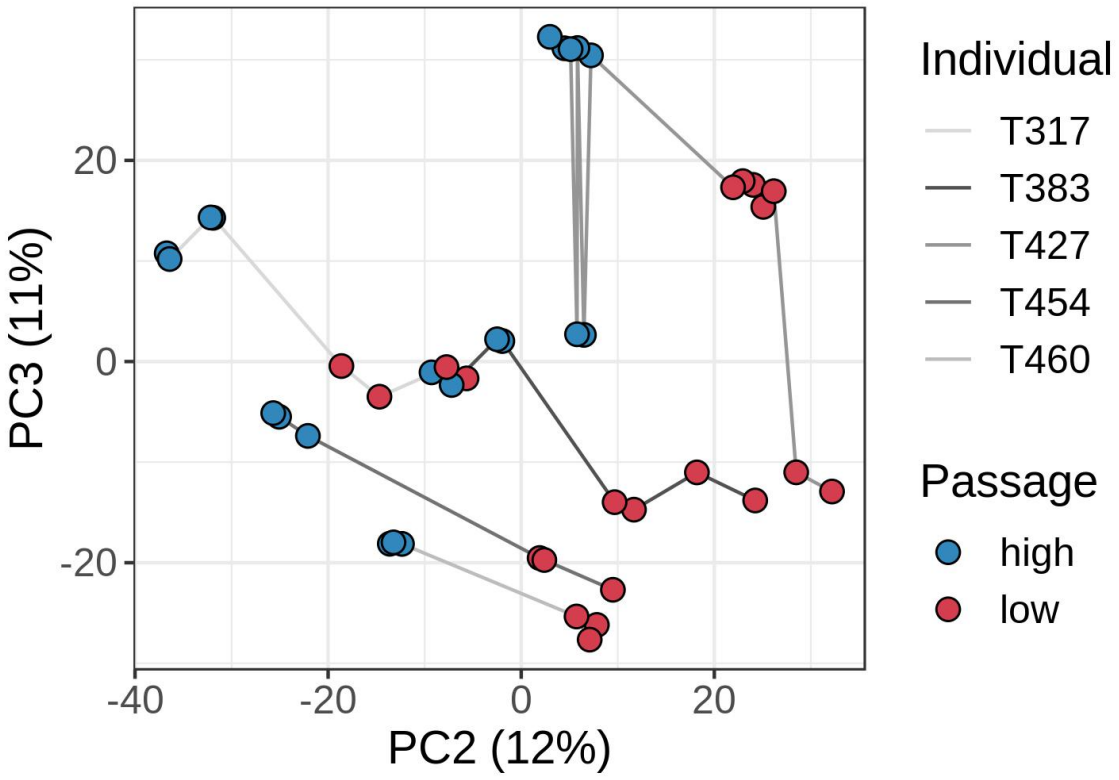


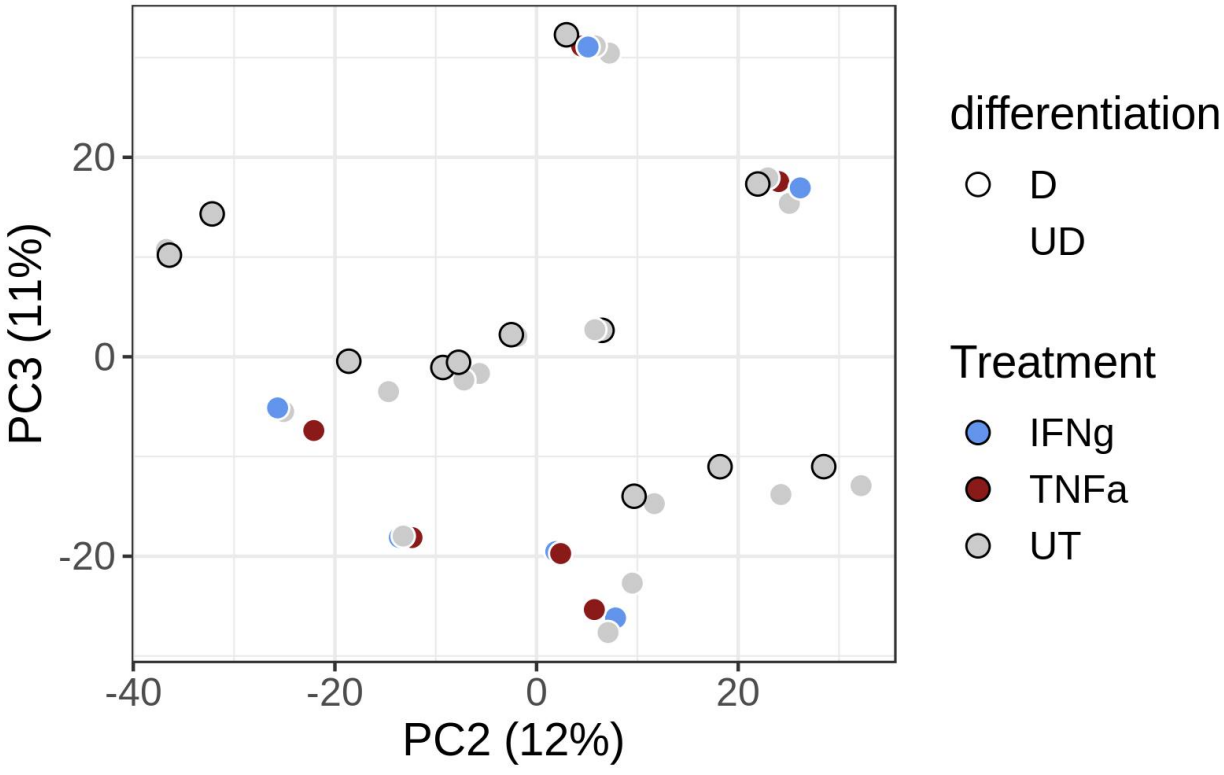


DNA Methylation

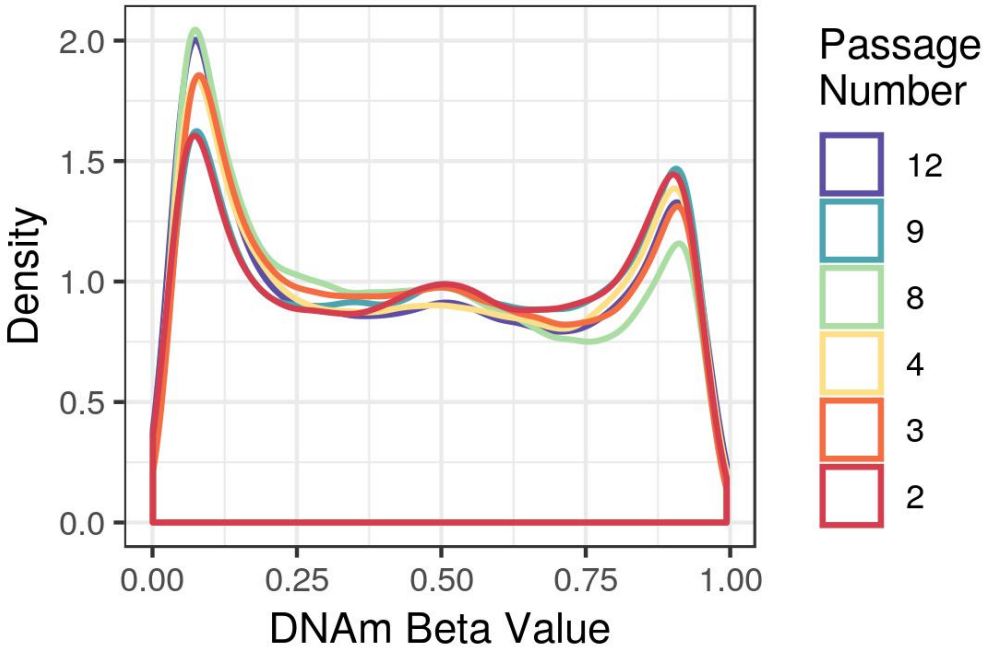

Variance
0.4
0.3
0.2
0.1
0.0
Plate
Sentrix ID
Gender
Segment
Individual
passage_hilo
treatment
differentiation
age
passage_numeric
det_pval
PC1
PC2
PC3
PC4
PC5
PC6
PC7
PC8
PC9
PC10
Principal Component
P Value
<=0.001
<=0.01
<=0.05
>0.05

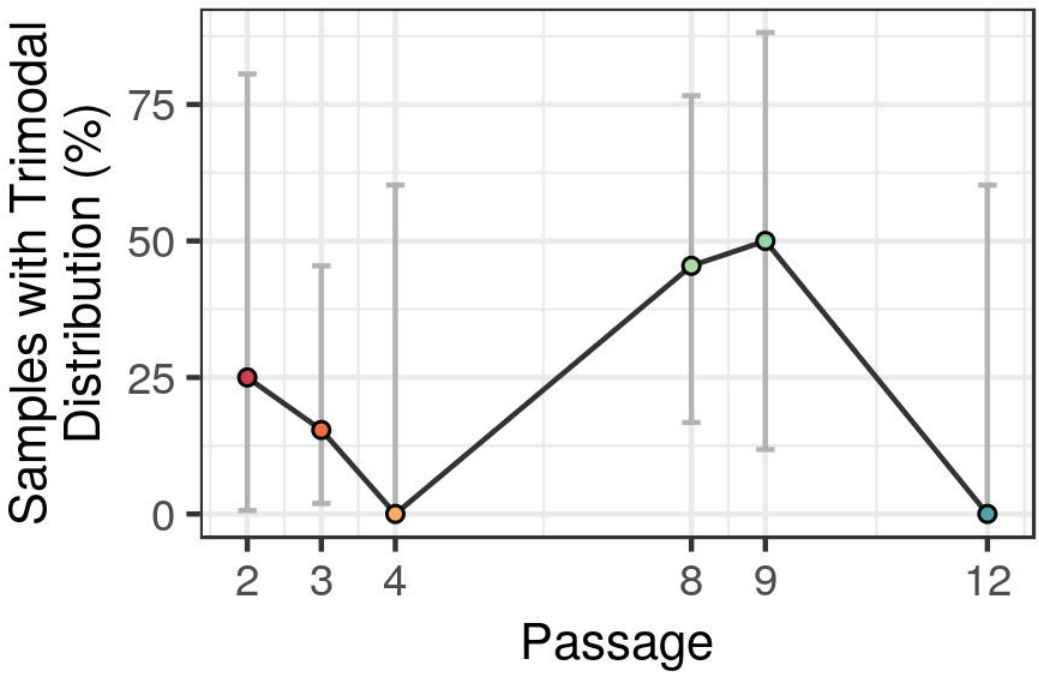

PC2 (12%)
20
0
-20
-40
-50
-25
0
25
PC1 (42%)
Segment
SC
TI



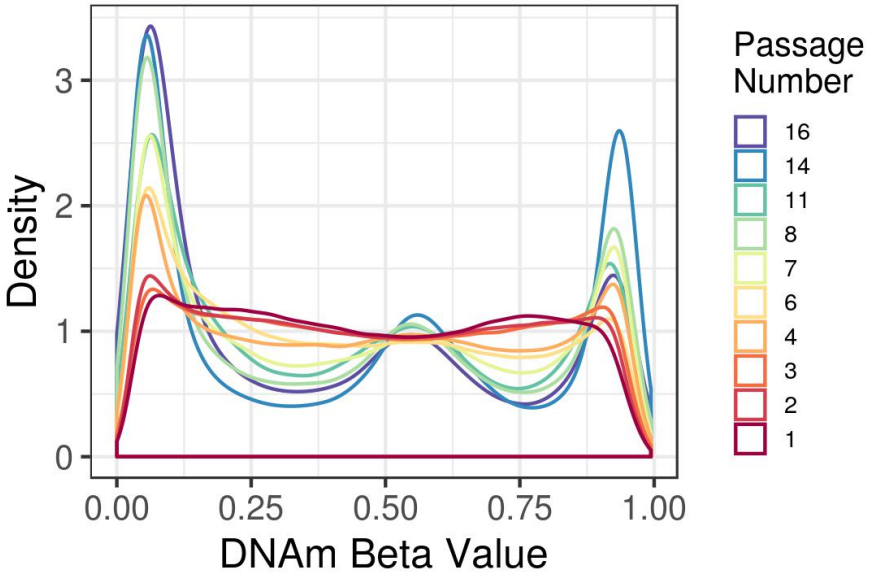


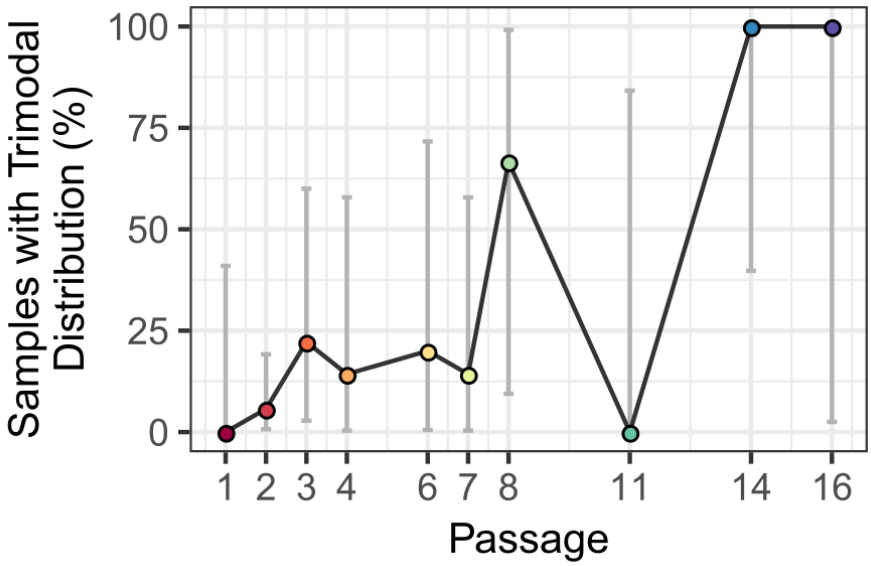
Validation Cohort



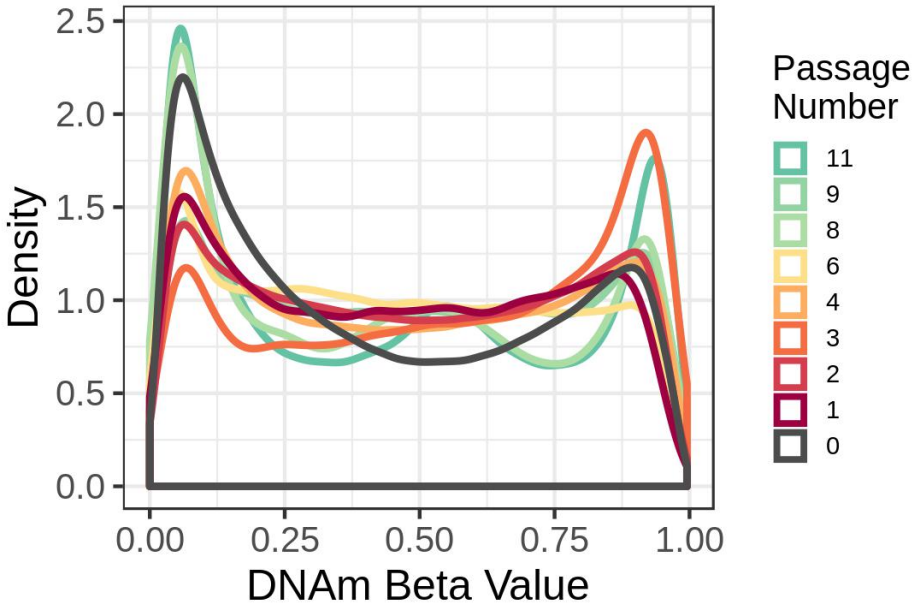


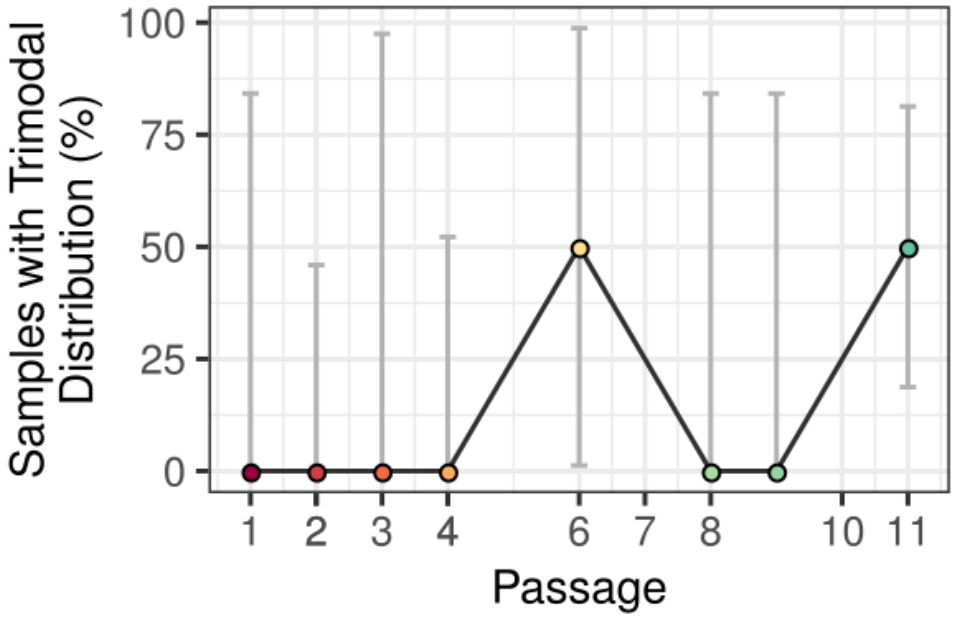
Original Cohort


Density
3
2
1
0
0.00
0.25
0.50
0.75
1.00
DNAm Beta Value
Passage Number
16
14
11
8
7
6
4
3
2
1

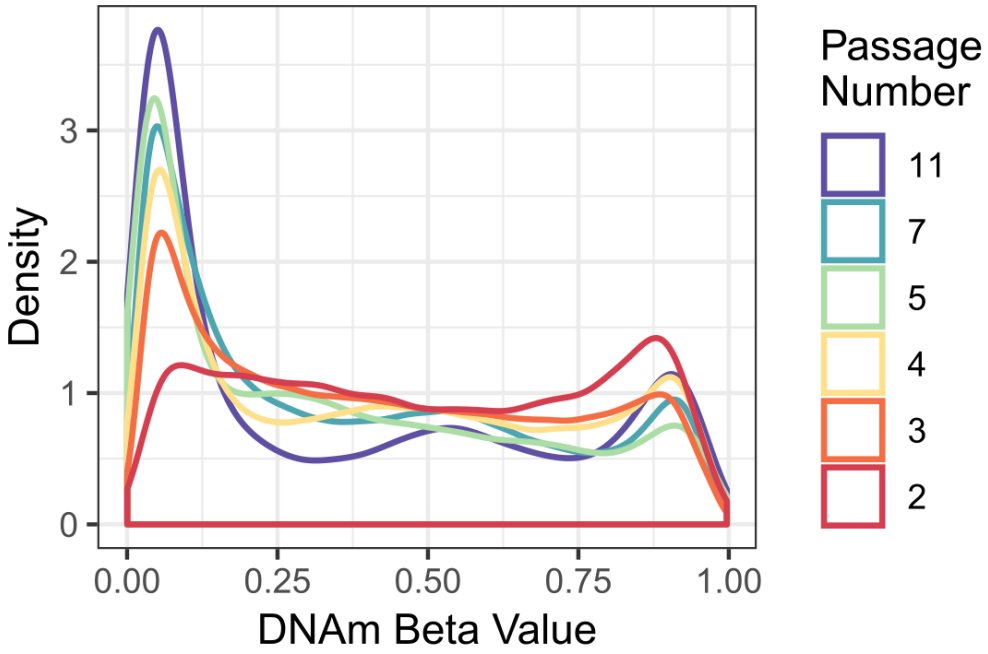

Samples with Trimodal Distribution (%)
100
75
50
25
0
1 2 3 4 6 7 8 11 14 16
Passage

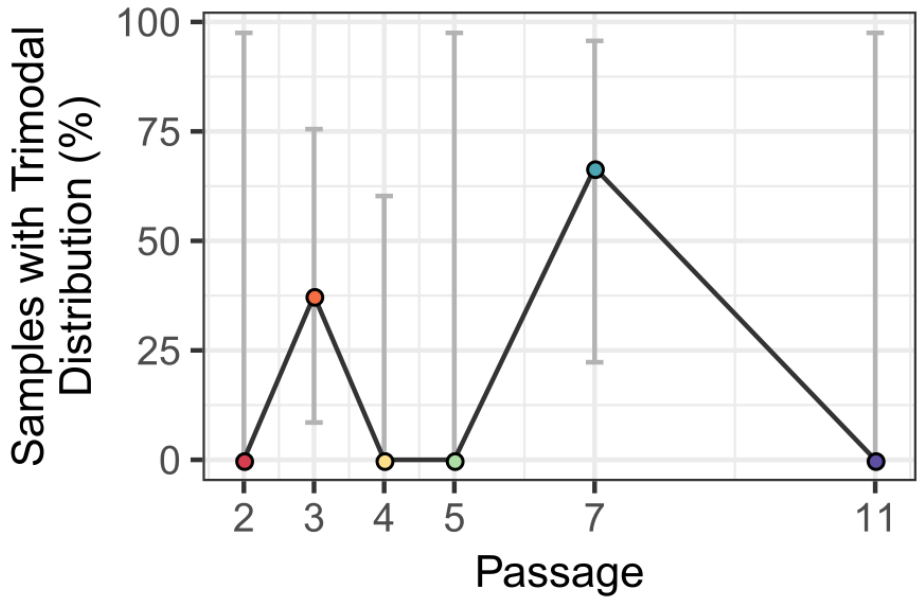
MTAB4957


Density
2.5
2.0
1.5
1.0
0.5
0.0
0.00
0.25
0.50
0.75
1.00
DNAm Beta Value
Passage Number
11
9
8
6
4
3
2
1
0

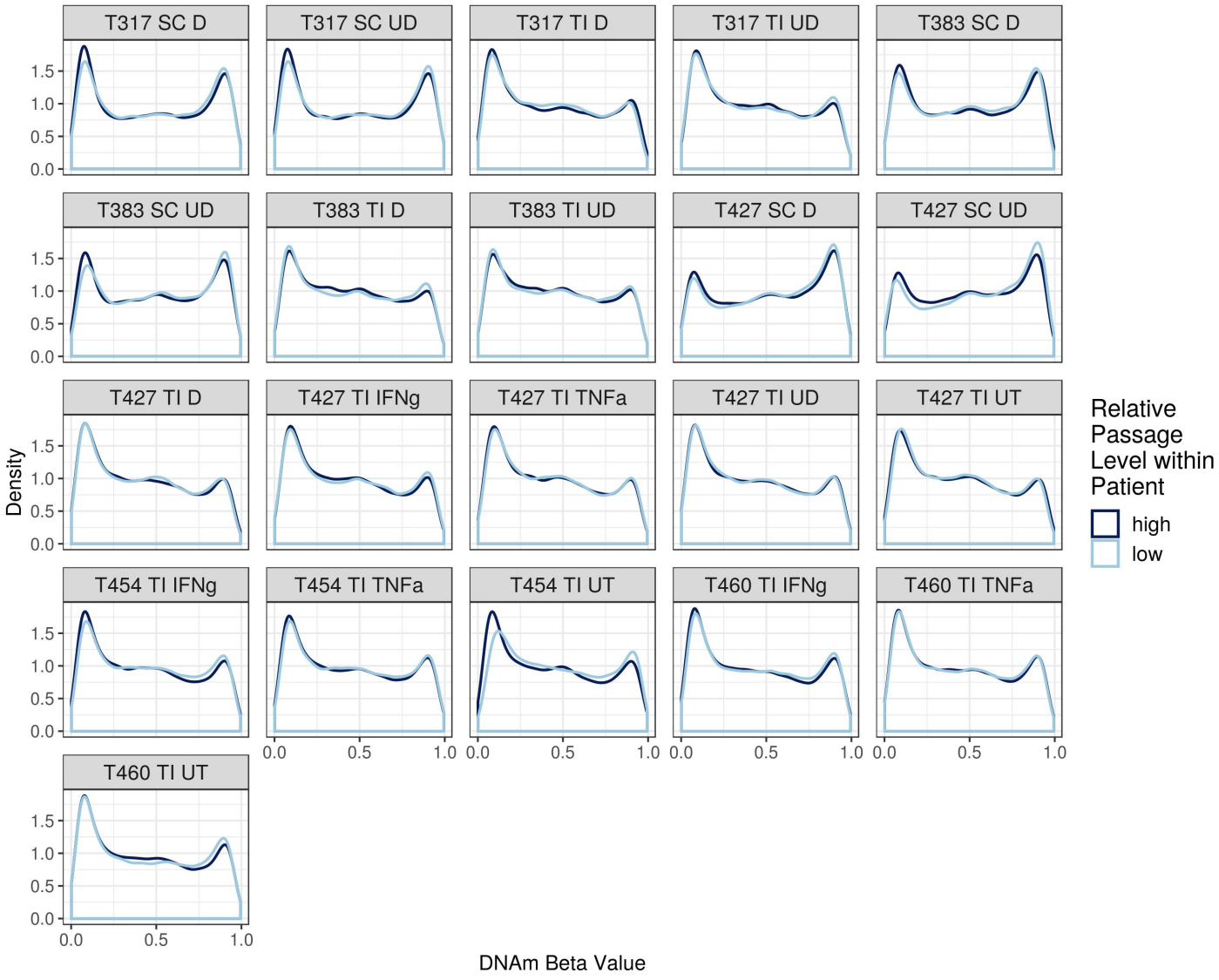

Samples with Trimodal Distribution (%)
100
75
50
25
0
1
2
3
4
6
7
8
10
11
Passage

GSE141256



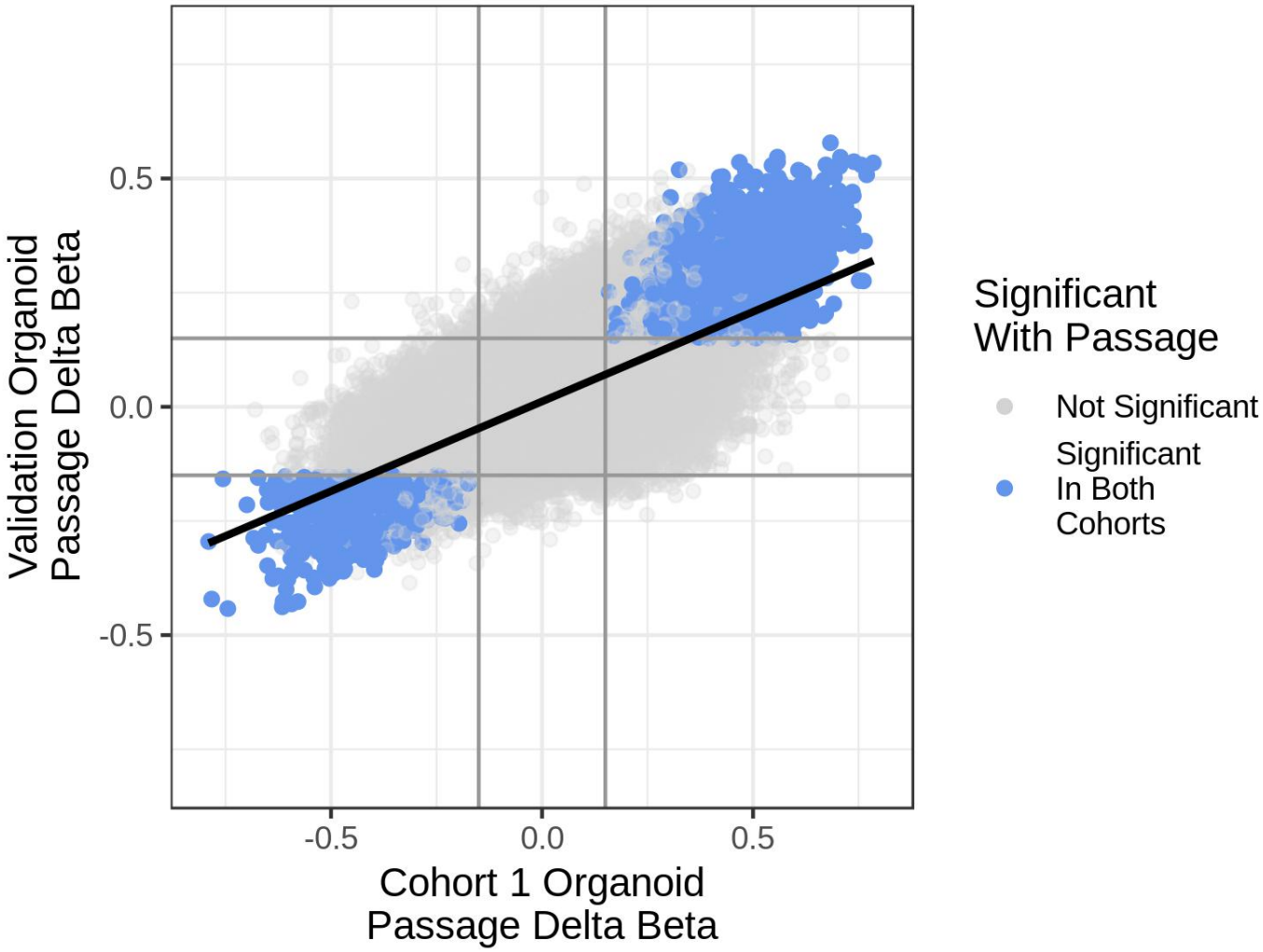


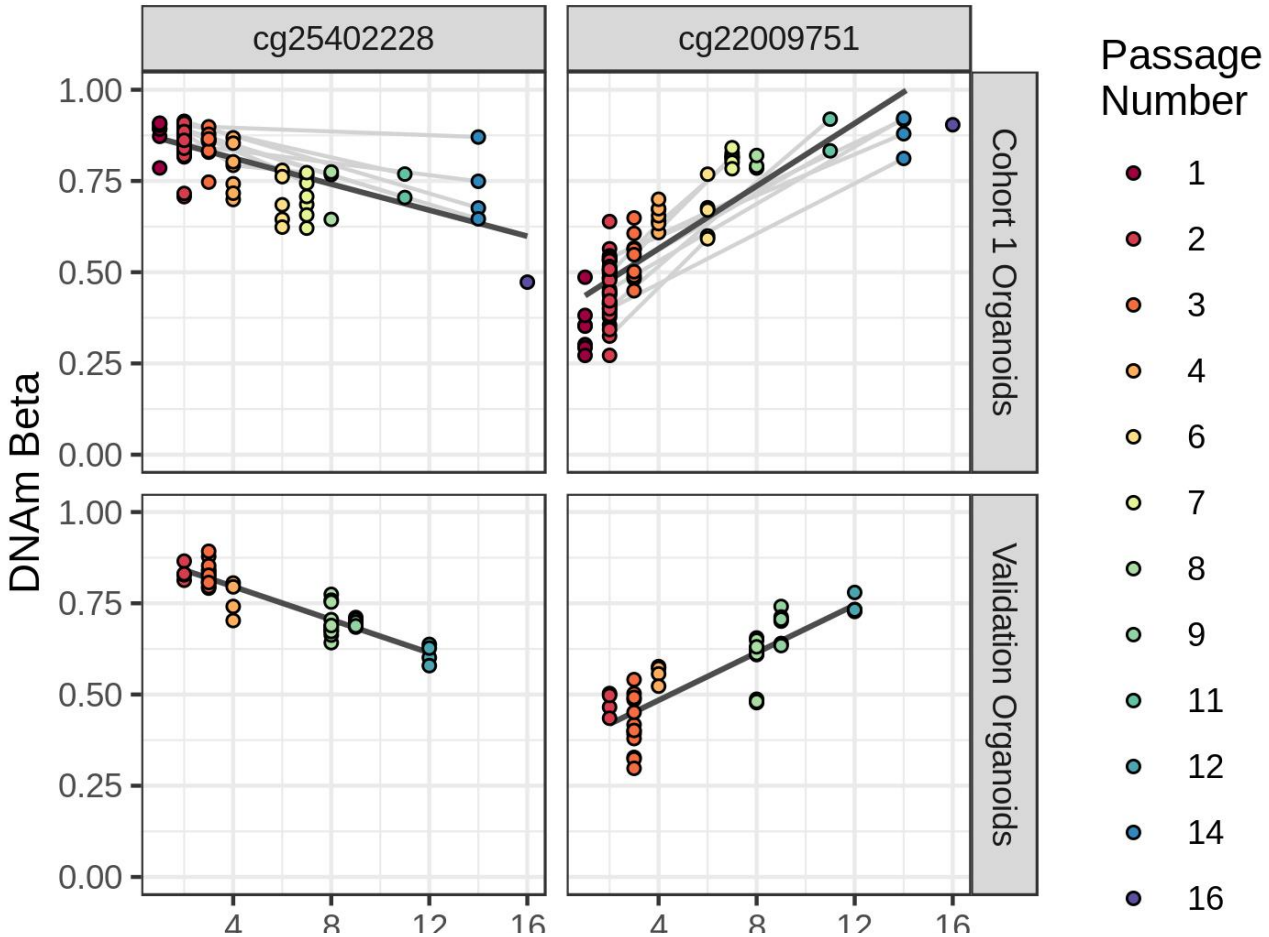
Validation data paired samples



Individual CpG Validation

Looking first at just untreated undifferentiated organoids (comparable conditions to Cohort 1, n=18), of the 23,766 CpGs differentially DNAm with passage in Cohort 1, 23,737 were measured in the validation cohort. These are split into 17,327 hypomethylated and 6,410 hypermethylated CpGs, of which 54% and 7%, respectively, were also significantly different with passage in the validation cohort

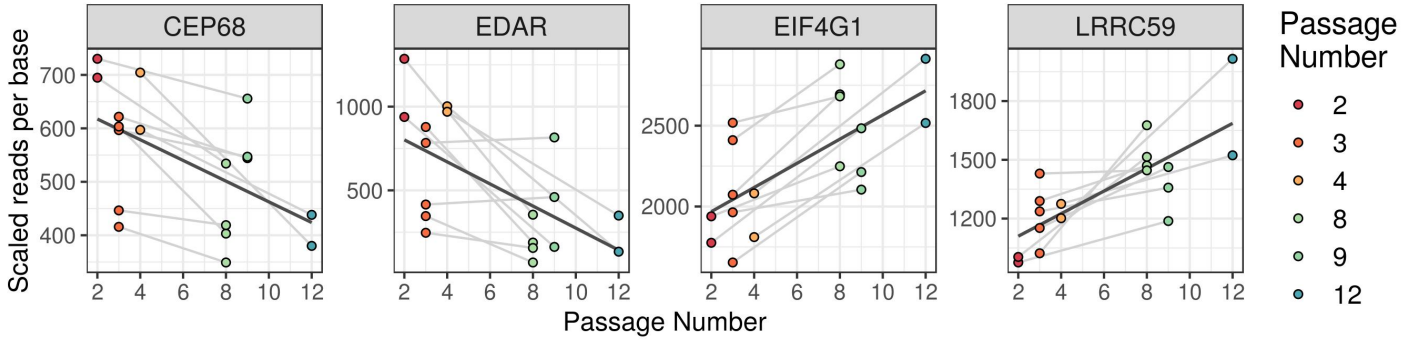

Validation Organoid
Passage Delta Beta
0.5
0.0
-0.5
-0.5
0.0
0.5
Cohort 1 Organoid
Passage Delta Beta
Significant
With Passage
Not Significant
Significant
In Both
Cohorts

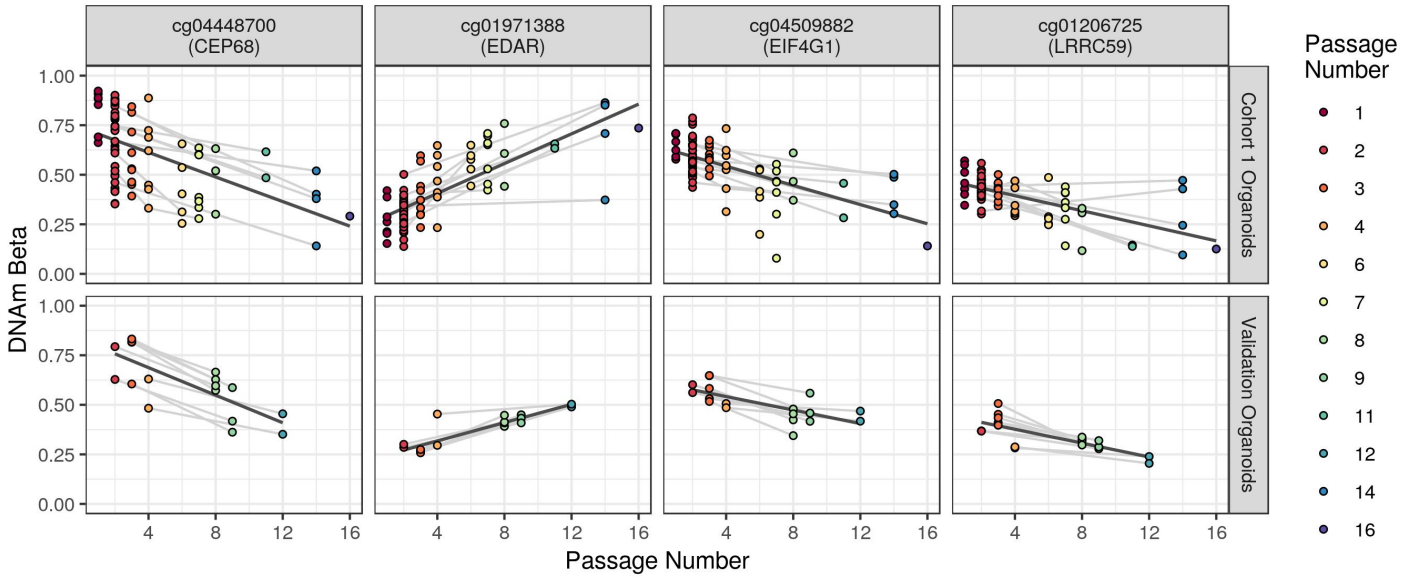

cg25402228
cg22009751
Cohort 1 Organoids
Validation Organoids
DNAm Beta
1.00
0.75
0.50
0.25
0.00
4
8
12
16
Passage
Number
1
2
3
4
6
7
8
9
11
12
14
16

Passage Number

Genes Differentially Expressed and DNAm

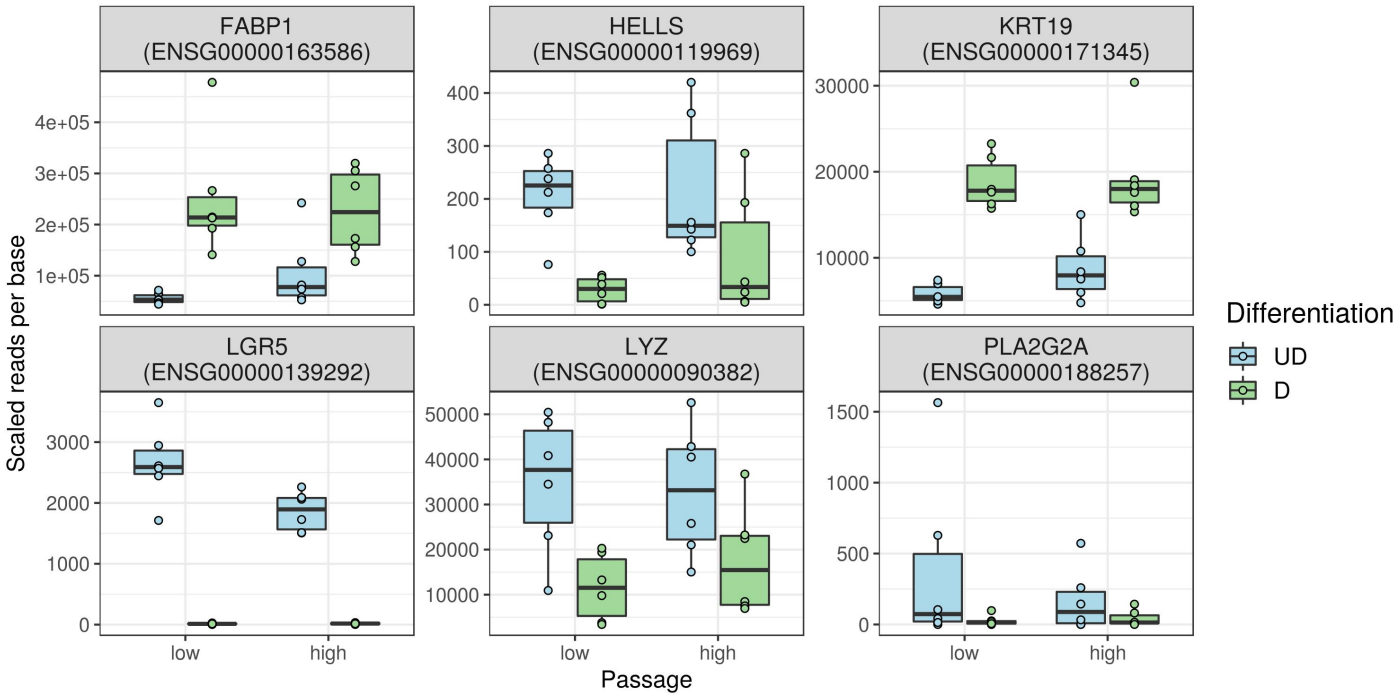
Of the 4,722 hypo and 402 hyper methylated genes in the original and validation DNAm data, 10 were differentially expressed (8 hypo, 2 hyper) of 114 which were differentially expressed with passage. Four are shown below.
Of the 8 hypo, 6 become more expressed with passage (ie LRRC59, EIF4G1) and 2 become less expressed (ie CEP68). Of the hypermethylated genes one increases expression one lowers expression (i.e EDAR).





Differentiation

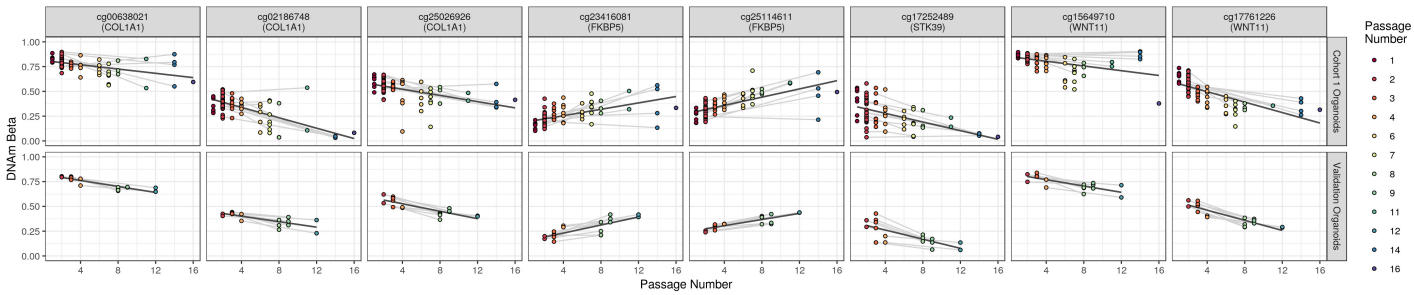
In low passage organoids there are 11,088 genes differenitally expressed with differentiation. In high passage there are 9,429 genes. Established markers of differentiation ("ASCL2","MKI67","LGR5","CA2","OLFM4","LYZ","MUC2","MUC1","CYP3A4","PLA2G2A","FABP1","KRT19","HELLS", "SPINK4","FCGBP","NEAT1","TOP2A") are significantly differentially expressed in both low and high passage. Except LYZ which is only significant in low passage, but is a paneth cell marker so should be more highly expressed so maybe not a good marker to highlight.

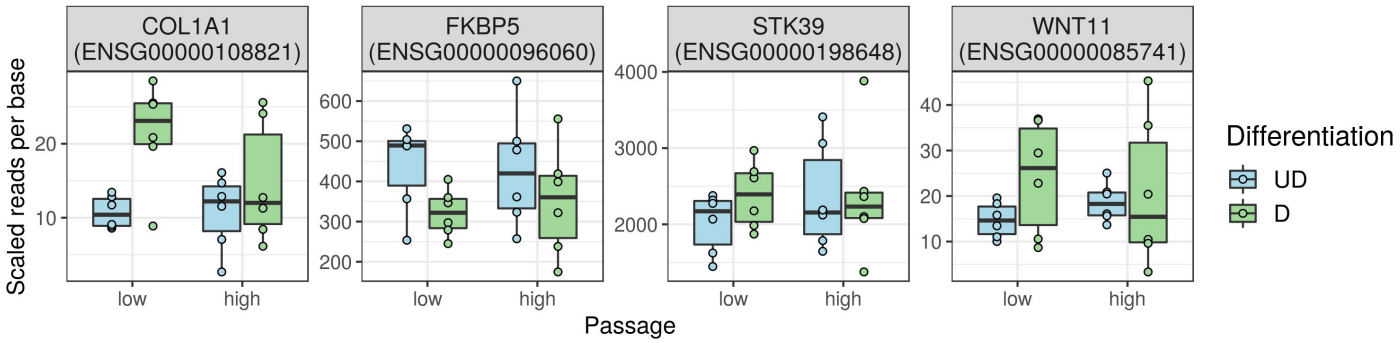


Of the 2,075 genes only differential in low passage, 175 were also differentially DNAm with passage. They are not differential DNAm with differentiation, as there were basically no CpGs different with differentiation (need to summarize that analysis later).

Genes of note (differentially expressed but not DNAm): Several WNT pathway gene only significant in low "DISC1" "AMER1" "WNT8B" "FRAT2" "MED12" "TCF7L1" "BCL9L" "KLF4" "RECK" "WNT11" "FZD8" "RARG"
Also VEGFB

Some the the representative genes in the 175 differenital DNAm with passage and differentailly expressed in low only

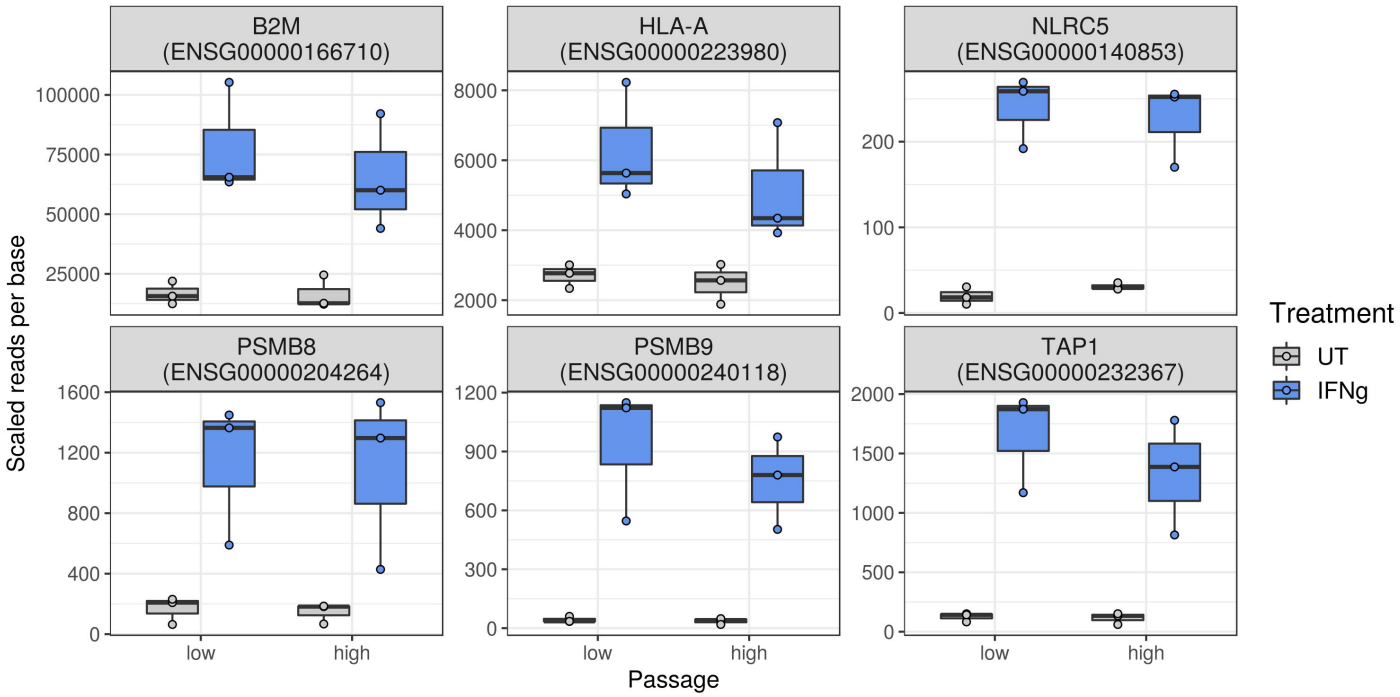




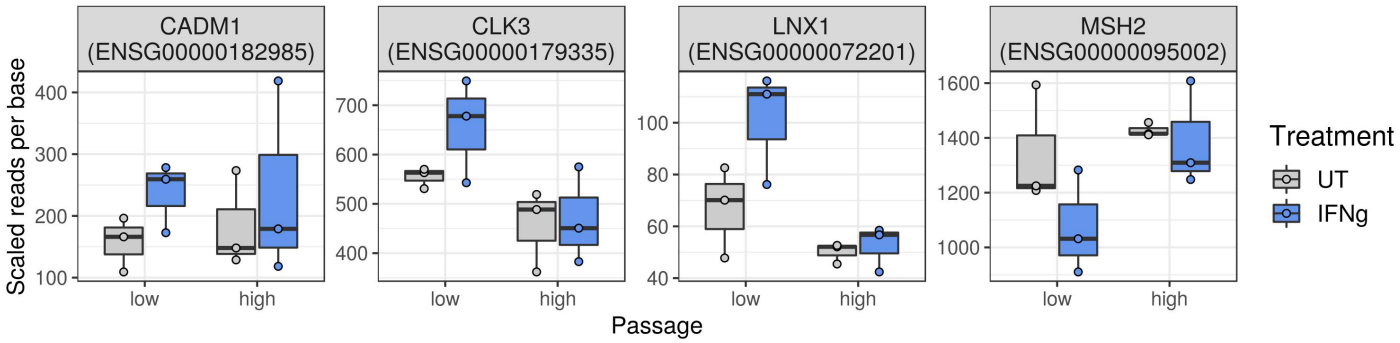
Proinflammatory Cytokine Treatments

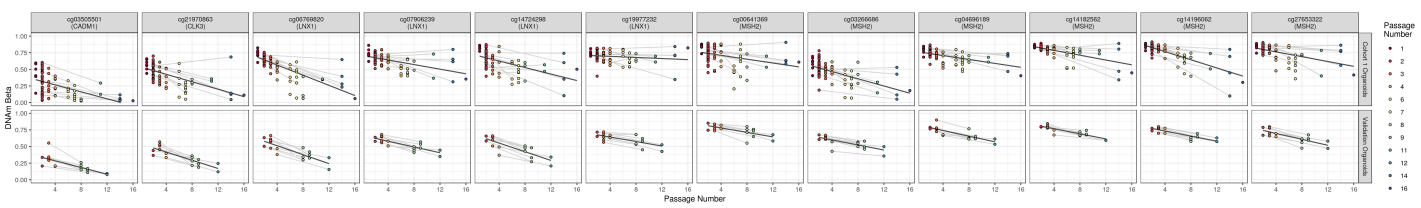
Looking at **IFNg** treatment, 6,425 and 4,565 genes changed with treatment in low and high passage organoids repectively. All candidate MHCI genes changed in both conditions
'HLA-F', 'HLA-G', 'HLA-A', 'HLA-E', 'HLA-C', 'HLA-B',"TAP1","TAP2","PSMB9","PSMB8","B2M","MR1","CD1D","IRF1","NLRC5"

Here I am showing the gene version with the lowest pvalue for each gene as some genes have many IDs.



Of the 2,781 genes differentially expressed in low but not high passage organoids, 194 were also differentially DNAm with passage. Like differentiation these were not differentially DNAm with treatment, but that will be fleshed out below. Here are some representative genes.





Looking at **TNFa** treatment 1,975 and 2,299 genes changed with treatment in low and high passage organoids respectively.

Of the 571 genes differentially expressed in low but not high passage organoids, 59 were also differentially DNAm with passage. Like differentiation these were not differentially DNAm with treatment, but that will be fleshed out below. Here are some representative genes.