

#DNA-methylation analyze baby-sitting by Yifan

Data citation

Corley MJ, Vargas-Maya N, Pang APS, Lum-Jones A et al. Epigenetic Delay in the Neurodevelopmental Trajectory of DNA Methylation States in Autism Spectrum Disorders. Front Genet 2019;10:907. PMID: 31681403

Data

GSE131706

```
## obtain the necessary library
library(GEOquery)
```

```
## Loading required package: Biobase
## Loading required package: BiocGenerics
##
## Attaching package: 'BiocGenerics'
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
## The following objects are masked from 'package:base':
##
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, setdiff, table,
##     tapply, union, unique, unsplit, which.max, which.min
## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname)".
## Setting options('download.file.method.GEOquery'='auto')
## Setting options('GEOquery.inmemory.gpl'=FALSE)
library(minfi)

## Loading required package: GenomicRanges
## Loading required package: stats4
## Loading required package: S4Vectors
##
## Attaching package: 'S4Vectors'
## The following object is masked from 'package:utils':
##
##     findMatches
## The following objects are masked from 'package:base':
##
```

```

##      expand.grid, I, unname
## Loading required package: IRanges
## Loading required package: GenomeInfoDb
## Loading required package: SummarizedExperiment
## Loading required package: MatrixGenerics
## Loading required package: matrixStats
##
## Attaching package: 'matrixStats'
## The following objects are masked from 'package:Biobase':
##
##      anyMissing, rowMedians
##
## Attaching package: 'MatrixGenerics'
## The following objects are masked from 'package:matrixStats':
##
##      colAlls, colAnyNAs, colAnys, colAveragesPerRowSet, colCollapse,
##      colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
##      colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
##      colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
##      colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
##      colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
##      colWeightedMeans, colWeightedMedians, colWeightedSds,
##      colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAveragesPerColSet,
##      rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
##      rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
##      rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
##      rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
##      rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
##      rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
##      rowWeightedSds, rowWeightedVars
## The following object is masked from 'package:Biobase':
##
##      rowMedians
## Loading required package: Biostrings
## Loading required package: XVector
##
## Attaching package: 'Biostrings'
## The following object is masked from 'package:base':
##
##      strsplit
## Loading required package: bumpHunter
## Loading required package: foreach
## Loading required package: iterators
## Loading required package: parallel

```

```
## Loading required package: locfit
```

```
## locfit 1.5-9.9      2024-03-01
```

```
library(stringr)
```

```
library(ggplot2)
```

```
library(data.table)
```

```
##
```

```
## Attaching package: 'data.table'
```

```
## The following object is masked from 'package:SummarizedExperiment':
```

```
##
```

```
##      shift
```

```
## The following object is masked from 'package:GenomicRanges':
```

```
##
```

```
##      shift
```

```
## The following object is masked from 'package:IRanges':
```

```
##
```

```
##      shift
```

```
## The following objects are masked from 'package:S4Vectors':
```

```
##
```

```
##      first, second
```

```
getGEOSuppFiles("GSE131706")
```

```
##
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_Matrix_processed.csv.gz      8109247
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_RAW.tar                  48619520
```

```
##
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_Matrix_processed.csv.gz FALSE
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_RAW.tar                  FALSE
```

```
##
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_Matrix_processed.csv.gz 644
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_RAW.tar                  644
```

```
##
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_Matrix_processed.csv.gz 2024-08-5
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_RAW.tar                  2024-08-5
```

```
##
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_Matrix_processed.csv.gz 2024-08-5
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_RAW.tar                  2024-08-5
```

```
##
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_Matrix_processed.csv.gz 2024-08-0
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_RAW.tar                  2024-08-5
```

```
##
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_Matrix_processed.csv.gz 501
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_RAW.tar                  501
```

```
##
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_Matrix_processed.csv.gz 20
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_RAW.tar                  20
```

```
##
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_Matrix_processed.csv.gz yifan
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_RAW.tar                  yifan
```

```
##
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_Matrix_processed.csv.gz staff
```

```
## /Users/yifan/Desktop/git_hub_content/DNA-methly/GSE131706/GSE131706_RAW.tar
```

staff

```
untar("GSE131706/GSE131706_RAW.tar", exdir = "GSE131706/idad")  
length(list.files("GSE131706/idad", pattern = "idad"))
```

```
## [1] 136
```

```
idadFiles = list.files("GSE131706/idad", pattern = "idad.gz$", full = TRUE)  
sapply(idatFiles, gunzip, overwrite = TRUE)
```

```
## GSE131706/idad/GSM3814394_9296931030_R01C01_Grn.idat.gz  
## 8095228  
## GSE131706/idad/GSM3814394_9296931030_R01C01_Red.idat.gz  
## 8095228  
## GSE131706/idad/GSM3814395_9296931030_R02C01_Grn.idat.gz  
## 8095228  
## GSE131706/idad/GSM3814395_9296931030_R02C01_Red.idat.gz  
## 8095228  
## GSE131706/idad/GSM3814396_9296931030_R03C01_Grn.idat.gz  
## 8095228  
## GSE131706/idad/GSM3814396_9296931030_R03C01_Red.idat.gz  
## 8095228  
## GSE131706/idad/GSM3814397_9296931030_R04C01_Grn.idat.gz  
## 8095228  
## GSE131706/idad/GSM3814397_9296931030_R04C01_Red.idat.gz  
## 8095228  
## GSE131706/idad/GSM3814398_9296931030_R05C01_Grn.idat.gz  
## 8095223  
## GSE131706/idad/GSM3814398_9296931030_R05C01_Red.idat.gz  
## 8095223  
## GSE131706/idad/GSM3814399_9296931030_R06C01_Grn.idat.gz  
## 8095226  
## GSE131706/idad/GSM3814399_9296931030_R06C01_Red.idat.gz  
## 8095226  
## GSE131706/idad/GSM3814400_9296931030_R01C02_Grn.idat.gz  
## 8095227  
## GSE131706/idad/GSM3814400_9296931030_R01C02_Red.idat.gz  
## 8095227  
## GSE131706/idad/GSM3814401_9296931030_R02C02_Grn.idat.gz  
## 8095226  
## GSE131706/idad/GSM3814401_9296931030_R02C02_Red.idat.gz  
## 8095226  
## GSE131706/idad/GSM3814402_9296931030_R03C02_Grn.idat.gz  
## 8095220  
## GSE131706/idad/GSM3814402_9296931030_R03C02_Red.idat.gz  
## 8095220  
## GSE131706/idad/GSM3814403_9296931030_R04C02_Grn.idat.gz  
## 8095220  
## GSE131706/idad/GSM3814403_9296931030_R04C02_Red.idat.gz  
## 8095220  
## GSE131706/idad/GSM3814404_9305216040_R02C01_Grn.idat.gz  
## 8095238  
## GSE131706/idad/GSM3814404_9305216040_R02C01_Red.idat.gz  
## 8095238  
## GSE131706/idad/GSM3814405_9305216040_R03C01_Grn.idat.gz
```

```

## 8095239
## GSE131706/odat/GSM3814405_9305216040_R03C01_Red.idat.gz
## 8095239
## GSE131706/odat/GSM3814406_9305216040_R04C01_Grn.idat.gz
## 8095236
## GSE131706/odat/GSM3814406_9305216040_R04C01_Red.idat.gz
## 8095236
## GSE131706/odat/GSM3814407_9305216040_R05C01_Grn.idat.gz
## 8095238
## GSE131706/odat/GSM3814407_9305216040_R05C01_Red.idat.gz
## 8095238
## GSE131706/odat/GSM3814408_9305216040_R06C01_Grn.idat.gz
## 8095236
## GSE131706/odat/GSM3814408_9305216040_R06C01_Red.idat.gz
## 8095236
## GSE131706/odat/GSM3814409_9305216040_R02C02_Grn.idat.gz
## 8095246
## GSE131706/odat/GSM3814409_9305216040_R02C02_Red.idat.gz
## 8095246
## GSE131706/odat/GSM3814410_9305216040_R03C02_Grn.idat.gz
## 8095249
## GSE131706/odat/GSM3814410_9305216040_R03C02_Red.idat.gz
## 8095249
## GSE131706/odat/GSM3814411_9305216040_R04C02_Grn.idat.gz
## 8095239
## GSE131706/odat/GSM3814411_9305216040_R04C02_Red.idat.gz
## 8095239
## GSE131706/odat/GSM3814412_9305216040_R05C02_Grn.idat.gz
## 8095236
## GSE131706/odat/GSM3814412_9305216040_R05C02_Red.idat.gz
## 8095236
## GSE131706/odat/GSM3814413_9305216040_R06C02_Grn.idat.gz
## 8095238
## GSE131706/odat/GSM3814413_9305216040_R06C02_Red.idat.gz
## 8095238
## GSE131706/odat/GSM3814414_9373550079_R01C01_Grn.idat.gz
## 8095225
## GSE131706/odat/GSM3814414_9373550079_R01C01_Red.idat.gz
## 8095225
## GSE131706/odat/GSM3814415_9373550079_R02C01_Grn.idat.gz
## 8095220
## GSE131706/odat/GSM3814415_9373550079_R02C01_Red.idat.gz
## 8095220
## GSE131706/odat/GSM3814416_9373550079_R03C01_Grn.idat.gz
## 8095220
## GSE131706/odat/GSM3814416_9373550079_R03C01_Red.idat.gz
## 8095220
## GSE131706/odat/GSM3814417_9373550079_R04C01_Grn.idat.gz
## 8095224
## GSE131706/odat/GSM3814417_9373550079_R04C01_Red.idat.gz
## 8095224
## GSE131706/odat/GSM3814418_9373550079_R05C01_Grn.idat.gz
## 8095220
## GSE131706/odat/GSM3814418_9373550079_R05C01_Red.idat.gz

```

```

##                                     8095220
## GSE131706/idad/GSM3814419_9373550079_R06C01_Grn.idat.gz
##                                     8095225
## GSE131706/idad/GSM3814419_9373550079_R06C01_Red.idat.gz
##                                     8095225
## GSE131706/idad/GSM3814420_9373550079_R01C02_Grn.idat.gz
##                                     8095236
## GSE131706/idad/GSM3814420_9373550079_R01C02_Red.idat.gz
##                                     8095236
## GSE131706/idad/GSM3814421_9373550079_R02C02_Grn.idat.gz
##                                     8095228
## GSE131706/idad/GSM3814421_9373550079_R02C02_Red.idat.gz
##                                     8095228
## GSE131706/idad/GSM3814424_9373550079_R03C02_Grn.idat.gz
##                                     8095228
## GSE131706/idad/GSM3814424_9373550079_R03C02_Red.idat.gz
##                                     8095228
## GSE131706/idad/GSM3814427_9373550079_R04C02_Grn.idat.gz
##                                     8095228
## GSE131706/idad/GSM3814427_9373550079_R04C02_Red.idat.gz
##                                     8095228
## GSE131706/idad/GSM3814430_9373550079_R05C02_Grn.idat.gz
##                                     8095221
## GSE131706/idad/GSM3814430_9373550079_R05C02_Red.idat.gz
##                                     8095221
## GSE131706/idad/GSM3814433_9373550079_R06C02_Grn.idat.gz
##                                     8095226
## GSE131706/idad/GSM3814433_9373550079_R06C02_Red.idat.gz
##                                     8095226
## GSE131706/idad/GSM3814435_9426020146_R05C02_Grn.idat.gz
##                                     8095261
## GSE131706/idad/GSM3814435_9426020146_R05C02_Red.idat.gz
##                                     8095261
## GSE131706/idad/GSM3814438_9426020146_R06C02_Grn.idat.gz
##                                     8095260
## GSE131706/idad/GSM3814438_9426020146_R06C02_Red.idat.gz
##                                     8095260

```

```

rg_reads = read.metharray.exp("GSE131706/idad")
info = getGEO("GSE131706")

```

```
## Found 1 file(s)
```

```
## GSE131706_series_matrix.txt.gz
```

```
pdata = pData(info[[1]])
```

```
## assign group and sex
```

```
pdata$group = as.factor(str_remove(pdata$characteristics_ch1.3, "^disease state:"))
pdata$age <- as.numeric(str_extract(pdata$`age:ch1`, "^\\d+"))
```

```
## define the age group
```

```
pdata$age_group <- cut(pdata$age,
                      breaks = c(-Inf, 16, 33, Inf),
                      labels = c("young", "middle", "old"),
                      right = FALSE)
```

```

new_names = list()
for (i in seq_along(sampleNames(rg_reads))){
  new_names[[i]] =strsplit(sampleNames(rg_reads),"_")[[i]][[1]]
}

sampleNames(rg_reads) = new_names
rownames(pdata)

## [1] "GSM3814394" "GSM3814395" "GSM3814396" "GSM3814397" "GSM3814398"
## [6] "GSM3814399" "GSM3814400" "GSM3814401" "GSM3814402" "GSM3814403"
## [11] "GSM3814404" "GSM3814405" "GSM3814406" "GSM3814407" "GSM3814408"
## [16] "GSM3814409" "GSM3814410" "GSM3814411" "GSM3814412" "GSM3814413"
## [21] "GSM3814414" "GSM3814415" "GSM3814416" "GSM3814417" "GSM3814418"
## [26] "GSM3814419" "GSM3814420" "GSM3814421" "GSM3814424" "GSM3814427"
## [31] "GSM3814430" "GSM3814433" "GSM3814435" "GSM3814438"

## check name
all.equal(sampleNames(rg_reads), rownames(pdata))

## [1] TRUE

update_pdata = pdata[,c("group","age_group")]
pData(rg_reads) = as(update_pdata, "DataFrame")

## check phenotype data
pData(rg_reads)

## DataFrame with 34 rows and 2 columns
##           group age_group
##           <factor> <factor>
## GSM3814394 ASD/Autism   young
## GSM3814395 ASD/Autism   young
## GSM3814396 Control      young
## GSM3814397 ASD/Autism   young
## GSM3814398 Control      young
## ...          ...      ...
## GSM3814427 Control      middle
## GSM3814430 Control      middle
## GSM3814433 ASD/Autism   old
## GSM3814435 ASD/Autism   old
## GSM3814438 ASD/Autism   old

## check part of raw data
head(assay(rg_reads))

##           GSM3814394 GSM3814395 GSM3814396 GSM3814397 GSM3814398 GSM3814399
## 10600313          254          396          430          561          717          526
## 10600322         7302         12535         12287         11629         13054         16810
## 10600328         5922          6292          7333          8020          8597          4280
## 10600336         1735          1870          2673          4022          2817          4199
## 10600345          8101         10292         12110         12061         13050         11591
## 10600353          5748          6807          7377          8761          9468          7072
##           GSM3814400 GSM3814401 GSM3814402 GSM3814403 GSM3814404 GSM3814405
## 10600313          419          402          418          339          439          451
## 10600322          7231          8884         13020         12595          8022         11666
## 10600328          5664          5893          6711          7316          8082          8865

```

##	10600336	1632	2066	3842	2473	1990	4424
##	10600345	10027	9956	10327	12503	9582	12984
##	10600353	6281	7805	7390	9966	7468	7786
##		GSM3814406	GSM3814407	GSM3814408	GSM3814409	GSM3814410	GSM3814411
##	10600313	476	678	643	306	784	508
##	10600322	9927	8906	6672	10183	9119	9589
##	10600328	8624	8174	7845	7877	9956	8096
##	10600336	4343	2896	1877	2128	2860	2428
##	10600345	10532	10017	8685	10517	13007	7796
##	10600353	8423	7958	7290	7456	8749	8712
##		GSM3814412	GSM3814413	GSM3814414	GSM3814415	GSM3814416	GSM3814417
##	10600313	479	628	360	461	517	641
##	10600322	7476	7590	8042	7796	8054	8099
##	10600328	8840	8399	7053	6636	6284	6836
##	10600336	2722	2125	2796	3476	1587	1728
##	10600345	12598	11552	6666	9860	8012	7589
##	10600353	8200	7708	6533	6965	7495	6845
##		GSM3814418	GSM3814419	GSM3814420	GSM3814421	GSM3814424	GSM3814427
##	10600313	676	558	385	402	411	646
##	10600322	10095	8621	8422	5992	8802	7194
##	10600328	7826	6427	7139	6499	6914	7253
##	10600336	3795	2385	2285	1814	3515	2074
##	10600345	10591	8474	5437	8993	10757	9951
##	10600353	8333	6456	5467	6414	6883	7794
##		GSM3814430	GSM3814433	GSM3814435	GSM3814438		
##	10600313	430	367	744	660		
##	10600322	8708	8958	11471	12906		
##	10600328	7102	7150	9154	10173		
##	10600336	2246	2886	3495	3874		
##	10600345	10142	7183	10905	12474		
##	10600353	7704	6937	10189	9067		

```
## preprocess the data
```

```
mset = preprocessIllumina(rg_reads)
```

```
## Loading required package: IlluminaHumanMethylation450kmanifest
```

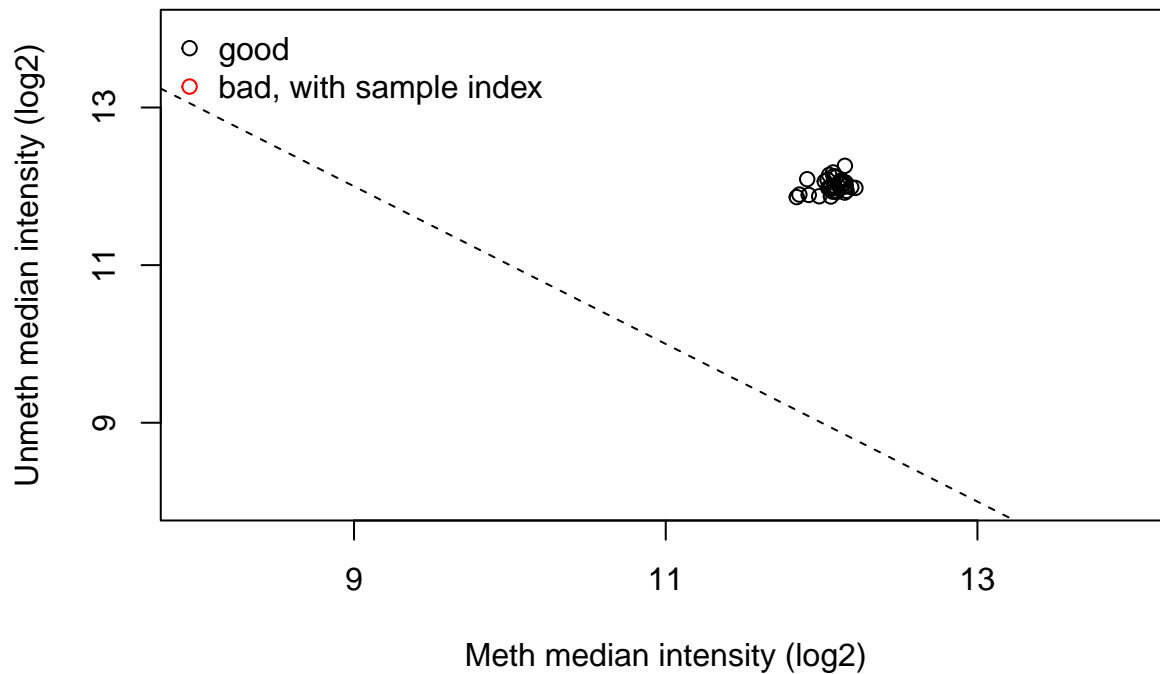
```
mset = mapToGenome(mset)
```

```
## Loading required package: IlluminaHumanMethylation450kanno.ilmn12.hg19
```

```
qc = getQC(mset)
```

```
## check QC
```

```
plotQC(qc)
```

```
## get normalized value
beta_values <- getBeta(mset)

## define groups
conditions = as.factor(pdata$group)
ages = as.factor(pdata$age_group)

## combine groups
groups = as.factor(paste(update_pdata$group, "_", update_pdata$age_group))

design = model.matrix(~ 0 + groups)
#colnames(design)

## redefine the colnames
colnames(design) = c("ASD_middle", "ASD_old", "ASD_young", "control_middle", "control_old", "control_young")

## check design
design
```

```
##   ASD_middle ASD_old ASD_young control_middle control_old control_young
## 1         0      0      1           0           0           0
## 2         0      0      1           0           0           0
## 3         0      0      0           0           0           1
## 4         0      0      1           0           0           0
## 5         0      0      0           0           0           1
## 6         0      0      1           0           0           0
## 7         0      0      0           0           0           1
## 8         0      0      0           0           0           1
## 9         0      0      0           0           0           1
## 10        0      0      1           0           0           0
## 11        0      0      0           1           0           0
## 12        0      1      0           0           0           0
## 13        0      0      0           0           1           0
```

```
## 14      0      0      0      0      1      0
## 15      0      0      0      0      1      0
## 16      0      1      0      0      0      0
## 17      0      0      0      0      1      0
## 18      0      1      0      0      0      0
## 19      0      0      0      0      1      0
## 20      1      0      0      0      0      0
## 21      0      0      0      1      0      0
## 22      1      0      0      0      0      0
## 23      0      1      0      0      0      0
## 24      0      0      0      1      0      0
## 25      1      0      0      0      0      0
## 26      0      0      0      1      0      0
## 27      1      0      0      0      0      0
## 28      0      0      0      1      0      0
## 29      1      0      0      0      0      0
## 30      0      0      0      1      0      0
## 31      0      0      0      1      0      0
## 32      0      1      0      0      0      0
## 33      0      1      0      0      0      0
## 34      0      1      0      0      0      0
## attr("assign")
## [1] 1 1 1 1 1 1
## attr("contrasts")
## attr("contrasts")$groups
## [1] "contr.treatment"
```

```
library(limma)
```

```
##
```

```
## Attaching package: 'limma'
```

```
## The following object is masked from 'package:BiocGenerics':
```

```
##
```

```
##      plotMA
```

```
## define contrast
```

```
contrasts <- makeContrasts(
```

```
  AvC_young= ASD_young - control_young,
  AvC_middle = ASD_middle - control_middle,
  AvC_old = ASD_old - control_old,
  A_old_young = ASD_old - ASD_young,
  A_middle_young = ASD_middle - ASD_young,
  ASD_control = (ASD_young + ASD_middle + ASD_old)/3 - (control_young + control_middle + control_old)/3,
  levels = colnames(design))
```

```
## filter NA data
```

```
beta_values = na.omit(beta_values)
```

```
## fit the model
```

```
fit = lmFit(beta_values, design)
fit_contrast = contrasts.fit(fit, contrasts)
fit2 = eBayes(fit_contrast)
```

```
update_pdata$group <- trimws(update_pdata$group)
```

```

update_pdata$age_group <- trimws(update_pdata$age_group)

dmp_1 <- topTable(fit2, coef = "AvC_young", number = Inf, adjust.method = "BH")
dmp_2 <- topTable(fit2, coef = "AvC_middle", number = Inf, adjust.method = "BH")
dmp_3 <- topTable(fit2, coef = "AvC_old", number = Inf, adjust.method = "BH")

dml_young = rownames(dmp_1[dmp_1$P.Value < 0.05 & abs(dmp_1$logFC) >= 0.1,])
dml_middle = rownames(dmp_2[dmp_2$P.Value < 0.05 & abs(dmp_2$logFC) >= 0.1,])
dml_old = rownames(dmp_3[dmp_3$P.Value < 0.05 & abs(dmp_3$logFC) >= 0.1,])

## seperate into groups
young_asd <- rownames(update_pdata[update_pdata$age_group == "young" & update_pdata$group == "ASD/Autism",])
young_control <- rownames(update_pdata[update_pdata$age_group == "young" & update_pdata$group == "Control",])

middle_asd <- rownames(update_pdata[update_pdata$age_group == "middle" & update_pdata$group == "ASD/Autism",])
middle_control <- rownames(update_pdata[update_pdata$age_group == "middle" & update_pdata$group == "Control",])

old_asd <- rownames(update_pdata[update_pdata$age_group == "old" & update_pdata$group == "ASD/Autism",])
old_control <- rownames(update_pdata[update_pdata$age_group == "old" & update_pdata$group == "Control",])

## assign data
young_asd_data = beta_values[dml_young, young_asd]
young_con_data = beta_values[dml_young, young_control]

middle_asd_data = beta_values[dml_middle, middle_asd]
middle_con_data = beta_values[dml_middle, middle_control]

old_asd_data = beta_values[dml_old, old_asd]
old_con_data = beta_values[dml_old, old_control]

# Convert all data to data.tables and add necessary columns

young_asd_dt <- data.table(Value = as.vector(young_asd_data), Age_Group = "Young", Condition = "ASD/Autism")
young_con_dt <- data.table(Value = as.vector(young_con_data), Age_Group = "Young", Condition = "Control")

middle_asd_dt <- data.table(Value = as.vector(middle_asd_data), Age_Group = "Middle", Condition = "ASD/Autism")
middle_con_dt <- data.table(Value = as.vector(middle_con_data), Age_Group = "Middle", Condition = "Control")

old_asd_dt <- data.table(Value = as.vector(old_asd_data), Age_Group = "Old", Condition = "ASD/Autism")
old_con_dt <- data.table(Value = as.vector(old_con_data), Age_Group = "Old", Condition = "Control")

# Combine all data tables into one

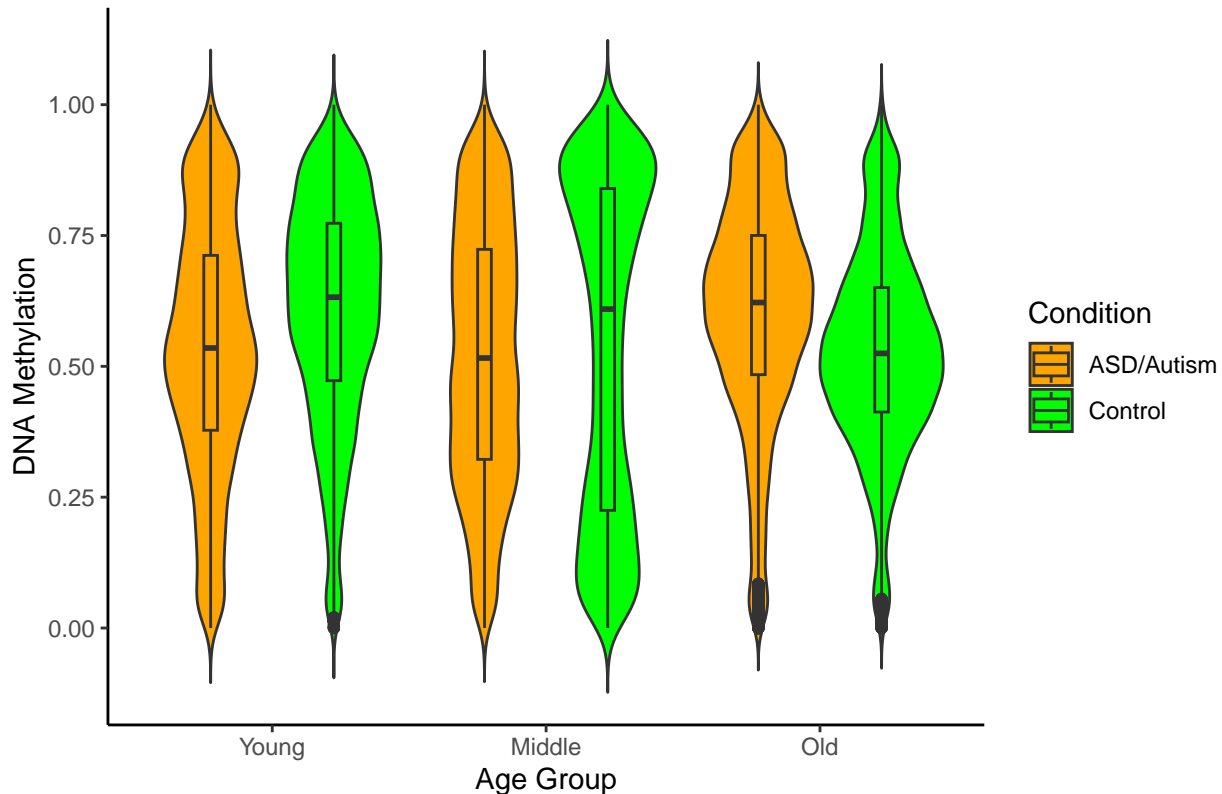
combined_data <- rbind(young_asd_dt, young_con_dt, middle_asd_dt, middle_con_dt, old_asd_dt, old_con_dt)
combined_data$Age_Group <- factor(combined_data$Age_Group, levels = c("Young", "Middle", "Old"))
combined_data$Condition <- factor(combined_data$Condition, levels = c("ASD/Autism", "Control"))

ggplot(combined_data, aes(x = Age_Group, y = Value, fill = Condition)) +
  geom_violin(trim = FALSE) +

```

```
geom_boxplot(width = 0.1, position = position_dodge(width = 0.9)) + # Optional: Add boxplot for median
labs(title = "Violin Plot of DNA Methylation by Age Group and Condition",
     x = "Age Group", y = "DNA Methylation") +
scale_fill_manual(values = c("ASD/Autism" = "orange", "Control" = "green")) +
theme_classic()
```

Violin Plot of DNA Methylation by Age Group and Condition



```
## example of ks_test
ks_test_result <- ks.test(combined_data$Value[combined_data$Age_Group == "Old"& combined_data$Condition == "Control"],
                          combined_data$Value[combined_data$Age_Group == "Old"& combined_data$Condition == "ASD/Autism"])

## Warning in ks.test.default(combined_data$Value[combined_data$Age_Group == "Old"& combined_data$Condition == "Control"],
## p-value will be approximate in the presence of ties
ks_test_result

##
## Asymptotic two-sample Kolmogorov-Smirnov test
##
## data: combined_data$Value[combined_data$Age_Group == "Old" & combined_data$Condition == "ASD/Autism"]
## D = 0.21559, p-value < 2.2e-16
## alternative hypothesis: two-sided

library(umap)
data_matrix <- t(beta_values)
# Set UMAP configuration parameters if desired
umap_config <- umap.defaults
umap_config$n_neighbors <- 4 # Number of neighbors (can be adjusted)
umap_config$min_dist <- 0.1 # Minimum distance between points (can be adjusted)
```

```

# Perform UMAP
umap_result <- umap(data_matrix, config = umap_config)

umap_data <- as.data.frame(umap_result$layout)

# Add sample information (condition, age_group) to the UMAP data
umap_data$Sample <- rownames(umap_data)
update_pdata$Sample = rownames(update_pdata)
umap_data

```

```

##           V1          V2      Sample
## GSM3814394 -0.28408767 -2.125571 GSM3814394
## GSM3814395  1.55450181 -1.750346 GSM3814395
## GSM3814396  1.24624355 -2.200475 GSM3814396
## GSM3814397  0.83786429 -2.481730 GSM3814397
## GSM3814398  1.42351974 -2.544649 GSM3814398
## GSM3814399  1.02513544 -1.711396 GSM3814399
## GSM3814400 -1.59765761  7.279467 GSM3814400
## GSM3814401  1.60126473 -3.009044 GSM3814401
## GSM3814402  0.41061154 -2.110708 GSM3814402
## GSM3814403  1.61230727 -1.973235 GSM3814403
## GSM3814404 -1.08905887  8.206042 GSM3814404
## GSM3814405 -1.28514499 -5.044772 GSM3814405
## GSM3814406 -1.09401921 -5.198183 GSM3814406
## GSM3814407 -0.41584966 -4.656123 GSM3814407
## GSM3814408 -0.92481157  7.939649 GSM3814408
## GSM3814409  1.14777942 -3.773730 GSM3814409
## GSM3814410 -1.16912346  7.814413 GSM3814410
## GSM3814411 -0.94560624  7.525007 GSM3814411
## GSM3814412 -0.66660096 -5.050113 GSM3814412
## GSM3814413 -1.49655222  7.888153 GSM3814413
## GSM3814414 -0.84643914 -2.600734 GSM3814414
## GSM3814415 -1.54027976 -4.558046 GSM3814415
## GSM3814416  0.19017672  8.458106 GSM3814416
## GSM3814417  1.45878192 -3.301777 GSM3814417
## GSM3814418 -1.02850395 -2.475137 GSM3814418
## GSM3814419  0.63660663 -4.077400 GSM3814419
## GSM3814420  0.46946534 -2.912478 GSM3814420
## GSM3814421  0.04453675  8.311839 GSM3814421
## GSM3814424 -1.50576839 -4.872187 GSM3814424
## GSM3814427 -0.55064542  8.389127 GSM3814427
## GSM3814430 -0.35900143  8.246426 GSM3814430
## GSM3814433  0.75710652 -4.283616 GSM3814433
## GSM3814435  1.16164452 -3.017976 GSM3814435
## GSM3814438  1.22160439 -4.328804 GSM3814438

```

```
update_pdata
```

```

##           group age_group      Sample
## GSM3814394 ASD/Autism      young GSM3814394
## GSM3814395 ASD/Autism      young GSM3814395
## GSM3814396   Control      young GSM3814396
## GSM3814397 ASD/Autism      young GSM3814397
## GSM3814398   Control      young GSM3814398

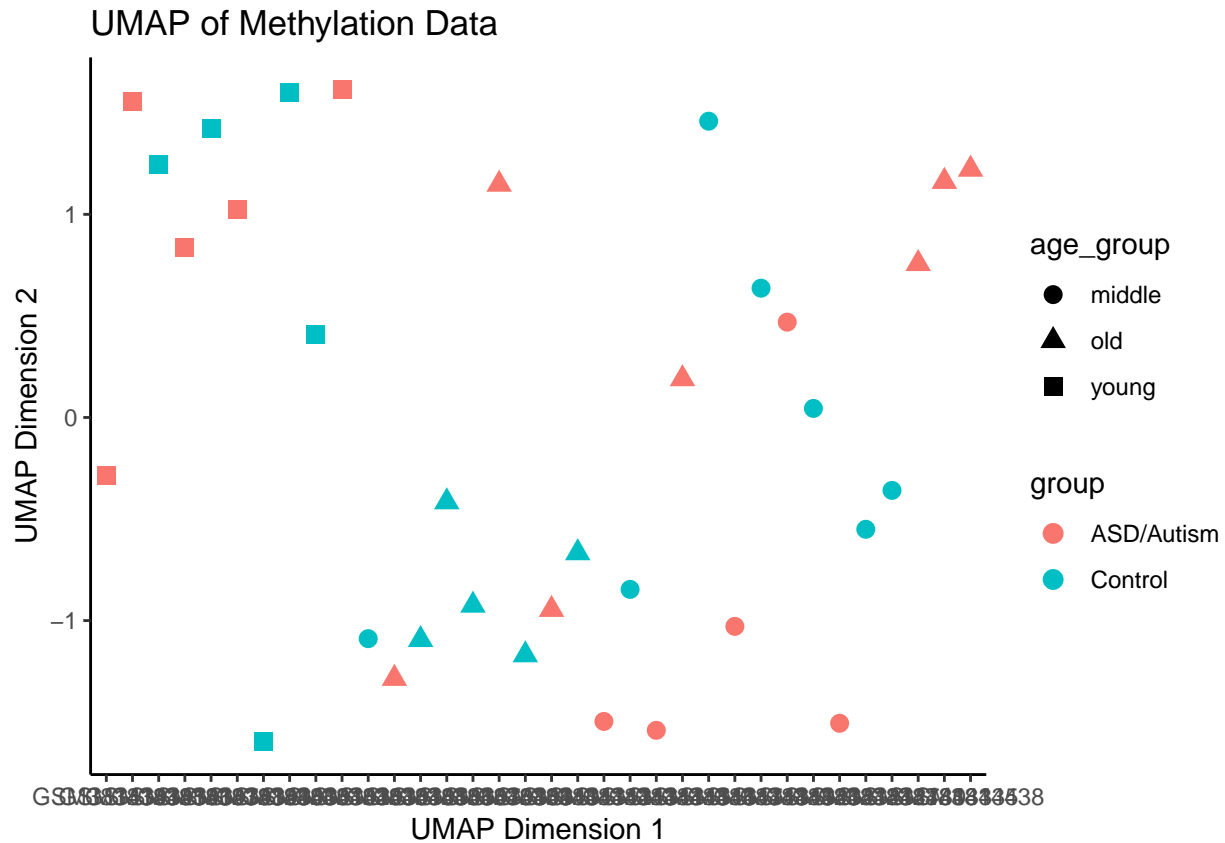
```

```
## GSM3814399 ASD/Autism      young GSM3814399
## GSM3814400      Control      young GSM3814400
## GSM3814401      Control      young GSM3814401
## GSM3814402      Control      young GSM3814402
## GSM3814403 ASD/Autism      young GSM3814403
## GSM3814404      Control      middle GSM3814404
## GSM3814405 ASD/Autism      old GSM3814405
## GSM3814406      Control      old GSM3814406
## GSM3814407      Control      old GSM3814407
## GSM3814408      Control      old GSM3814408
## GSM3814409 ASD/Autism      old GSM3814409
## GSM3814410      Control      old GSM3814410
## GSM3814411 ASD/Autism      old GSM3814411
## GSM3814412      Control      old GSM3814412
## GSM3814413 ASD/Autism      middle GSM3814413
## GSM3814414      Control      middle GSM3814414
## GSM3814415 ASD/Autism      middle GSM3814415
## GSM3814416 ASD/Autism      old GSM3814416
## GSM3814417      Control      middle GSM3814417
## GSM3814418 ASD/Autism      middle GSM3814418
## GSM3814419      Control      middle GSM3814419
## GSM3814420 ASD/Autism      middle GSM3814420
## GSM3814421      Control      middle GSM3814421
## GSM3814424 ASD/Autism      middle GSM3814424
## GSM3814427      Control      middle GSM3814427
## GSM3814430      Control      middle GSM3814430
## GSM3814433 ASD/Autism      old GSM3814433
## GSM3814435 ASD/Autism      old GSM3814435
## GSM3814438 ASD/Autism      old GSM3814438
```

```
umap_data <- merge(umap_data, update_pdata, by = "Sample")
```

```
# Rename the UMAP dimensions for easier plotting
colnames(umap_data)[1:2] <- c("UMAP1", "UMAP2")
```

```
ggplot(umap_data, aes(x = UMAP1, y = UMAP2, color = group, shape = age_group)) +
  geom_point(size = 3) +
  labs(title = "UMAP of Methylation Data",
       x = "UMAP Dimension 1",
       y = "UMAP Dimension 2") +
  theme_classic()
```



```
# Perform PCA
pca_result <- prcomp(data_matrix, center = TRUE, scale. = TRUE)
# Summary of PCA results
summary(pca_result)
```

```
## Importance of components:
##
##          PC1          PC2          PC3          PC4          PC5
## Standard deviation 307.8475 246.1230 200.75498 166.19981 156.49317
## Proportion of Variance 0.1953 0.1248 0.08306 0.05692 0.05047
## Cumulative Proportion 0.1953 0.3201 0.40320 0.46012 0.51059
##
##          PC6          PC7          PC8          PC9         PC10
## Standard deviation 133.18777 127.20445 119.85984 109.06776 106.29164
## Proportion of Variance 0.03656 0.03335 0.02961 0.02451 0.02328
## Cumulative Proportion 0.54715 0.58049 0.61010 0.63461 0.65790
##
##          PC11         PC12         PC13         PC14         PC15         PC16
## Standard deviation 104.33509 95.74883 92.62693 91.60622 91.39254 90.08848
## Proportion of Variance 0.02243 0.01889 0.01768 0.01729 0.01721 0.01673
## Cumulative Proportion 0.68033 0.69922 0.71690 0.73420 0.75141 0.76814
##
##          PC17         PC18         PC19         PC20         PC21         PC22
## Standard deviation 89.71587 88.44606 87.19260 86.37697 85.0264 84.7352
## Proportion of Variance 0.01659 0.01612 0.01567 0.01538 0.0149 0.0148
## Cumulative Proportion 0.78472 0.80084 0.81651 0.83189 0.8468 0.8616
##
##          PC23         PC24         PC25         PC26         PC27         PC28
## Standard deviation 83.40272 82.79781 81.64499 81.2468 80.65349 78.95402
## Proportion of Variance 0.01433 0.01413 0.01374 0.0136 0.01341 0.01285
## Cumulative Proportion 0.87592 0.89004 0.90378 0.9174 0.93079 0.94364
##
##          PC29         PC30         PC31         PC32         PC33         PC34
```

```
## Standard deviation      78.41052 77.35863 76.47685 71.31064 65.44704 4.226e-12
## Proportion of Variance 0.01267 0.01233 0.01205 0.01048 0.00883 0.000e+00
## Cumulative Proportion 0.95631 0.96864 0.98069 0.99117 1.00000 1.000e+00
```

```
pca_data <- as.data.frame(pca_result$x)
```

```
# Add sample information (condition, age_group) to the PCA data
```

```
pca_data$Sample <- rownames(pca_data)
```

```
pca_data <- merge(pca_data, update_pdata, by = "Sample")
```

```
# Plot the first two principal components with ggplot2
```

```
ggplot(pca_data, aes(x = PC1, y = PC2, color = group, shape = age_group)) +
```

```
  geom_point(size = 3) +
```

```
  labs(title = "PCA of Methylation Data",
```

```
        x = "Principal Component 1",
```

```
        y = "Principal Component 2") +
```

```
  theme_classic()
```

