

Supplementary Figures for: Dysregulation of DNA Methylation in Human Intestinal Epithelial Organoids with Increasing Passage

Rachel Edgar

June 1, 2020

Table S1. Top over-represented GO groups in the list of genes adjacent to heteroskedastic CpGs. Columns are: GO ID, name of the GO gene set, number of genes annotated to that gene set and associated to a heteroskedastic CpG, Benjamini-Hochberg corrected enrichment p value (which the data is sorted by), Multifunctionality (MF) score and the names of genes annotated to that gene set and associated to a heteroskedastic CpG.

Table S2. Top over-represented GO groups in the list of genes adjacent to hypomethylated CpGs. Columns are: GO ID, name of the GO gene set, number of genes annotated to that gene set and associated to a hypomethylated CpG, Benjamini-Hochberg corrected enrichment p value (which the data is sorted by), Multifunctionality (MF) score and the names of genes annotated to that gene set and associated to a hypomethylated CpG.

Table S3. Top over-represented GO groups in the list of genes adjacent to hypermethylated CpGs. Columns are: GO ID, name of the GO gene set, number of genes annotated to that gene set and associated to a hypermethylated CpG, Benjamini-Hochberg corrected enrichment p value (which the data is sorted by), Multifunctionality (MF) score and the names of genes annotated to that gene set and associated to a hypermethylated CpG.

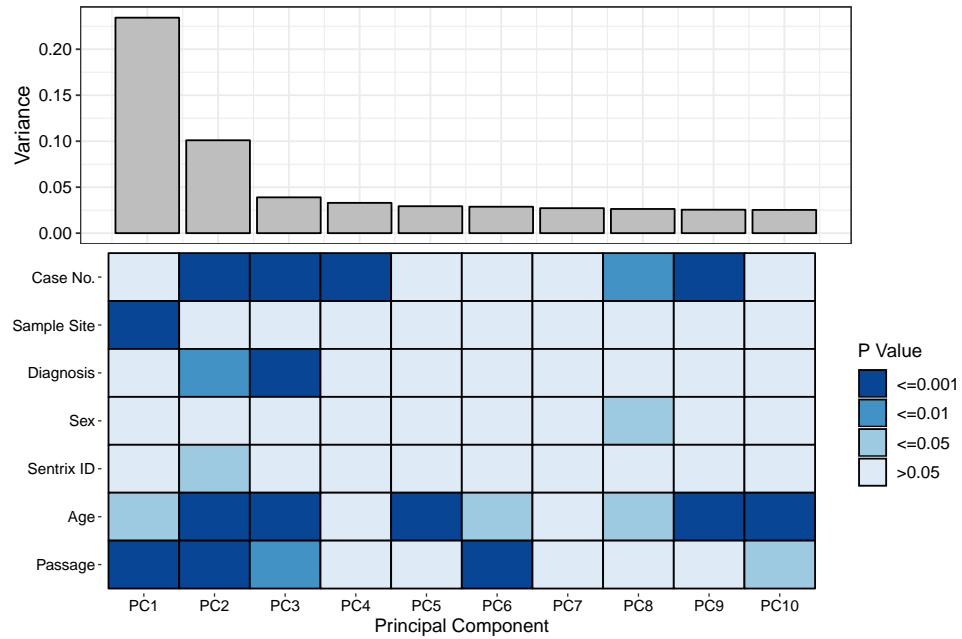


Figure S1: Sample site and passage contribute to the major components of variance in DNAm. The scree plot shows the amount of DNAm variance accounted for by each PC. The heat map shows the associations between sample variables and each PC. P values were generated with a Spearman correlation for continuous variables or an ANOVA for categorical.

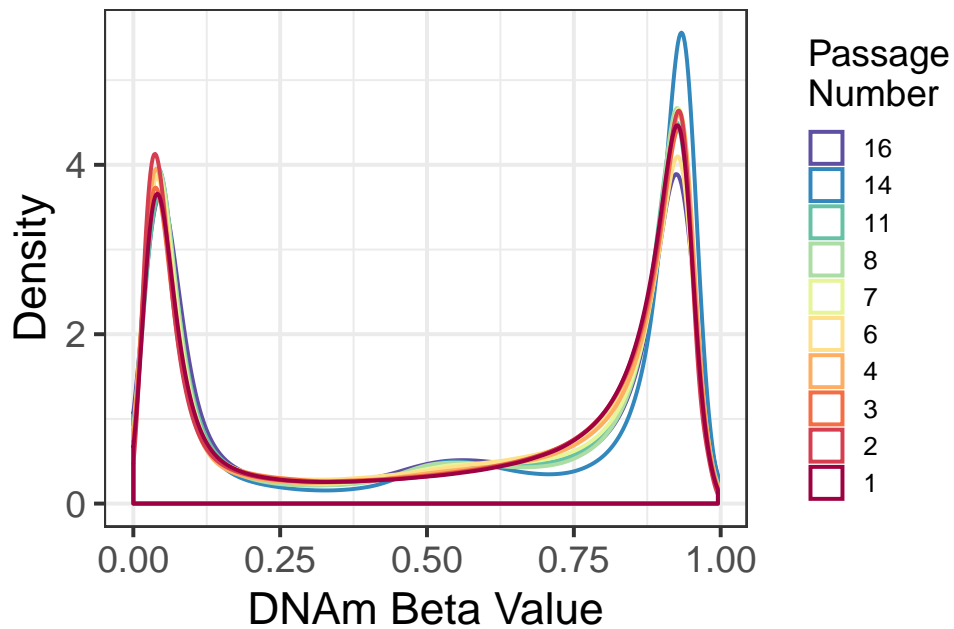


Figure S2: The DNAm beta value distribution for high passage samples is trimodal while for low passage samples it is bimodal. All beta distributions displayed are for all 800,383 CpGs. (a) Beta value distributions for all samples at a given passage number. Curves are coloured by the passage number.

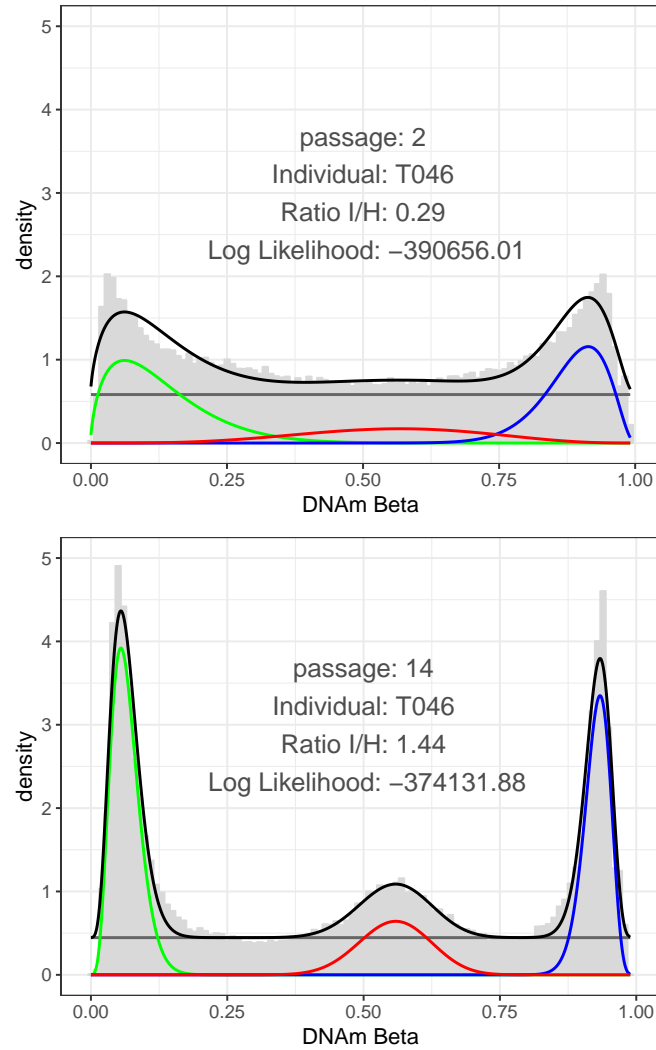


Figure S3: Fit of the mixture of beta-binomial distributions in two representative samples, one bimodal and one trimodal. Both samples shown are from one individual which were passaged 2 and 14 times. The grey histogram represents the actual beta value distribution, the coloured curves are the three beta-binomial distributions fit and the horizontal grey line represents the background uniform distribution. Ratio I/H is the ratio of the intermediate distribution peak to the background uniform distribution, this is the ratio for selection of trimodal samples.

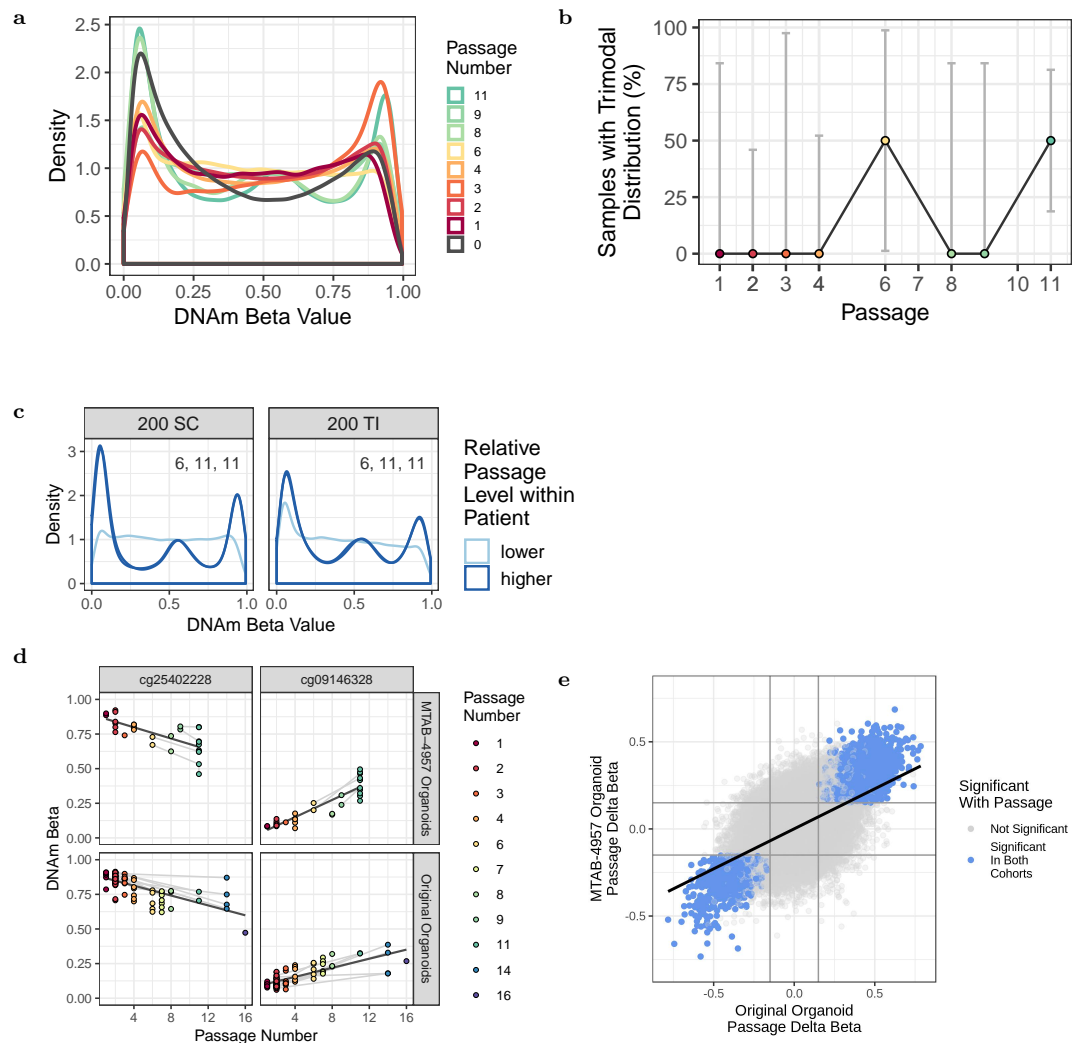


Figure S4: In E-MTAB-4957 pediatric organoids the DNAm beta value distribution for high passage samples is trimodal while for low passage samples it is bimodal. All beta distributions displayed are for the 47,924 most variable CpGs. (a) Beta value distributions for all samples at a given passage number. Curves are coloured by the passage number. Included is the distribution for primary samples for comparison where passage is labelled as 0. (b) The percent of samples considered trimodal. Points are coloured by passage and the bars around each point represent the confidence interval for that percent. (c) Beta value distributions for samples derived from the same patient but cultured to a different number of passages. Plots are labelled with the patient ID number, sampling site of origin and the passage number of each sample derived from that patient and sampling site. Curves are coloured by high or low passage relative to the other sample(s).

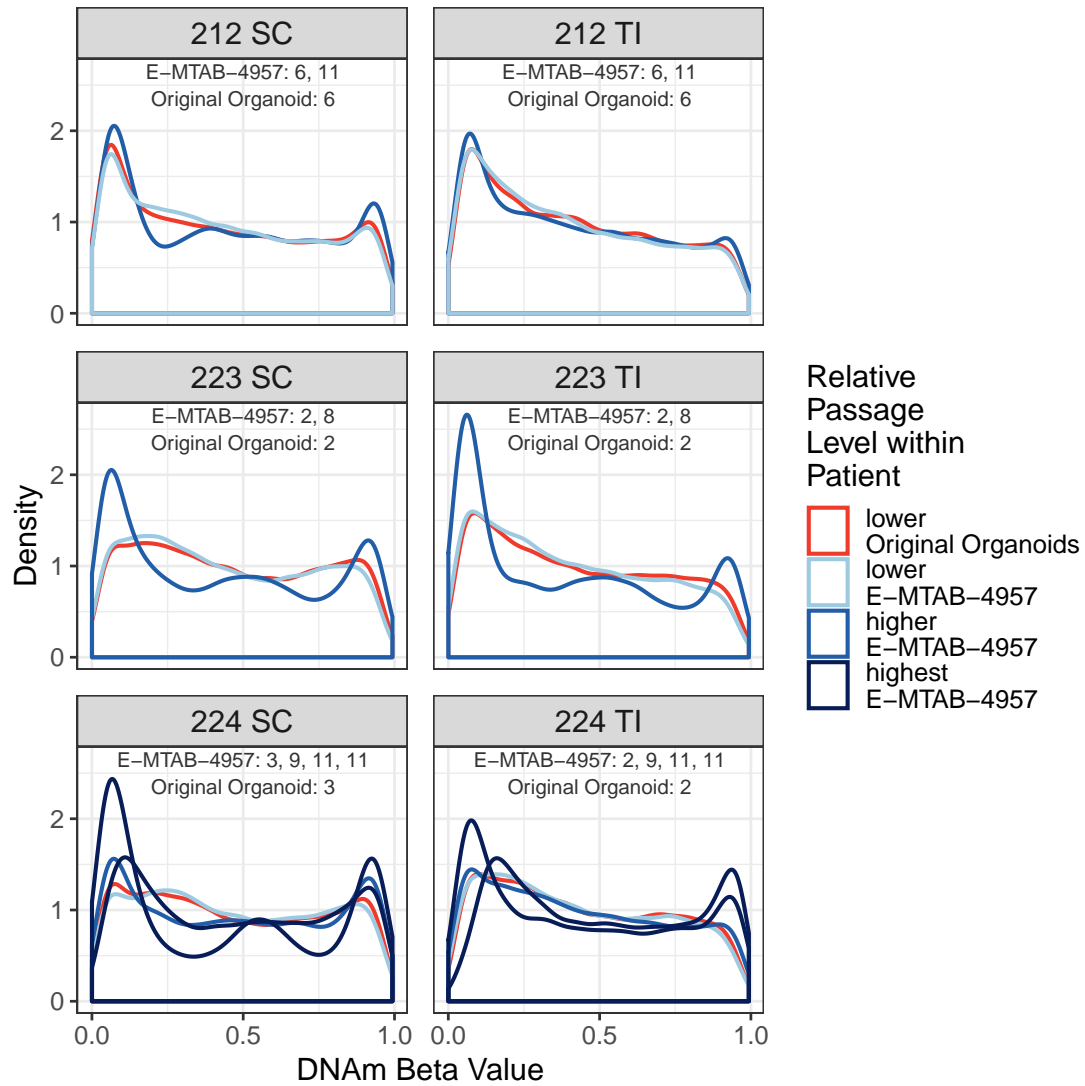


Figure S5: In E-MTAB-4957 pediatric organoids the DNAm beta value distribution for high passage samples is trimodal while for low passage samples it is bimodal. Beta value distributions for samples derived from the same patient but cultured to a different number of passages. Samples are shown which here if an individual's samples were run in both our organoids and E-MTAB-4957. All beta distributions displayed are for the 19,452 variable CpGs on both the 450K and EPIC. Plots are labelled with the patient ID number, sampling site of origin and the passage number of each sample derived from that patient and sampling site split by study. Curves are coloured by high or low passage relative to the other sample(s) and samples from our 165 organoid cohort are in red.

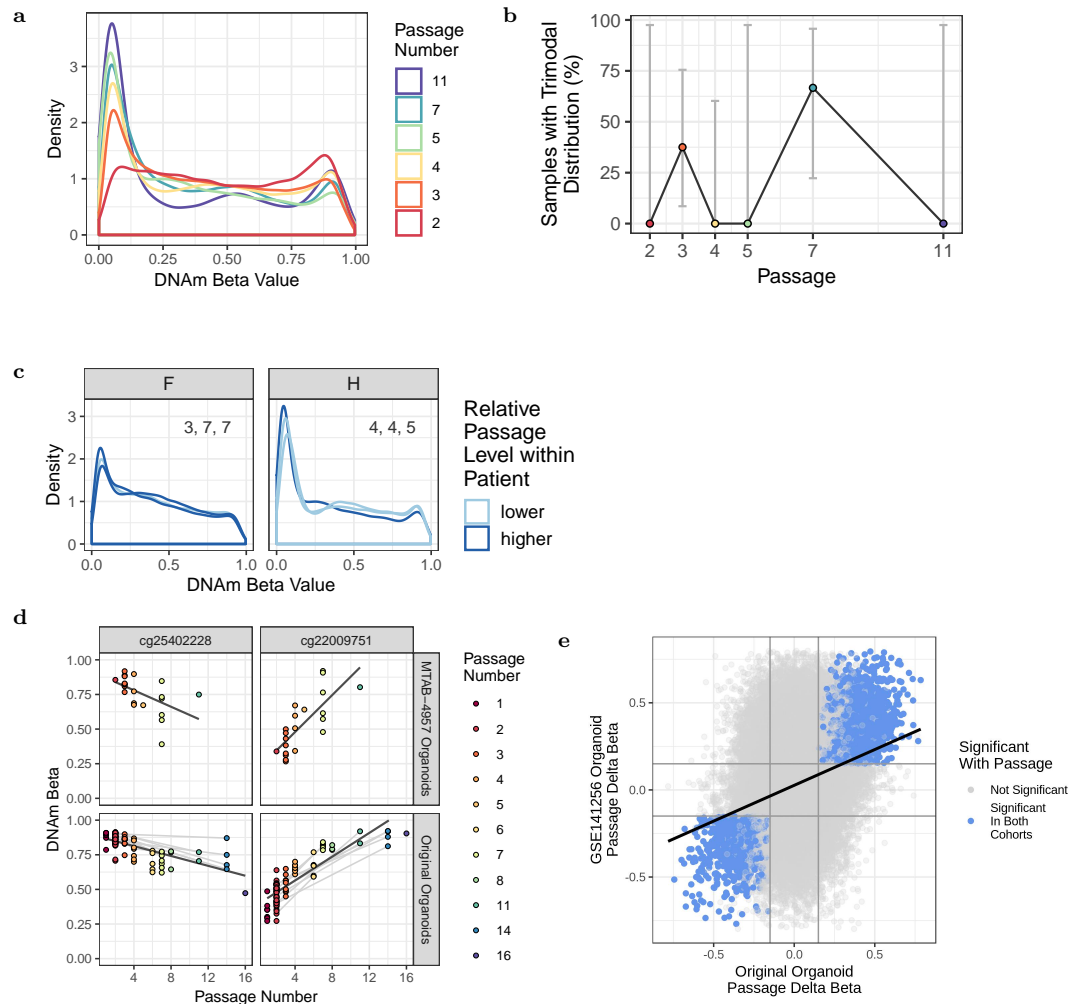


Figure S6: In GSE141256 organoids the DNAm beta value distribution for high passage samples is trimodal while for low passage samples it is bimodal. All beta distributions displayed are for the 55,379 most variable CpGs in fetal organoids. (a) Beta value distributions for all samples at a given passage number. Curves are coloured by the passage number. (b) The percent of samples considered trimodal. Points are coloured by passage and the bars around each point represent the confidence interval for that percent. (c) Beta value distributions for samples derived from the same patient but cultured to a different number of passages. Plots are labelled with the patient ID and the passage number of each sample derived from that patient. Curves are coloured by high or low passage relative to the other sample(s).

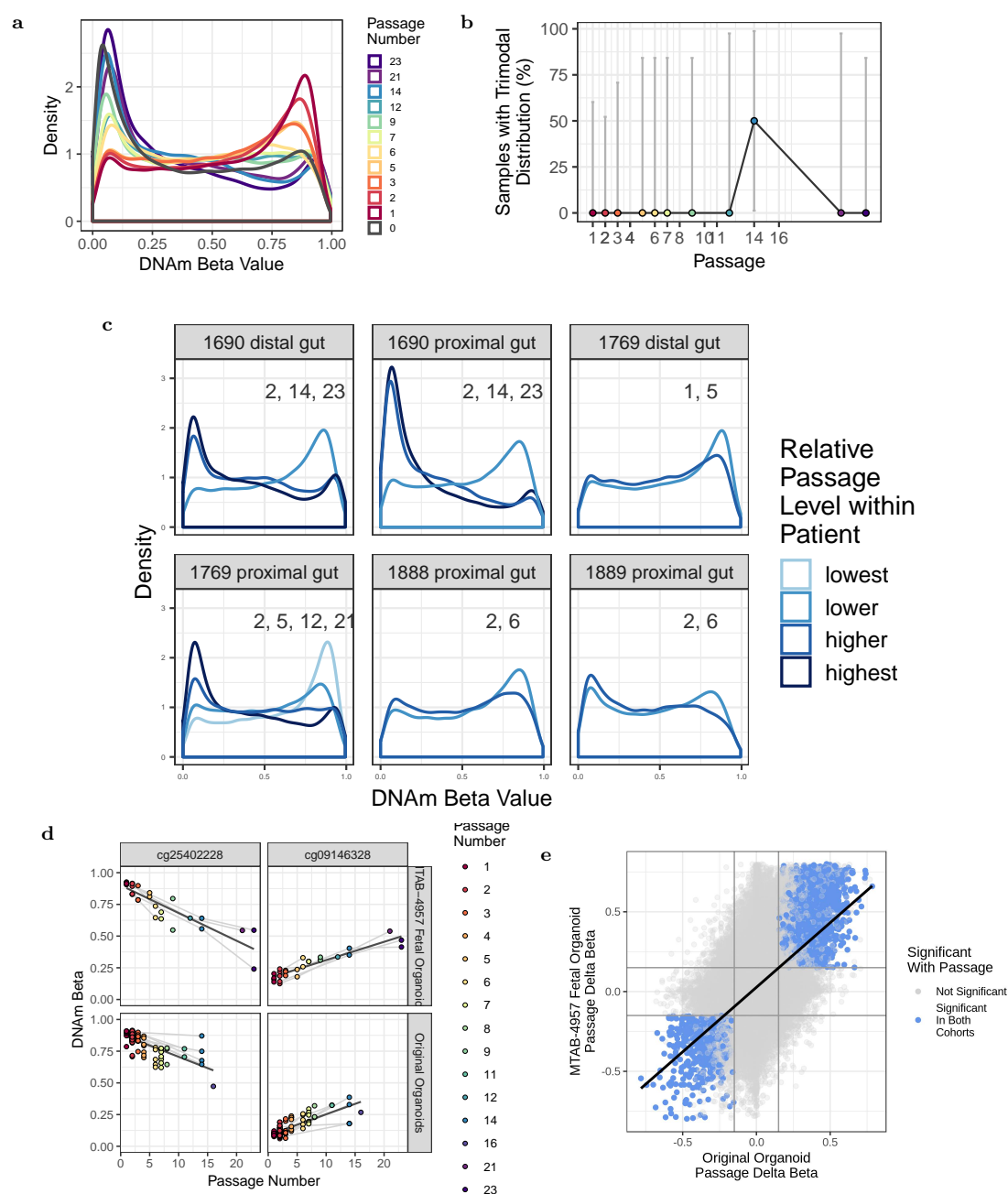


Figure S7: In E-MTAB-4957 fetal organoids the DNAm beta value distribution for high passage samples is more unmethylated than low passage samples. All beta distributions displayed are for the 28,418 most variable CpGs in fetal organoids. (a) Beta value distributions for all samples at a given passage number. Curves are coloured by the passage number. Included is the distribution for primary samples for comparison where passage is labelled as 0. (b) The percent of samples considered trimodal. Points are coloured by passage and the bars around each point represent the confidence interval for that percent. (c) Beta value distributions for samples derived from the same patient but cultured to a different number of passages. Plots are labelled with the patient ID number, sampling site of origin and the passage number of each sample derived from that patient and sampling site. Curves are coloured by high or low passage relative to the other sample(s).

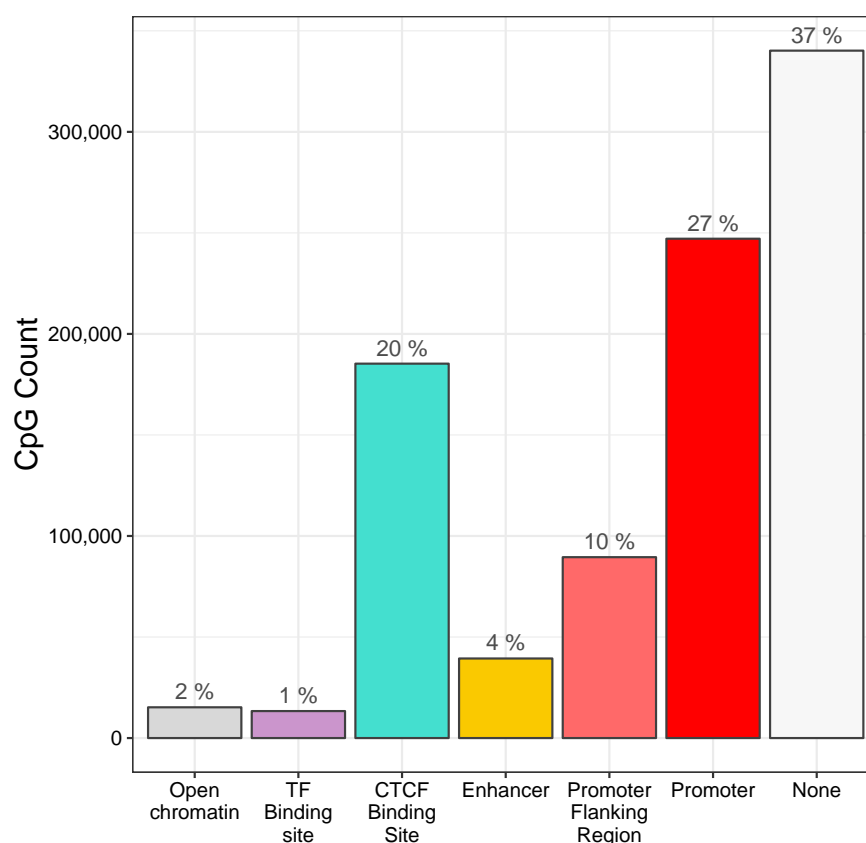


Figure S8: Background distribution of CpGs on the EPIC array. Number of 800,383 EPIC array CpGs (used in this analysis) in regulatory genomic features.

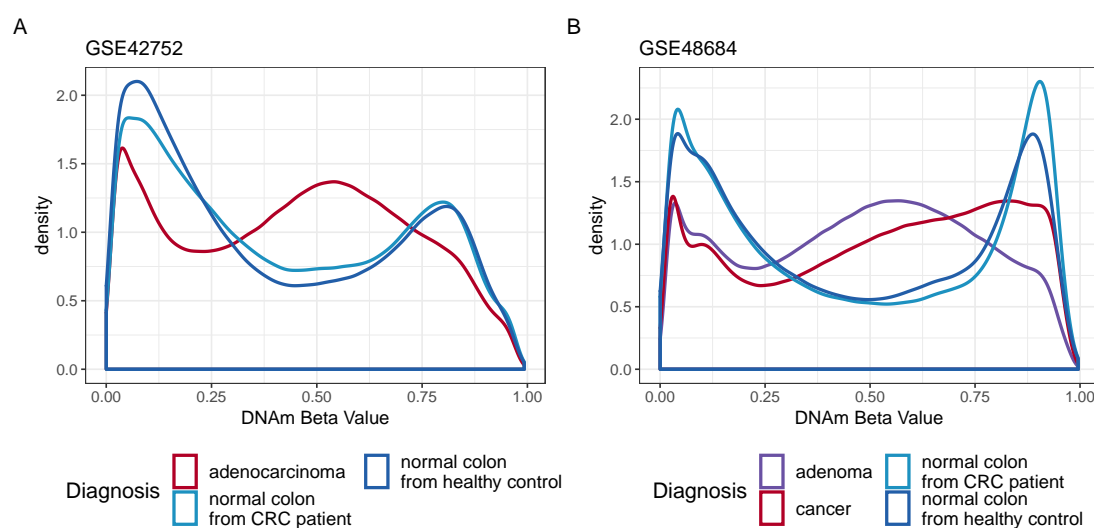


Figure S9: Appearance of the trimodal effect in three published cancer datasets. In all datasets samples grouped by cancer diagnosis, curves show the DNAm distribution the top most variable CpGs. (a) GSE42752; 63 samples; colon samples (b) GSE48684; 147 samples; colon samples (c) GSE53051; 220 samples; For clarity plots are also split into tissue of origin.

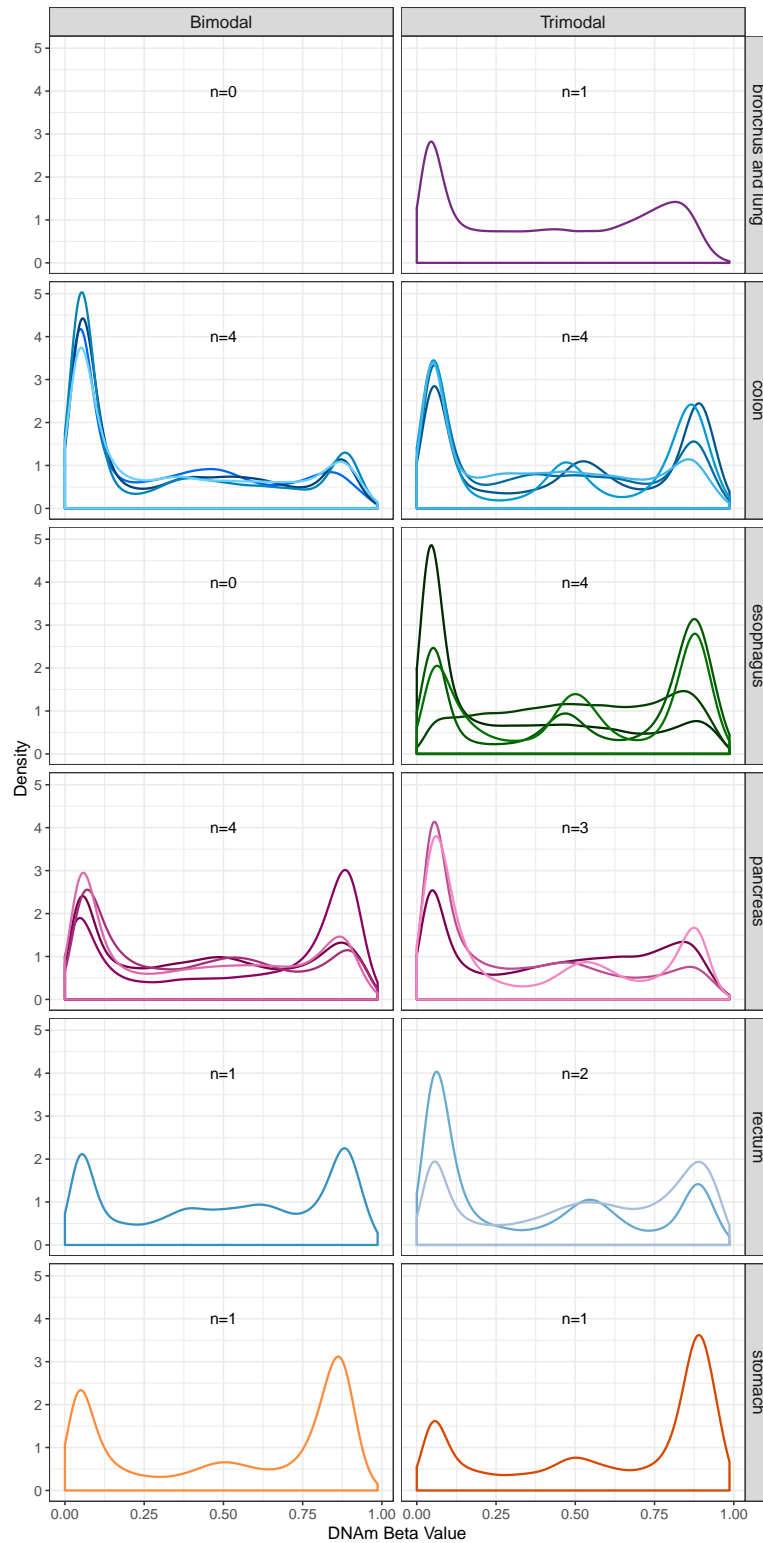


Figure S10: Appearance of the trimodal effect in publicly available cancer organoid data (GSE142213). Curves show the DNAm distribution the top 71,384 most variable CpGs. Samples are grouped by tissue of origin and separated by whether they are classified as tri or bimodal.