COMMSBIO pesticide mixture toxicity

Load the packages used to make this statistical analysis

```
Libraryloadpacks()
Libraryloadpacks()
```

```
# Checking package: dplyr
Libraryloadpacks()

## Checking required packages: dplyr
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

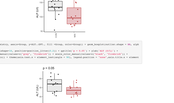
```
Libraryloadpacks()
Libraryloadpacks()
```

```
## Checking required packages: Hmisc
##
## Checking required packages: Formula
##
## Checking required packages: ggplot2
##
## Checking required packages: survival
##
## Checking required packages: lattice
##
## The following objects are masked from 'package:base':
##
##     format.pvalue, units
##
## Attaching package: 'Hmisc'
##
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
##
## The following objects are masked from 'package:base':
##
##     format.pvalue, units
##
## Attaching package: 'survival'
##
## The following object is masked from 'package:Hmisc':
##
##     cluster
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
##
## Loaded ggsurvplot functions
##
##  Attaching package: 'survminer'
##
##  The following object is masked from 'package:survival':
##
##     ggsurvplot
```

```
Libraryloadpacks()
Libraryloadpacks()
```

```
## Checking required packages: ggsurvplot
```

```
# Checking package: ggsurvplot
Libraryloadpacks()
```



```
wilcoxtest(Genus[Ketogenic, Gastrectomy], y=GLP1_1M[C.jejuni/coli=0, Gastrectomy=1], alternative="greater")$p.value
wilcoxtest(Genus[Ketogenic, Gastrectomy], y=GLP1_1M[C.jejuni/coli=1, Gastrectomy=1], alternative="greater")$p.value
Genus[,"Ketogenic"] ~ Gastrectomy*C.jejuni/coli + Age + Sex + Diabetes + Metformin + Statins + Proton.pump.inhibitor
Genus[,"Ketogenic"] ~ Gastrectomy*C.jejuni/coli + Age + Sex + Diabetes + Metformin + Statins + Proton.pump.inhibitor + Genus[,"Lactobacillus"]
```

```
>>> ls -lrt /usr/local/bin/*
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldsc
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/munge_sumstats.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/make_annot.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/make_annot_mafbins.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/make_annot_maftrees.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/make_annot_mafbins2.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/make_annot_mafbins_ldscores.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/make_annot_mafbins_ldscores_chr.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore_fromlist.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore_sldsc.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore_sldsc_fromlist.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore_sldsc_fromlist_mafbins.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore_sldsc_fromlist_mafbins2.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore_sldsc_fromlist_mafbins_ldscores.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore_sldsc_fromlist_mafbins_ldscores_chr.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore_sldsc_fromlist_mafbins_ldscores_chr_sldsc.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore_sldsc_fromlist_mafbins_ldscores_sldsc.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore_sldsc_fromlist_mafbins_sldsc.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore_sldsc_fromlist_sldsc.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldscore_sldsc.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/sldsc.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/sldsc_fromlist.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/sldsc_fromlist_mafbins.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/sldsc_fromlist_mafbins2.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/sldsc_fromlist_mafbins_ldscores.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/sldsc_fromlist_mafbins_ldscores_chr.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/sldsc_fromlist_mafbins_ldscores_chr_sldsc.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/sldsc_fromlist_mafbins_ldscores_sldsc.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/sldsc_fromlist_mafbins_sldsc.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/sldsc_fromlist_sldsc.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/sldsc_sldsc.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/munge_sumstats.py
-rwxr-xr-x  1 root  wheel  1.5K 12 Jan 21:11 /usr/local/bin/ldsc
>>> ldsc -h
usage: ldsc [-h] [--bfile BFILE] [--l2] [--ld-wind-snps LD_WIND_SNPS] [--ld-wind-kb LD_WIND_KB] [--ld-wind-cm LD_WIND_CM] [--print-snps PRINT_SNPS] [--annot ANNOT] [--thin-annot] [--cts-bin CTS_BIN] [--cts-bin-add CTS_BIN_ADD] [--cts-breaks CTS_BREAKS] [--cts-names CTS_NAMES] [--per-allele] [--pq-exp PQ_EXP] [--maf MAF] [--chunk-size CHUNK_SIZE] [--pickle] [--yes-really] [--no-print-annot] [--maf-wind MAF_WIND] [--h2 H2] [--rg RG] [--ref-ld REF_LD] [--ref-ld-chr REF_LD_CHR] [--w-ld W_LD] [--w-ld-chr W_LD_CHR] [--overlap-annot] [--hm3-snps HM3_SNPS] [--hm3-quick HM3_QUICK] [--print-cov] [--print-delete] [--sum-weights SUM_WEIGHTS] [--two-sample] [--no-intercept] [--intercept-h2 INTERCEPT_H2] [--intercept-gcov INTERCEPT_GCOV] [--no-scale-suffix] [--frqfile FRQFILE] [--frqfile-full FRQFILE_FULL] [--frqfile-chr FRQFILE_CHR] [--l1] [--l2 L2] [--l4] [--dominance DOMINANCE] [--maf-beta MAF_BETA] [--nsnp NSNP] [--nindel NINDEL] [--print-chr PRINT_CHR] [--maf-exp MAF_EXP] [--no-intercept-se] [--no-hm3-snps] [--h2-cov H2_COV] [--rg-cov RG_COV] [--print-coefficients] [--print-trait-sum] [--print-total-cov] [--print-ldscores] [--print-gencov] [--print-cov-cov] [--print-delete-vals] [--chisq-max CHISQ_MAX] [--samp-prev SAMP_PREV] [--pop-prev POP_PREV] [--invalid-cov-thresh INVALID_COV_THRESH] [--invalid-h2-thresh INVALID_H2_THRESH] [--invalid-genetic-cov-thresh INVALID_GENETIC_COV_THRESH] [--invalid-gcov-thresh INVALID_GCOV_THRESH] [--invalid-ldscore-thresh INVALID_LDSCORE_THRESH] [--no-check] [--no-check-alleles] [--one-tail-gc] [--n-blocks N_BLOCKS] [--chr CHR] [--no-ldscores] [--no-intercepts] [--no-hess] [--no-output] [--no-obs-h2] [--no-obs-gc] [--no-obs-cov] [--no-obs-int] [--no-obs-ldscores] [--no-obs-hess] [--no-exp-h2] [--no-exp-gc] [--no-exp-cov] [--no-exp-int] [--no-exp-ldscores] [--no-exp-hess] [--no-gc] [--no-cov] [--no-int] [--no-ldscores] [--no-hess] [--no-h2] [--no-gcov] [--no-h2-cov] [--no-gcov-cov] [--no-h2-int] [--no-gcov-int] [--no-h2-ldscores] [--no-gcov-ldscores] [--no-h2-hess] [--no-gcov-hess] [--no-cov-ldscores] [--no-cov-hess] [--no-int-ldscores] [--no-int-hess] [--no-ldscores-hess] [--no-obs-h2-cov] [--no-obs-gc-cov] [--no-obs-cov-cov] [--no-obs-int-cov] [--no-obs-ldscores-cov] [--no-obs-hess-cov] [--no-exp-h2-cov] [--no-exp-gc-cov] [--no-exp-cov-cov] [--no-exp-int-cov] [--no-exp-ldscores-cov] [--no-exp-hess-cov] [--no-gc-cov] [--no-cov-cov] [--no-int-cov] [--no-ldscores-cov] [--no-hess-cov] [--no-h2-cov-cov] [--no-gcov-cov-cov] [--no-h2-int-cov] [--no-gcov-int-cov] [--no-h2-ldscores-cov] [--no-gcov-ldscores-cov] [--no-h2-hess-cov] [--no-gcov-hess-cov] [--no-cov-ldscores-cov] [--no-cov-hess-cov] [--no-int-ldscores-cov] [--no-int-hess-cov] [--no-ldscores-hess-cov] [--no-obs-h2-gc] [--no-obs-gc-gc] [--no-obs-cov-gc] [--no-obs-int-gc] [--no-obs-ldscores-gc] [--no-obs-hess-gc] [--no-exp-h2-gc] [--no-exp-gc-gc] [--no-exp-cov-gc] [--no-exp-int-gc] [--no-exp-ldscores-gc] [--no-exp-hess-gc] [--no-gc-gc] [--no-cov-gc] [--no-int-gc] [--no-ldscores-gc] [--no-hess-gc] [--no-h2-cov-gc] [--no-gcov-cov-gc] [--no-h2-int-gc] [--no-gcov-int-gc] [--no-h2-ldscores-gc] [--no-gcov-ldscores-gc] [--no-h2-hess-gc] [--no-gcov-hess-gc] [--no-cov-ldscores-gc] [--no-cov-hess-gc] [--no-int-ldscores-gc] [--no-int-hess-gc] [--no-ldscores-hess-gc] [--no-obs-h2-int] [--no-obs-gc-int] [--no-obs-cov-int] [--no-obs-int-int] [--no-obs-ldscores-int] [--no-obs-hess-int] [--no-exp-h2-int] [--no-exp-gc-int] [--no-exp-cov-int] [--no-exp-int-int] [--no-exp-ldscores-int] [--no-exp-hess-int] [--no-gc-int] [--no-cov-int] [--no-int-int] [--no-ldscores-int] [--no-hess-int] [--no-h2-cov-int] [--no-gcov-cov-int] [--no-h2-int-int] [--no-gcov-int-int] [--no-h2-ldscores-int] [--no-gcov-ldscores-int] [--no-h2-hess-int] [--no-gcov-hess-int] [--no-cov-ldscores-int] [--no-cov-hess-int] [--no-int-ldscores-int] [--no-int-hess-int] [--no-ldscores-hess-int] [--no-obs-h2-ldscores] [--no-obs-gc-ldscores] [--no-obs-cov-ldscores] [--no-obs-int-ldscores] [--no-obs-ldscores-ldscores] [--no-obs-hess-ldscores] [--no-exp-h2-ldscores] [--no-exp-gc-ldscores] [--no-exp-cov-ldscores] [--no-exp-int-ldscores] [--no-exp-ldscores-ldscores] [--no-exp-hess-ldscores] [--no-gc-ldscores] [--no-cov-ldscores] [--no-int-ldscores] [--no-ldscores-ldscores] [--no-hess-ldscores] [--no-h2-cov-ldscores] [--no-gcov-cov-ldscores] [--no-h2-int-ldscores] [--no-gcov-int-ldscores] [--no-h2-ldscores-ldscores] [--no-gcov-ldscores-ldscores] [--no-h2-hess-ldscores] [--no-gcov-hess-ldscores] [--no-cov-ldscores-ldscores] [--no-cov-hess-ldscores] [--no-int-ldscores-ldscores] [--no-int-hess-ldscores] [--no-ldscores-hess-ldscores] [--no-obs-h2-hess] [--no-obs-gc-hess] [--no-obs-cov-hess] [--no-obs-int-hess] [--no-obs-ldscores-hess] [--no-obs-hess-hess] [--no-exp-h2-hess] [--no-exp-gc-hess] [--no-exp-cov-hess] [--no-exp-int-hess] [--no-exp-ldscores-hess] [--no-exp-hess-hess] [--no-gc-hess] [--no-cov-hess] [--no-int-hess] [--no-ldscores-hess] [--no-hess-hess] [--no-h2-cov-hess] [--no-gcov-cov-hess] [--no-h2-int-hess] [--no-gcov-int-hess] [--no-h2-ldscores-hess] [--no-gcov-ldscores-hess] [--no-h2-hess-hess] [--no-gcov-hess-hess] [--no-cov-ldscores-hess] [--no-cov-hess-hess] [--no-int-ldscores-hess] [--no-int-hess-hess] [--no-ldscores-hess-hess] [--no-h2-cov-ldscores] [--no-gcov-cov-ldscores] [--no-h2-int-ldscores] [--no-gcov-int-ldscores] [--no-h2-ldscores-ldscores] [--no-gcov-ldscores-ldscores] [--no-h2-hess-ldscores] [--no-gcov-hess-ldscores] [--no-cov-ldscores-ldscores] [--no-cov-hess-ldscores] [--no-int-ldscores-ldscores] [--no-int-hess-ldscores] [--no-ldscores-hess-ldscores
```

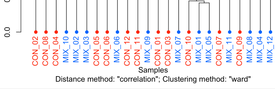
```
> dmr_gene_annotation[1:5,]
```

```
GRCh38.p13.genome.fa.gz.bismark.cov.gz.CpG_report.txt.gz.CX_report.txt.gz.bismark.bam.deduplicated.bam.sorted.bam.splitting_report.txt.M-bias.txt.png.png.png.png.png.png.png.png.png.png.png.png.png.png.
```

```
> dmr_gene_annotation[1:5,]
```

```
> dmr_gene_annotation[1:5,]
```

```
> dmr_gene_annotation[1:5,]
```




```
##
## Call:
## hclust(d = d, method = HCLUST.METHODS[hclust.method])
##
## Cluster method   : ward.D
## Distance         : pearson
## Number of objects: 24
```

```
# Volcano plots of differentially methylated CpG sites
annotateWithGeneParts(as(myDiff,"GRanges"),gene.obj)
```

```
## Summary of target set annotation with genic parts
```

```
## Rows in target set: 77008
```

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

```
## percentage of target features overlapping with annotation:
```

```
##   promoter       exon     intron intergenic
##      12.65       9.02      40.05      49.30
```

```
##
## percentage of target features overlapping with annotation:
```

```
## (with promoter > exon > intron precedence):
```

```
##   promoter       exon     intron intergenic
##      12.65       5.08      32.96      49.30
```

```
##
```

```
## percentage of annotation boundaries with feature overlap:
```

```
## promoter     exon   intron
##     9.03     0.83     3.92
```

```
##
```

```
## summary of distances to the nearest TSS:
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0    5112   21372   63780   64529 1665135
```

```
##
```

```
diffAnn=annotateWithGeneParts(as(myDiff,"GRanges"),gene.obj)
diffAnn_table <- getAssociationWithTSS(diffAnn)

# Match annotations
result_methylation_table <- data.frame(myDiff, diffAnn_table)
result_methylation_table$feature.name<- sub("\\..*","\\",result_methylation_table$feature.name)
result_methylation_table$gene.name <- mapIds(org.Rn.eg.db, result_methylation_table$feature.name, keytype="REFSEQ
", column="GENENAME")
```

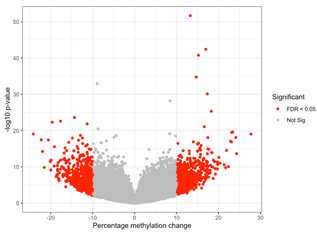
```
## 'select()' returned 1:1 mapping between keys and columns
```

```
result_methylation_table$go.name <- mapIds(org.Rn.eg.db, result_methylation_table$feature.name, keytype="REFSEQ",
column="GO")
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
result_methylation_table$Significant <- ifelse(abs(result_methylation_table$meth.diff) > 10 & result_methylation_
table$qvalue < 0.05, "FDR < 0.05", "Not Sig")

ggplot(result_methylation_table, aes(x = meth.diff, y = -log10(pvalue))) +
  geom_point(aes(color = Significant)) +
  xlab("Percentage methylation change") + ylab("-log10 p-value") +
  scale_color_manual(values = c("red", "grey")) +
  theme_bw(base_size = 12) + theme(legend.position = "bottom") + theme_bw()
```




```
# Promoter specific analysis
promoters=regionCounts(myobj,gene.obj$promoters)
meth_promoters=unite(promoters, destrand=FALSE)
```

```
## uniting...
```

```
# Finding differentially methylated bases or regions
myDiff_meth_promoters=calculateDiffMeth(meth_promoters)
```

```
## two groups detected:
##  will calculate methylation difference as the difference of
## treatment (group: 1) - control (group: 0)
```

```
# Summary of target set annotation with genic parts
annotateWithGeneParts(as(myDiff_meth_promoters,"GRanges"),gene.obj)
```

```
## Summary of target set annotation with genic parts
```

```
## Rows in target set: 11228
```

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

```
## percentage of target features overlapping with annotation:
```

```
##   promoter       exon     intron intergenic
##     100.00     100.00      95.54       0.00
```

```
##
## percentage of target features overlapping with annotation:
```

```
## (with promoter > exon > intron precedence):
```

```
##   promoter       exon     intron intergenic
##        100          0          0          0
```

```
##
```

```
## percentage of annotation boundaries with feature overlap:
```

```
## promoter     exon   intron
##    22.89     3.18     3.36
```

```
##
```

```
## summary of distances to the nearest TSS:
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0       0       0       0       0       0
```

```
##
```

```
diffAnn_promoters=annotateWithGeneParts(as(myDiff_meth_promoters,"GRanges"),gene.obj)
diffAnn_table_promoters <- getAssociationWithTSS(diffAnn_promoters)

# Match annotations
result_methylation_table_promoters <- data.frame(myDiff_meth_promoters, diffAnn_table_promoters)
result_methylation_table_promoters$feature.name<- sub("\\..*","\\",result_methylation_table_promoters$feature.nam
e)
result_methylation_table_promoters$gene.name <- mapIds(org.Rn.eg.db, result_methylation_table_promoters$feature.n
ame, keytype="REFSEQ", column="GENENAME")
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
result_methylation_table_promoters$go.name <- mapIds(org.Rn.eg.db, result_methylation_table_promoters$feature.nam
e, keytype="REFSEQ", column="GO")
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
result_methylation_table_promoters <- result_methylation_table_promoters %>% distinct(feature.name, meth.diff, .k
eep_all = TRUE)

i <- abs(result_methylation_table_promoters$meth.diff) > 10
result_methylation_table_promoters <- result_methylation_table_promoters[i,]
i <- result_methylation_table_promoters$qvalue < 0.05
result_methylation_table_promoters <- result_methylation_table_promoters[i,]
```

```
# Match annotations
result_methylation_table_promoters <- data.frame(myDiff_meth_promoters, diffAnn_table_promoters)
result_methylation_table_promoters$feature.name<- sub("\\..*","\\",result_methylation_table_promoters$feature.nam
e)
result_methylation_table_promoters$gene.name <- mapIds(org.Rn.eg.db, result_methylation_table_promoters$feature.n
ame, keytype="REFSEQ", column="GENENAME")
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
result_methylation_table_promoters$go.name <- mapIds(org.Rn.eg.db, result_methylation_table_promoters$feature.nam
e, keytype="REFSEQ", column="GO")
```

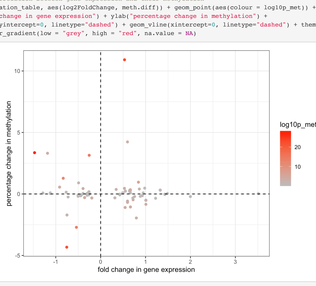
```
## 'select()' returned 1:many mapping between keys and columns
```

```
# Fetch transcriptome results
rna_table <-data.frame(transcriptome$res_mixtureOrdered)
rna_table$feature.name <- mapIds(org.Rn.eg.db, rna_table$gene_id, keytype="ENSEMBL", column="REFSEQ")
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
# Make correlation tables with gene methylation and gene expression
correlation_table <- left_join(result_methylation_table_promoters, rna_table, by = 'feature.name')
correlation_table <- correlation_table[!is.na(correlation_table$gene_id),]
correlation_table <- correlation_table[correlation_table$padj < 0.05,]
correlation_table$log10p_met <- -log10(correlation_table$pvalue.x)

# Plot the correlation between gene expression and their methylation
ggplot(correlation_table, aes(log2FoldChange, meth.diff)) + geom_point(aes(colour = log10p_met)) +
  xlab("fold change in gene expression") + ylab("percentage change in methylation") +
  geom_hline(yintercept=0, linetype="dashed") + geom_vline(xintercept=0, linetype="dashed") + theme_bw() +
  scale_colour_gradient(low = "grey", high = "red", na.value = NA)
```




```
# Calculate and test the correlation
cor.test(correlation_table$meth.diff, correlation_table$FC)
```

```
##
##  Pearson's product-moment correlation
##
## data:  correlation_table$meth.diff and correlation_table$FC
## t = -0.12895, df = 101, p-value = 0.8977
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence intervals:
##  -0.2058439  0.1811441
## sample estimates:
##        cor
## -0.01283034
```