Analysis of TCGA-CHOL methylation array data

Prakash Sah

2025-01-19

1. Introduction

In this analysis, DNA methylation profiling of cholangiocarcinoma (CCA) was performed using data from the TCGA-CHOL project. The methylation data were generated with the Illumina HumanMethylation450 BeadChip array, which interrogates over 450,000 CpG sites across the human genome. Preprocessed data were provided by TCGA through the Genomic Data Commons (GDC) as Masked Intensities and Beta Values, computed with the sesame pipeline. Signal masking and normalization were applied to improve reliability and reproducibility.

The data were downloaded and organized into a SummarizedExperiment object using the GDCprepare() function from the TCGAbiolinks package. This object contained sample-level metadata, probe annotations, and analysis-ready beta values derived via sesame. Alternatively, masked intensities can be downloaded and processed manually with sesame for more granular control over quality control and normalization, which mirrors the steps already applied to the GDC-provided beta values.

Differential methylation between normal and tumor tissues in TCGA-CHOL was investigated. For context and validation, two key resources were consulted:

1. "Identification of Prognostic Markers in Cholangiocarcinoma Using Altered DNA Methylation and Gene Expression Profiles" by Mishra et al. (2020), in which TCGA-CHOL methylation data were analyzed and the highest density of DMRs was reported on chromosome 1 and the lowest on chromosome 21 — findings that were corroborated in this analysis. DMRs on chromosome 19 were visualized using the Gviz package, producing a plot analogous to Figure 3 in Mishra et al. (2020).
2. The Bioconductor workflow "A cross-package Bioconductor workflow for analysing methylation array data" by Maksimovic, Phipson, and Oshlack, in which best practices for processing, quality control, differential methylation analysis, and gene ontology enrichment of 450k array data were described.

The workflow presented here applies Bioconductor tools to analyze TCGA-CHOL methylation array data and may serve as a template for analyzing other methylation datasets.

2. Methods Overview

After downloading the methylation data with TCGAbiolinks::GDCprepare(), the beta value matrix was extracted from the SummarizedExperiment using the assay() function. Problematic probes were filtered out following standard practice: probes with missing values (NA), probes mapping to sex chromosomes (chrX, chrY), probes overlapping common SNPs (minor allele frequency 0.05), and known cross-reactive probes (Chen et al., 2013) were removed. Quality control of the filtered beta values was assessed using density plots and multidimensional scaling (MDS) plots generated with functions from the minfi package.

To better approximate normality for statistical modeling, beta values were converted into M-values using the logit transformation. Differentially methylated probes (DMPs) were identified using the limma package, modeling tissue type (tumor vs. normal) as the primary covariate. Differentially methylated regions (DMRs)

were then identified using the DMRcate package, which aggregates nearby DMPs into contiguous regions of differential methylation.

For visualization of DMRs, the Gviz package was used to create custom genome tracks, including gene annotations and methylation signal profiles, with a focus on regions of interest such as chromosome 19. Finally, gene ontology enrichment analysis was performed on significant probes using the missMethyl package, which accounts for the varying number of CpG sites per gene when testing for pathway enrichment.

Libraries used: TCGAbiolinks
sesame
minfi
IlluminaHumanMethylation450kanno.ilmn12.hg19
limma
DMRcate
missMethyl
Gviz
GenomicRanges
RColorBrewer

3. Download and prepare data

```
# Download TCGA-CHOL dataset (methylation array dataset for cholangiocarcinoma samples).
TCGA_CHOL_query = GDCquery(project = "TCGA-CHOL", data.category = "DNA Methylation",
 ↪  data.type = "Masked Intensities", platform = "Illumina Human Methylation 450")
```

```
## ------------------------------------
```

```
## o GDCquery: Searching in GDC database
```

```
## ------------------------------------
```

```
## Genome of reference: hg38
```

```
## ------------------------------------------
```

```
## oo Accessing GDC. This might take a while...
```

```
## ------------------------------------------
```

```
## ooo Project: TCGA-CHOL
```

```
## --------------------
```

```
## oo Filtering results
```

```
## --------------------
```

```
## ooo By platform
```

```
## ooo By data.type
```

```
## ----------------

## oo Checking data

## ----------------

## ooo Checking if there are duplicated cases

## Warning: There are more than one file for the same case. Please verify query results. You can use the

## ooo Checking if there are results for the query

## -------------------

## o Preparing output

## -------------------
```

```
GDCdownload(TCGA_CHOL_query)
```

```
## Downloading data for project TCGA-CHOL

## Of the 90 files for download 90 already exist.

## All samples have been already downloaded
```

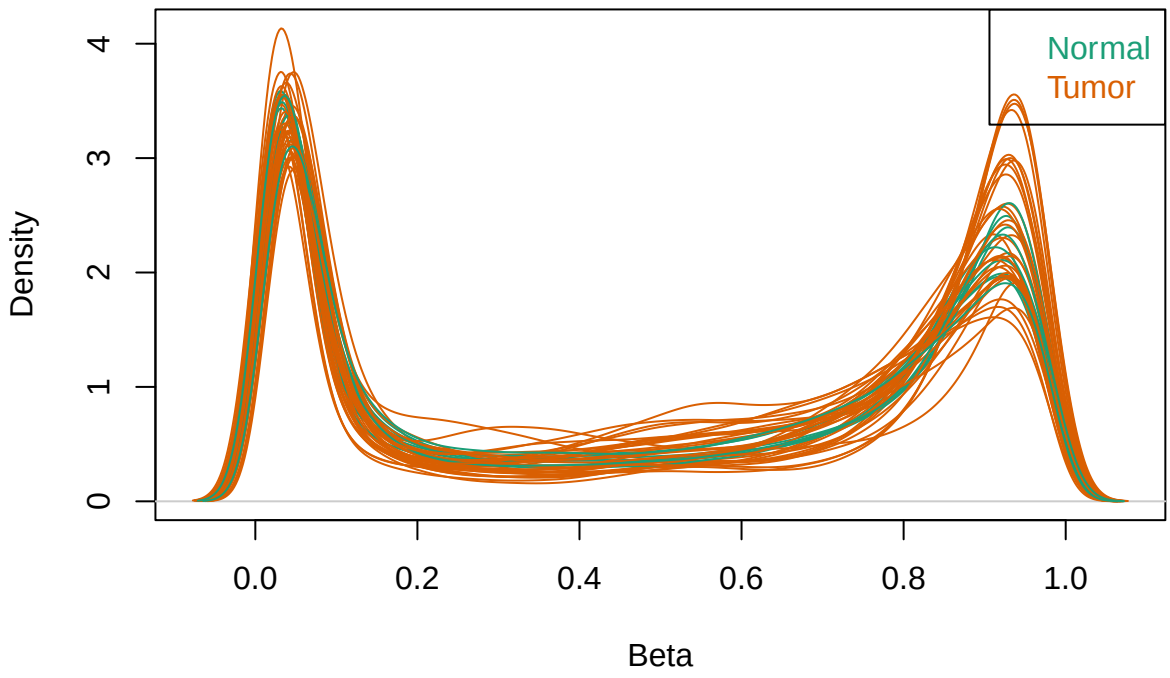
```
TCGA_CHOL_data = GDCprepare(TCGA_CHOL_query) # creates a summarized experiment object
 ↪  (requires semame package)
```

```
## Processing  IDATs with Sesame - http://bioconductor.org/packages/sesame/

## Running opensesame - applying quality masking and nondetection masking (threshold P-value 0.05)

## Please cite: doi: 10.1093/nar/gky691 and 10.1093/nar/gkt090

## This might take a while....

## =-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-

## Creating a SummarizedExperiment from DNA methylation input

## Accessing DNAm annotation from sesame package for: hg38 - HM450

## see ?sesameData and browseVignettes('sesameData') for documentation

## loading from cache
```

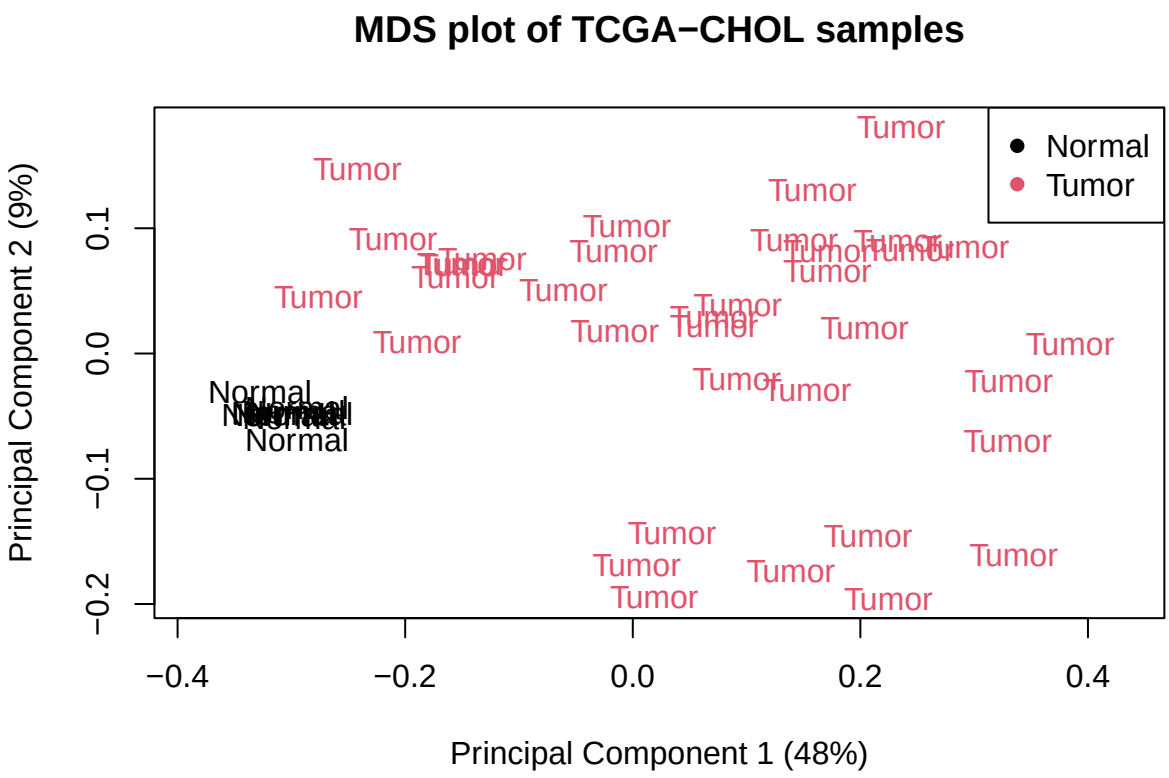
```
## Starting to add information to samples

##  => Add clinical information to samples

##  => Adding TCGA molecular information from marker papers

##  => Information will have prefix 'paper_'

## chol subtype information from:doi:10.1016/j.celrep.2017.02.033

## Starting to add information to samples

##  => Add clinical information to samples

##  => Adding TCGA molecular information from marker papers

##  => Information will have prefix 'paper_'

## chol subtype information from:doi:10.1016/j.celrep.2017.02.033
```

```r
saveRDS(TCGA_CHOL_data, file = "TCGA_CHOL_data.rds")

# Extract beta value matrix and sample information
beta_mat = assay(TCGA_CHOL_data) # access the matrix of beta values for each CpGs
coldata = colData(TCGA_CHOL_data) # sample information
```

```r
# QC plots
densityPlot(beta_mat, sampGroups = coldata$tissue_type) #density plot to examine beta
 ↪  values
```




```
plotMDS(beta_mat, top = 1000, labels = coldata$tissue_type, col =
 ↪  as.numeric(as.factor(coldata$tissue_type)), gene.selection = "common", main = "MDS
 ↪  plot of TCGA-CHOL samples")
legend("topright", legend = levels(as.factor(coldata$tissue_type)), col =
 ↪  1:length(unique(coldata$tissue_type)), pch = 16)
```

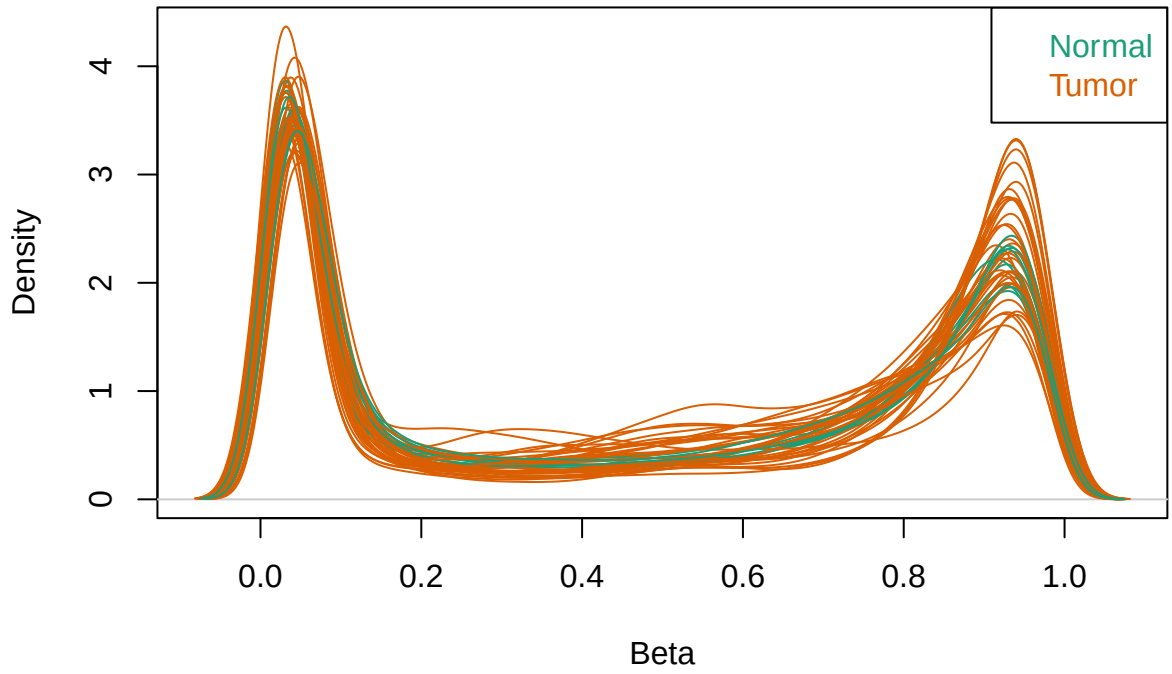



```
## Filter probes
# Load 450k annotation
ann450k = getAnnotation(IlluminaHumanMethylation450kanno.ilmn12.hg19)

# Remove probes with any NA
table(rowSums(is.na(beta_mat))==0) # examine number of probes with at one or more NA
```

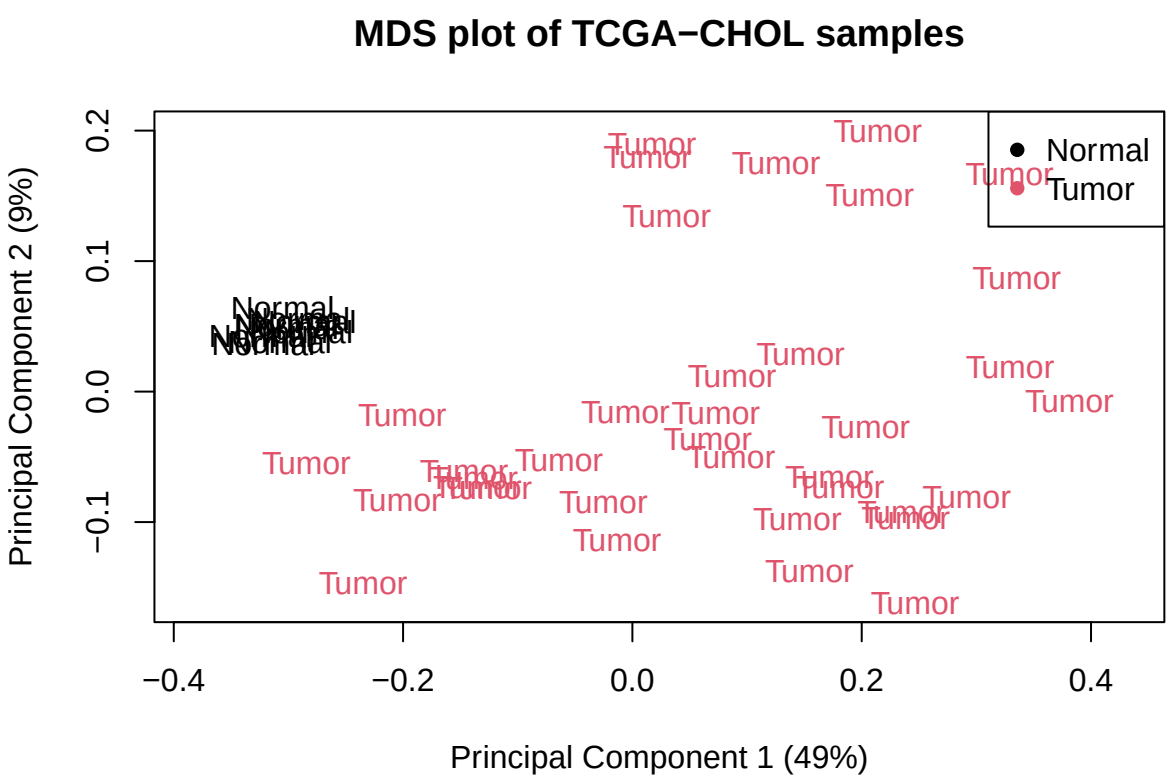
```
##
##  FALSE   TRUE
## 123697 361880
```

```
beta_mat = beta_mat[rowSums(is.na(beta_mat)) == 0, ]

# Remove probes on sex chromosomes
sex_probes = ann450k$Name[ann450k$chr %in% c("chrX", "chrY")]
beta_mat = beta_mat[!(rownames(beta_mat) %in% sex_probes), ]

# Remove probes overlapping known SNPs
no_snp_probes = ann450k$Name[is.na(ann450k$Probe_rs)]         # probes with no SNP
snp_probes = ann450k[!is.na(ann450k$Probe_rs), ]               # probes with SNPs
good_snp_probes = snp_probes$Name[snp_probes$Probe_maf <= 0.05] # SNPs with MAF  0.05

keep_probes = c(no_snp_probes, good_snp_probes)
beta_mat = beta_mat[rownames(beta_mat) %in% keep_probes, ]
```

```r
# Remove cross-reactive probes (Chen et al. 2013)
# Download and remove cross-reactive probes
url <-
 ↪  "https://raw.githubusercontent.com/hamidghaedi/Methylation_Analysis/master/cross_reactive_probe.chen
download.file(url, destfile = "cross_reactive_probe.chen2013.csv", mode = "wb")
cross_reactive = read.csv("cross_reactive_probe.chen2013.csv")
cross_reactive_ids = cross_reactive$TargetID[-1]  # remove header
beta_mat = beta_mat[!rownames(beta_mat) %in% cross_reactive_ids, ]

# QC plots with filtered beta values matrix
densityPlot(beta_mat, sampGroups = coldata$tissue_type) #density plot to examine beta
 ↪  values after filtering
```




```r
plotMDS(beta_mat, top = 1000, labels = coldata$tissue_type, col =
 ↪  as.numeric(as.factor(coldata$tissue_type)), gene.selection = "common", main = "MDS
 ↪  plot of TCGA-CHOL samples")
legend("topright", legend = levels(as.factor(coldata$tissue_type)), col =
 ↪  1:length(unique(coldata$tissue_type)), pch = 16)
```




```
# Transforms beta values to M values using the logit transformation: M = log2( / (1 - ))
 ↪  with an offset to prevent infinite values when  = 0 or 1.
beta2m = function(beta_mat, offset = 1e-6) {
  beta_mat = pmin(pmax(beta_mat, offset), 1 - offset)
  log2(beta_mat / (1 - beta_mat))
}

mval_mat = beta2m(beta_mat)

# Or use MValueToBetaValue() function from the sesame package
# mval_mat = BetaValueToMValue(beta_mat)

# Save Cleaned Matrices
saveRDS(beta_mat, "TCGA-CHOL_beta_matrix_filtered.rds")
saveRDS(mval_mat, "TCGA-CHOL_mvalue_matrix_filtered.rds")

# QC plots can now also be performed using the M value matrix
```

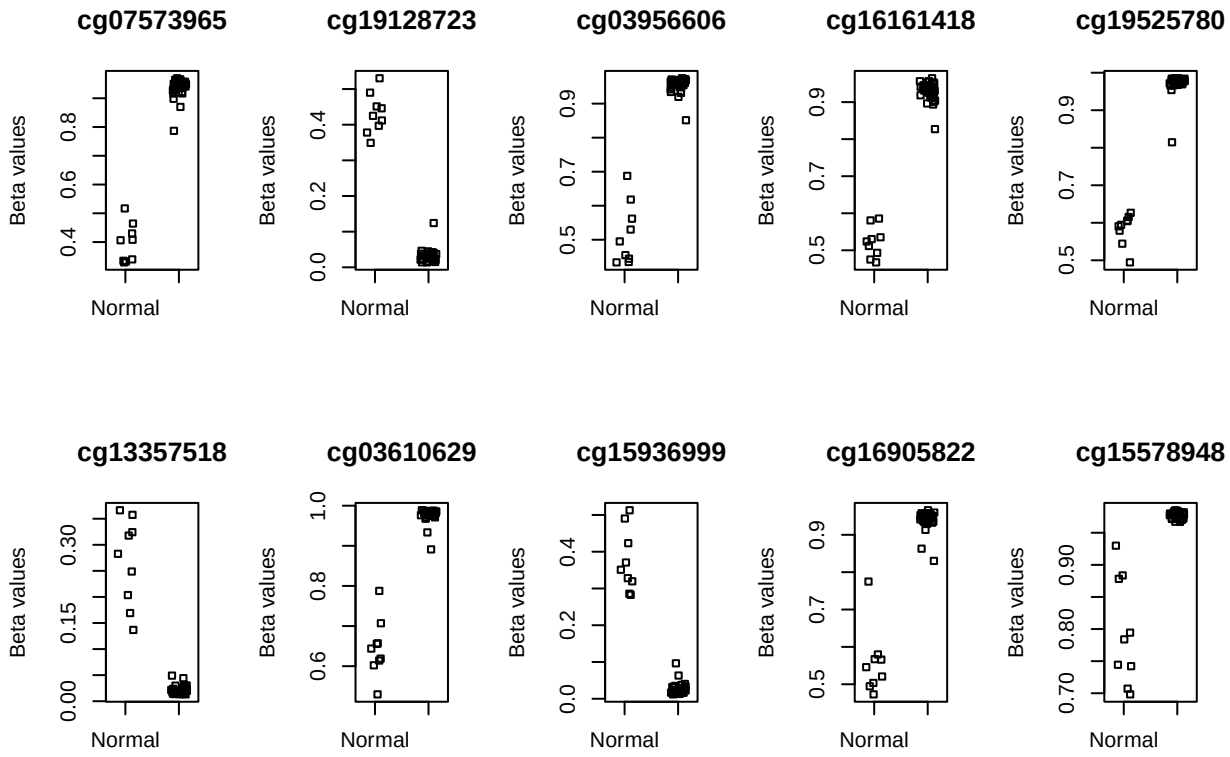
4. Differential methylation analysis using limma

```
# Differential methylation analysis
design = model.matrix(~ coldata$tissue_type)
fit = lmFit(mval_mat, design)
fit2 = eBayes(fit)
# extract results with annotation
```

```r
ann450k_sub <- ann450k[match(rownames(mval_mat), ann450k$Name),
                       c(1:4, 12:19, 24:ncol(ann450k))]
DMPs = topTable(fit2, coef=2, number=Inf, genelist = ann450k_sub)
head(DMPs)
```

```
##              chr       pos strand       Name Probe_rs Probe_maf CpG_rs CpG_maf
## cg07573965 chr22  35656596      - cg07573965     <NA>        NA   <NA>      NA
## cg19128723 chr10 105668237      - cg19128723     <NA>        NA   <NA>      NA
## cg03956606  chr1  11249057      - cg03956606     <NA>        NA   <NA>      NA
## cg16161418  chr6  30644798      + cg16161418     <NA>        NA   <NA>      NA
## cg19525780 chr14 105964243      - cg19525780     <NA>        NA   <NA>      NA
## cg13357518  chr1  22263347      + cg13357518     <NA>        NA   <NA>      NA
##            SBE_rs SBE_maf             Islands_Name Relation_to_Island
## cg07573965   <NA>      NA   chr22:35653361-35654028            S_Shelf
## cg19128723   <NA>      NA                                     OpenSea
## cg03956606   <NA>      NA                                     OpenSea
## cg16161418   <NA>      NA    chr6:30640431-30640853            S_Shelf
## cg19525780   <NA>      NA chr14:105964988-105965304            N_Shore
## cg13357518   <NA>      NA                                     OpenSea
##                      UCSC_RefGene_Name                 UCSC_RefGene_Accession
## cg07573965               HMGXB4;HMGXB4                 NM_001003681;NR_027780
## cg19128723                       OBFC1                              NM_024928
## cg03956606                MTOR;ANGPTL7                   NM_004958;NM_021146
## cg16161418           KIAA1949;KIAA1949                 NM_001134870;NM_133471
## cg19525780 C14orf80;C14orf80;C14orf80 NM_001134877;NM_001134876;NM_001134875
## cg13357518                       HSPG2                              NM_005529
##            UCSC_RefGene_Group Phantom DMR Enhancer             HMM_Island
## cg07573965         5'UTR;Body
## cg19128723               Body                 TRUE
## cg03956606       Body;TSS1500
## cg16161418        3'UTR;3'UTR
## cg19525780     Body;Body;Body                      14:105035254-105035778
## cg13357518               Body                 TRUE    1:22135873-22137111
##            Regulatory_Feature_Name           Regulatory_Feature_Group  DHS
## cg07573965
## cg19128723  10:105668034-105668371    Unclassified_Cell_type_specific TRUE
## cg03956606
## cg16161418     6:30644477-30645026 Gene_Associated_Cell_type_specific
## cg19525780
## cg13357518     1:22263094-22264418                       Unclassified TRUE
##                logFC   AveExpr         t      P.Value    adj.P.Val        B
## cg07573965  4.641138  3.093011  21.77702 1.019010e-25 1.588103e-20 46.96890
## cg19128723 -4.716821 -4.178821 -21.62076 1.375084e-25 1.588103e-20 46.70260
## cg03956606  4.430743  3.654519  21.58588 1.470581e-25 1.588103e-20 46.64288
## cg16161418  3.674036  3.070346  21.43886 1.953555e-25 1.588103e-20 46.39002
## cg19525780  4.847449  4.369264  21.13694 3.518105e-25 2.287978e-20 45.86501
## cg13357518 -4.067014 -4.771294 -20.87904 5.846827e-25 3.168707e-20 45.41033
```

```r
# plot top 10 differentially methylated probes
par(mfrow=c(2,5))
sapply(rownames(DMPs)[1:10], function(cpg){
plotCpg(beta_mat, cpg=cpg, pheno=coldata$tissue_type, ylab="Beta values")
})
```




```
## $cg07573965
## NULL
##
## $cg19128723
## NULL
##
## $cg03956606
## NULL
##
## $cg16161418
## NULL
##
## $cg19525780
## NULL
##
## $cg13357518
## NULL
##
## $cg03610629
## NULL
##
## $cg15936999
## NULL
##
## $cg16905822
## NULL
```

```
##
## $cg15578948
## NULL
```

5. Identify differential methylated regions using DMRcate

```
## Differentially methylated regions
my_annotation = cpg.annotate(object = mval_mat, datatype = "array", what = "M",
 ↪  analysis.type = "differential", design = design, contrasts = FALSE, coef = 2,
 ↪  arraytype = "450K")
```

```
## Your contrast returned 38960 individually significant probes. We recommend the default setting of pcu
```

```
my_annotation
```

```
## CpGannotated object describing 325172 CpG sites, with independent
## CpG threshold indexed at fdr=0.05 and 38960 significant CpG sites.
```

```
DMRs = dmrcate(my_annotation, lambda = 1000, C = 2)
```

```
## Fitting chr1...

## Fitting chr2...

## Fitting chr3...

## Fitting chr4...

## Fitting chr5...

## Fitting chr6...

## Fitting chr7...

## Fitting chr8...

## Fitting chr9...

## Fitting chr10...

## Fitting chr11...

## Fitting chr12...

## Fitting chr13...

## Fitting chr14...
```

```
## Fitting chr15...

## Fitting chr16...

## Fitting chr17...

## Fitting chr18...

## Fitting chr19...

## Fitting chr20...

## Fitting chr21...

## Fitting chr22...

## Demarcating regions...

## Done!
```

```
result.ranges = extractRanges(DMRs)
```

```
## see ?DMRcatedata and browseVignettes('DMRcatedata') for documentation

## loading from cache
```

```
result.ranges
```

```
## GRanges object with 5690 ranges and 8 metadata columns:
##          seqnames              ranges strand |   no.cpgs min_smoothed_fdr
##             <Rle>           <IRanges>  <Rle> | <integer>        <numeric>
##      [1]     chr6   33239694-33247509      * |        85     6.04721e-307
##      [2]    chr14 105963610-105965186      * |         9     1.33423e-276
##      [3]     chr6   31702632-31705409      * |        15     3.33837e-273
##      [4]    chr16   88717134-88717850      * |        13     1.82537e-234
##      [5]     chr2 216877750-216878510      * |         9     1.03637e-223
##      ...      ...                 ...    ... .       ...              ...
##   [5686]    chr20       866011-866087      * |         2      3.09964e-06
##   [5687]    chr19   21646447-21646470      * |         2      3.13164e-06
##   [5688]     chr6 164172504-164172519      * |         2      3.18382e-06
##   [5689]    chr10     7574780-7574782      * |         2      3.24630e-06
##   [5690]     chr8   25868045-25868076      * |         2      3.33432e-06
##              Stouffer       HMFDR       Fisher   maxdiff  meandiff
##             <numeric>   <numeric>    <numeric> <numeric> <numeric>
##      [1]  9.31418e-79 5.14242e-14 5.22409e-134 -0.567475 -0.125608
##      [2]  6.38828e-44 5.61953e-21  2.98986e-71  0.534812  0.214885
##      [3]  1.16310e-83 2.18579e-16 6.27948e-113 -0.536560 -0.223746
##      [4] 1.79074e-132 3.31872e-12 1.10097e-125 -0.604401 -0.498949
##      [5] 2.04129e-104 4.29463e-17 1.06349e-100 -0.382100 -0.217660
```

```
##      ...          ...         ...          ...       ...       ...
##   [5686]  1.36608e-04 0.000333464  1.14706e-04  0.279413  0.195730
##   [5687]  1.05875e-05 0.001244623  2.28309e-05  0.334659  0.294064
##   [5688]  2.80743e-06 0.000191624  4.81524e-06  0.172248  0.133636
##   [5689]  2.52287e-06 0.000488449  5.26762e-06  0.387678  0.347619
##   [5690]  1.67779e-05 0.001482962  3.55978e-05  0.270892  0.256812
##               overlapping.genes
##                     <character>
##      [1] RPS18, B3GALT4, VPS5..
##      [2]               C14orf80
##      [3]                  CLIC1
##      [4]                   CYBA
##      [5]             MREG, PECR
##      ...                    ...
##   [5686]                 ANGPT4
##   [5687] CTD-2561J22.3, CTD-2..
##   [5688]           RP1-230L10.1
##   [5689]                   <NA>
##   [5690]                   EBF2
##   -------
##   seqinfo: 22 sequences from an unspecified genome; no seqlengths
```

```
# DMRs by chromosome
table(seqnames(result.ranges))
```

```
##
##  chr1  chr2  chr3  chr4  chr5  chr6  chr7  chr8  chr9 chr10 chr11 chr12 chr13
##   565   390   249   244   314   418   343   243   104   283   376   303   121
## chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
##   178   160   271   387    78   348   159    48   108
```

```
# The highest density of DMRs was reported on chromosome 1 and the lowest on chromosome
 ↪  21 as also observed in analysis by Mishra et al.
```

6. Visualization of DMRs using Gviz

```
### custom plot of DMRs using Gviz
gen = "hg19" #genome version to be used
dmrIndex = 6 # DMR index. This DMR was also identified by Mishra et al in their TGCA-CHOL
 ↪  analysis (See reference).

# sample group colors
pal = brewer.pal(length(unique(coldata$tissue_type)), "Set1")
```

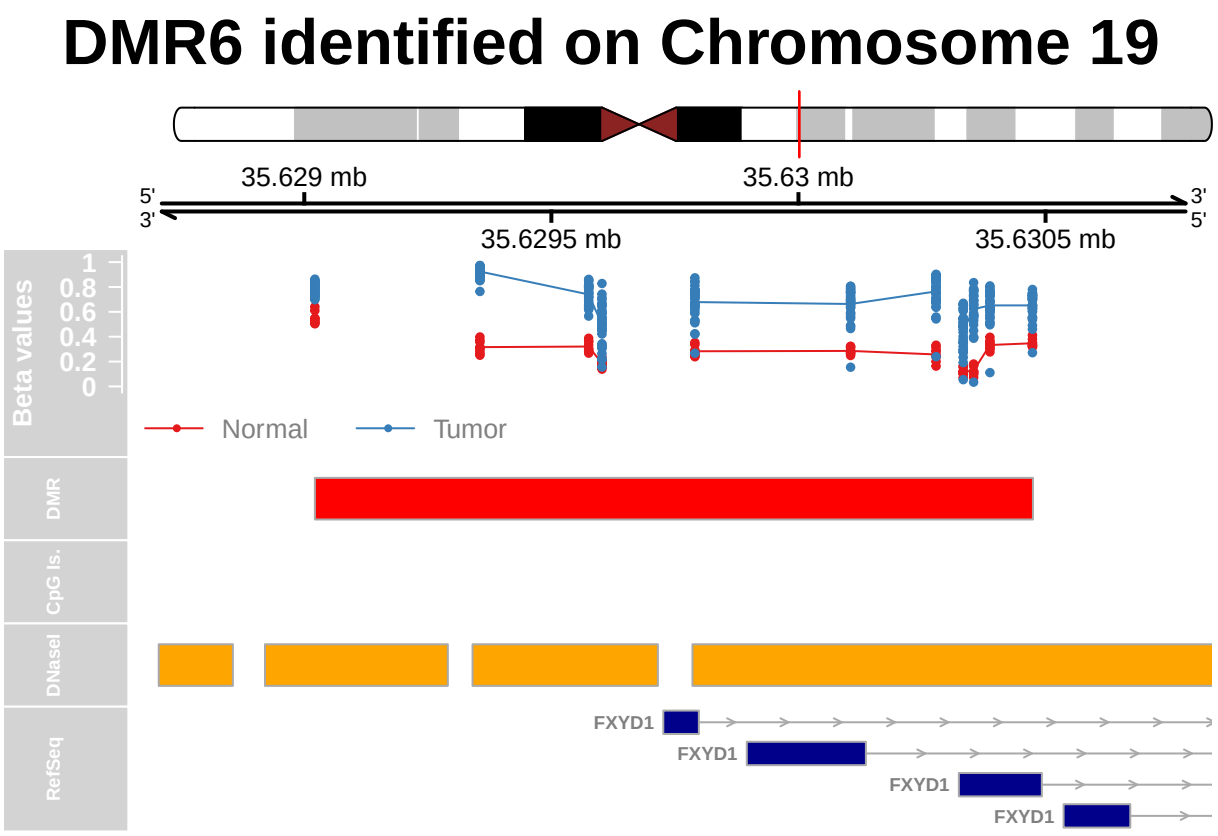
```
## Warning in brewer.pal(length(unique(coldata$tissue_type)), "Set1"): minimal value for n is 3, returni
```

```
names(pal) = unique(coldata$tissue_type)
cols = pal[coldata$tissue_type]

# Extract region of interest
chrom = as.character(seqnames(result.ranges[dmrIndex]))
```

```
start = as.numeric(start(result.ranges[dmrIndex]))
end   = as.numeric(end(result.ranges[dmrIndex]))
# Add some padding (25% extra space) to view context
minbase = start - 0.25 * (end - start)
maxbase = end   + 0.25 * (end - start)

## load annotation tracks
# CpG islands file (from UCSC)
islands = read.table("cpgIslandExt.txt", header=FALSE, stringsAsFactors=FALSE)
islandData = GRanges(seqnames = islands[,2],
              ranges = IRanges(start=islands[,3], end=islands[,4]),
              strand = "*")
# filter CpG islands to region of interest
roi <- GRanges(seqnames = chrom, ranges = IRanges(start = minbase, end = maxbase))
islandData_sub <- subsetByOverlaps(islandData, roi)

# DNase hypersensitive sites file (from UCSC)
dnase = read.table("wgEncodeRegDnaseClusteredV3.txt", header=FALSE,
 ↪  stringsAsFactors=FALSE)
dnaseData = GRanges(seqnames = dnase[,2],
                    ranges = IRanges(start=dnase[,3]+1, end=dnase[,4]),
                    strand = "*",
                    data = dnase[,5])
# filter DNase data to the region of interest
roi <- GRanges(seqnames = chrom, ranges = IRanges(start = minbase, end = maxbase))
dnaseData_sub <- subsetByOverlaps(dnaseData, roi)

## prepare methylation data
# make sure annotation and beta matrix are in same order
ann450kOrd = ann450k[order(ann450k$chr, ann450k$pos), ]
bValsOrd  = beta_mat[match(ann450kOrd$Name, rownames(beta_mat)), ]

# extract probes overlapping the DMR
cpgData = GRanges(seqnames = ann450kOrd$chr,ranges = IRanges(start=ann450kOrd$pos,
 ↪  end=ann450kOrd$pos), strand = "*", betas = bValsOrd)

cpgData = subsetByOverlaps(cpgData, result.ranges[dmrIndex])

## Create Gviz tracks
# ideogram and axis
iTrack = IdeogramTrack(genome=gen, chromosome=chrom, name="")
gTrack = GenomeAxisTrack(col="black", cex=1, name="", fontcolor="black")

# RefSeq track
rTrack = UcscTrack(genome=gen, chromosome=chrom, track ="NCBI RefSeq", table = "refGene",
 ↪  from=minbase, to=maxbase,  trackType="GeneRegionTrack", rstarts="exonStarts",
 ↪  rends="exonEnds", gene="name", symbol="name2", transcript="name", strand="strand",
 ↪  fill="darkblue", stacking="squish", name="RefSeq", showId=TRUE, geneSymbol=TRUE)
```

```
## Warning in .local(x, ...): 'track' parameter is deprecated now you go by the 'table' instead
##                 Use ucscTables(genome, track) to retrieve the list of tables for a track

## Warning in .local(x, ...): 'track' parameter is deprecated now you go by the 'table' instead
##                 Use ucscTables(genome, track) to retrieve the list of tables for a track
```

```
# CpG islands track
islandTrack = AnnotationTrack(range=islandData_sub, genome=gen, name="CpG Is.",
 ↪  chromosome=chrom, fill="darkgreen")

# DNase hypersensitive sites track
dnaseTrack = AnnotationTrack(range=dnaseData_sub, genome=gen, name="DNaseI",
 ↪  chromosome=chrom, fill="orange")

# DMR track
dmrTrack = AnnotationTrack(start=start, end=end, genome=gen, name="DMR",
 ↪  chromosome=chrom, fill="red")

# Methylation data track
methTrack = DataTrack(range=cpgData, genome=gen, chromosome=chrom,
                      groups=coldata$tissue_type, type=c("a","p"),
                      col=pal, name="Beta values", legend=TRUE,
                      background.panel="white", ylim=c(-0.05,1.05),
                      cex.title=0.8, cex.axis=0.8, cex.legend=0.8)

## combine all tracks and plot
tracks = list(iTrack, gTrack, methTrack, dmrTrack, islandTrack, dnaseTrack, rTrack)
sizes  = c(2, 2, 5, 2, 2, 2, 3)  # relative heights

plotTracks(tracks, from=minbase, to=maxbase, showTitle=TRUE, add53=TRUE,
 ↪  add35=TRUE,lty.grid=3, sizes=sizes, main="DMR6 identified on Chromosome 19")
```



7. Gene ontology analysis using missMethyl

```
## GO analysis
# get significant probes with adjusted p value <0.05
sigCpGs <- DMPs$Name[DMPs$adj.P.Val<0.05]
sigCpGs[1:10]
```

```
##  [1] "cg07573965" "cg19128723" "cg03956606" "cg16161418" "cg19525780"
##  [6] "cg13357518" "cg03610629" "cg15936999" "cg16905822" "cg15578948"
```

```
# get all probes
all = DMPs$Name

# run enrichment
gst <- gometh(sig.cpg=sigCpGs, all.cpg=all)
```

```
## All input CpGs are used for testing.
```

```
topGSA(gst)
```

```
##            ONTOLOGY                                TERM    N   DE         P.DE
## GO:0007275       BP multicellular organism development 4472 2707 2.356661e-36
## GO:0048731       BP                 system development 3821 2356 1.766131e-35
## GO:0007399       BP         nervous system development 2389 1569 5.321759e-34
## GO:0048856       BP   anatomical structure development 5615 3248 8.523211e-31
## GO:0009653       BP anatomical structure morphogenesis 2603 1649 4.268031e-28
## GO:0032502       BP              developmental process 6106 3471 4.623807e-28
## GO:0071944       CC                     cell periphery 5458 3046 7.825641e-27
## GO:0048513       BP           animal organ development 2869 1755 7.432889e-25
## GO:0022008       BP                       neurogenesis 1652 1101 1.105564e-23
## GO:0030182       BP             neuron differentiation 1346  916 8.649591e-23
## GO:0048699       BP              generation of neurons 1423  962 1.127848e-22
## GO:0032501       BP   multicellular organismal process 7173 3885 7.953098e-22
## GO:0009887       BP         animal organ morphogenesis  989  673 6.590940e-19
## GO:0005886       CC                     plasma membrane 5009 2752 7.037980e-19
## GO:0030154       BP               cell differentiation 4141 2387 1.342643e-18
## GO:0048869       BP     cellular developmental process 4143 2387 1.767170e-18
## GO:0048666       BP                 neuron development 1088  735 1.577114e-17
## GO:0007267       BP                cell-cell signaling 1614 1020 1.754172e-17
## GO:0030054       CC                      cell junction 2092 1308 2.068569e-17
## GO:0007155       BP                      cell adhesion 1442  897 2.917640e-17
##                     FDR
## GO:0007275 5.266431e-32
## GO:0048731 1.973387e-31
## GO:0007399 3.964178e-30
## GO:0048856 4.761705e-27
## GO:0009653 1.722137e-24
## GO:0032502 1.722137e-24
## GO:0071944 2.498280e-23
## GO:0048513 2.076285e-21
## GO:0022008 2.745116e-20
```

```
## GO:0030182 1.932924e-19
## GO:0048699 2.291275e-19
## GO:0032501 1.481066e-18
## GO:0009887 1.123412e-15
## GO:0005886 1.123412e-15
## GO:0030154 2.000270e-15
## GO:0048869 2.468184e-15
## GO:0048666 2.073163e-14
## GO:0007267 2.177804e-14
## GO:0030054 2.432963e-14
## GO:0007155 3.260025e-14
```

References

1. Mishra, N. K., Niu, M., Southekal, S., Bajpai, P., Elkholy, A., Manne, U., & Guda, C. (2020). Identification of Prognostic Markers in Cholangiocarcinoma Using Altered DNA Methylation and Gene Expression Profiles. Frontiers in Genetics, 11, 522125. https://doi.org/10.3389/fgene.2020.522125

2. Maksimovic, J., Phipson, B., & Oshlack, A. (2016). A cross-package Bioconductor workflow for analysing methylation array data. F1000Research, 5, 1281. https://doi.org/10.12688/f1000research.8839.2