

# The coMET User Guide

Tiphaine Martin <sup>\*</sup>, Idil Yet <sup>†</sup>, Pei-Chien Tsai <sup>‡</sup>, Jordana T. Bell <sup>§</sup>

Edited: September 2014; Compiled: October 22, 2014

## 1 Citation

---

```
citation(package='coMET')

##
## To cite package 'coMET' in publications use:
##
##   Tiphaine C. Martin, Idil Yet, Pei-Chien Tsai and Jordana T. Bell (2014).
##   coMET: Visualization regional plots of (epigenome/transcriptome)genome-wide
##   association scan results. R package version 0.99.3.
##   http://comet.epigen.kcl.ac.uk:3838/coMET/ or http://epigen.kcl.ac.uk/comet
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {coMET: Visualization regional plots of (epigenome/transcriptome)genome-wide assoc
##     author = {Tiphaine C. Martin and Idil Yet and Pei-Chien Tsai and Jordana T. Bell},
##     year = {2014},
##     note = {R package version 0.99.3},
##     url = {http://comet.epigen.kcl.ac.uk:3838/coMET/ or http://epigen.kcl.ac.uk/comet},
##   }
##
## ATTENTION: This citation information has been auto-generated from the package
## DESCRIPTION file and may need manual editing, see 'help("citation")'.
```

---

<sup>\*</sup>tiphaine.martin@kcl.ac.uk

<sup>†</sup>idil.yet@kcl.ac.uk

<sup>‡</sup>peichien.tsai@kcl.ac.uk

<sup>§</sup>jordana.bell@kcl.ac.uk

## Contents

---

<b>1</b>	<b>Citation</b>	<b>1</b>
<b>2</b>	<b>Introduction</b>	<b>3</b>
<b>3</b>	<b>Usage</b>	<b>3</b>
<b>4</b>	<b>Files formats</b>	<b>4</b>
4.1	Format of info file (mandatory) . . . . .	4
4.2	Format of correlation matrix (mandatory) . . . . .	5
4.3	Format of extra info file . . . . .	6
4.4	Format of annotation file . . . . .	6
4.5	Option of config.file . . . . .	6
<b>5</b>	<b>Creating a plot like webservice</b>	<b>8</b>
<b>6</b>	<b>Creating a plot with the generic function: comet</b>	<b>8</b>
6.1	coMET plot: pvalue plot, annotation tracks, and correlation matrix . . . . .	8
6.2	coMET plot: annotation tracks and correlation matrix . . . . .	11

## 2 Introduction

---

The CoMET package is a web-based plotting tool and R-based package to visualize EWAS (epigenome-wide association scan) results in a genomic region of interest. CoMET provides a plot of the EWAS association signal and visualisation of the methylation correlation between CpG sites (co-methylation). The CoMET package also provides the option to annotate the region using functional genomic information, including both user-defined features and pre-selected features based on the Encode project. The plot can be customized with different parameters, such as plot labels, colours, symbols, heatmap colour scheme, significance thresholds, and including reference CpG sites. Finally, the tool can also be applied to display the correlation patterns of other genomic data, e.g. gene expression array data.

coMET generates a multi-panel plot to visualize EWAS results, co-methylation patterns, and annotation tracks in a genomic region of interest. A coMET figure (cf. Fig. 1) includes three components:

1. the upper plot shows the strength and extent of EWAS association signal;
2. the middle panel provides customized annotation tracks;
3. the lower panel shows the correlation between selected CpG sites in the genomic region.

The structure of the plots builds on `snp.plotter` (Luna et al., 2007), with extensions to incorporate genomic annotation tracks and customized functions. coMET produces plots in PDF and Encapsulated Postscript (EPS) format.

## 3 Usage

---

CoMET requires the installation of R, the statistical computing software, freely available for Linux, Windows, or MacOS. CoMET can be downloaded from bioconductor. Packages can be installed using the `install.packages` command in R. The coMET R package includes two major functions `comet.web` and `comet`. The function `comet.web` generates output plot with the same settings of genomic annotation tracks as that of the webservice (<http://comet.epigen.kcl.ac.uk:3838/coMET/> or <http://www.epigen.kcl.ac.uk/comet>). The function `comet` generates output plots with the customized annotation tracks defined by user.

```
source("http://bioconductor.org/biocLite.R")
biocLite("coMET")
```

CoMET can be loaded into R using this command:

```
library(coMET)

## Loading required package: grid
## Loading required package: biomaRt
## Loading required package: Gviz
## Loading required package: BiocGenerics
## Loading required package: parallel
##
## Attaching package: 'BiocGenerics'
##
## The following objects are masked from 'package:parallel':
##
##   clusterApply, clusterApplyLB, clusterCall, clusterEvalQ, clusterExport,
```

```
##      clusterMap, parApply, parCapply, parLapply, parLapplyLB, parRapply,
##      parSapply, parSapplyLB
##
## The following object is masked from 'package:stats':
##
##      xtabs
##
## The following objects are masked from 'package:base':
##
##      anyDuplicated, append, as.data.frame, as.vector, cbind, colnames, do.call,
##      duplicated, eval, evalq, Filter, Find, get, intersect, is.unsorted, lapply,
##      Map, mapply, match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##      Position, rank, rbind, Reduce, rep.int, rownames, sapply, setdiff, sort,
##      table, tapply, union, unique, unlist
##
## Loading required package:  rtracklayer
## Loading required package:  GenomicRanges
## Loading required package:  IRanges
## Loading required package:  GenomeInfoDb
## Warning:  replacing previous import by 'hash::keys' when loading 'coMET'
```

The configuration file specifies the options for the coMET plot. Example configuration and input files are also provided on <http://www.epigen.kcl.ac.uk/comet>. Information about the package can be viewed from within R using this command:

```
?comet
?comet.web
```

## 4 Files formats

---

There are four types of files that the user should or can give to produce the plot:

1. info file is defined in the option DATA.FILE (mandatory)
2. correlation files is defined in the option CORMATRIX.FILE (mandatory)
3. extra info files are defined in the option DATA.FILE.LARGE.
4. Annotation info file is defined in the option BIOFEAT.USER.FILE.

### 4.1 Format of info file (mandatory)

Info file can be a list of CpG sites with/without Beta value (or direction sign). If it is a site file then it is mandatory to have the 4 columns as shown below with headers in the same order. Beta can be the 5th column(optional) and it can be either a numeric value (positive or negative values) or only direction sign ("+", "-")

```
extdata <- system.file("extdata", package="coMET", mustWork=TRUE)
infofile <- file.path(extdata, "cyp1b1_infofile.txt")
```

```
data_info <- read.csv(infile, header = TRUE,
                     sep = "\t", quote = "")
```

```
head(data_info)
```

```
##      TargetID CHR  MAPINFO      Pval
## 1 cg22248750   2 38294160 2.750e-01
## 2 cg11656478   2 38297759 7.795e-01
## 3 cg14407177   2 38298023 2.864e-01
## 4 cg02162897   2 38300537 3.148e-07
## 5 cg20408276   2 38300586 1.468e-06
## 6 cg00565882   2 38300707 7.563e-03
```

Alternatively, the info file can be region-based and if so, the region-based info file must have the 5 columns (see below) with headers in this order. The beta or direction can be included in the 6th column (optional).

```
extdata <- system.file("extdata", package="coMET", mustWork=TRUE)
infoexp <- file.path(extdata, "cyp1b1_infofile_exprGene_region.txt")
```

```
data_infoexp <- read.csv(infoexp, header = TRUE,
                       sep = "\t", quote = "")
```

```
head(data_infoexp)
```

```
##                                     TargetID CHR MAPINFO.START MAPINFO.STOP      Pval BETA
## 1 ENSG00000138061.7_38294652_38298453   2      38294652      38298453 3.064e-17   +
## 2 ENSG00000138061.7_38301489_38302532   2      38301489      38302532 1.145e-07   +
## 3 ENSG00000138061.7_38302919_38303323   2      38302919      38303323 1.014e-08   -
```

## 4.2 Format of correlation matrix (mandatory)

The data file used for the correlation matrix is described in the option CORMATRIX.FILE. This tab-delimited file can take 2 formats described in the option CORMATRIX.FORMAT:

1. CORMATRIX: pre-computed correlation matrix provided by the user; Dimension of matrix : CpG\_number X CpG\_number
2. RAW: Raw data format. Correlations of these can be computed by one of 3 methods Spearman, Pearson, Kendall (option CORMATRIX.METHOD). Dimension of matrix : sample\_size X CpG\_number ;

```
extdata <- system.file("extdata", package="coMET", mustWork=TRUE)
corfile <- file.path(extdata, "cyp1b1_res37_rawMatrix.txt")
```

```
data_cor <- read.csv(corfile, header = TRUE,
                   sep = "\t", quote = "")
```

```
data_cor[1:6, 1:6]
```

```
##      cg22248750 cg11656478 cg14407177 cg02162897 cg20408276 cg00565882
## 1 -0.086368    -0.4897      1.6719      0.52423      0.1659      0.224222
## 2 -0.001079    -0.6331      0.3151     -0.29821     -0.4339     -0.007795
```

## 3	0.316569	-0.2610	-0.4943	0.04657	0.1840	0.313967
## 4	-0.409150	0.6816	-0.3251	-0.58656	-0.2070	0.150720
## 5	1.299533	0.3986	0.1119	0.81182	0.1833	0.194928
## 6	-1.119488	0.3036	-1.2795	-0.49785	0.1076	-0.876012

### 4.3 Format of extra info file

The path file of extra info files can be described in the option `DATA.FILE.LARGE`. Different extra info files are separated by a comma.

This can be another type of info file (e.g expression or replication data) and should follow the same rules as the standard info file.

### 4.4 Format of annotation file

The path of file is defined in the option `BIOFEAT.USER.FILE` and the format of file is the format accepted by GViz (BED, GTF, and GFF3).

### 4.5 Option of config.file

If you would like to make your own changes to the plot you can download the configuration file, make changes to it, and upload it into R as shown in the example below.

The important options of a coMET figure include three components:

1. The upper plot shows the strength and extent of EWAS association signal.
  - `PVAL.THRESHOLD` : Significance threshold to be displayed as a red dashed line
  - `DISP.ASSOCIATION` : Optional direction of association if `MYDATA.FILE=SITE_ASSOC` or `REGION_ASSOC`. The value can be `TRUE` or `FALSE`: if `FALSE` (default), the co-methylation pattern is shown in the p-value plot; if `TRUE`, the effect direction is shown
  - `DISP.REGION` : Optional if `MYDATA.FILE=REGION` or `REGION_ASSOC`. The value can be `TRUE` or `FALSE` (default). If `TRUE`, the genomic element will be shown by a continuous line with the color of the element, in addition to the symbol at the center of the region
2. The middle panel provides customized annotation tracks;
  - `LIST.TRACKS` (for `comet.web` function): List of annotation tracks that can be visualised: `geneENSEMBL`, `CGI`, `ChromHMM`, `DNase`, `RegENSEMBL`, `SNP`, `transcriptENSEMBL`, `SNPstoma`, `SNPstru`, `SNPstrustoma`, `ISCA`, `COSMIC`, `GAD`, `ClinVar`, `GeneReviews`, `GWAS`, `Clin-VarCNV`, `GCcontent`, `genesUCSC`, `xenogenesUCSC`.
  - `TRACKS.GVIZ`, `TRACKS.GGBIO`, `TRACKS.TRACKVIEWER` (for `comet` function): For each option, it is possible to give a list of annotation tracks that is created by the `Gviz`, `GGBio`, and `TrackViewer` bioconductor packages.
3. The lower panel shows the correlation between selected CpG sites in the genomic region.
  - `CORMATRIX.FORMAT` : Format of the input file `CORMATRIX.FILE`: either raw data (option `RAW`) or correlation matrix (option `CORMATRIX`)
  - `CORMATRIX.METHOD` : If raw data are provided it will be necessary to produce the correlation matrix using one of 3 methods (`spearman`, `pearson` and `kendall`).
  - `CORMATRIX.COLOR.SCHEME` : There are 5 colors (`heat`, `bluwhitered`, `cm`, `topo`, `gray`, `bluetored`)

```
extdata <- system.file("extdata", package="coMET", mustWork=TRUE)
configfile <- file.path(extdata, "config_cyp1b1_zoom_4webserver.txt")

data_config <-read.csv(configfile, quote = "")
data_config

##                               DISP.MYDATA.TRUE
## 1                            MYDATA.FORMAT=SITE
## 2                             MYDATA.REF=cg02162897
## 3                      PVAL.THRESHOLD=4.720623e-06
## 4                    DISP.ASSOCIATION=FALSE
## 5                     DISP.REGION=FALSE
## 6             MYDATA.LARGE.FORMAT=REGION ASSO
## 7            DISP.ASSOCIATION.LARGE=TRUE
## 8          DISP.REGION.LARGE=TRUE
## 9      SAMPLE.LABELS.LARGE=Gene expression
## 10                COLOR.LIST.LARGE=green
## 11              SYMBOLS.LARGE=diamond-fill
## 12                   START=38290160
## 13                   END=38303219
## 14           SAMPLE.LABELS=CpG
## 15         SYMBOLS=circle-fill
## 16                 LAB.Y=log
## 17       DISP.COLOR.REF=TRUE
## 18     CORMATRIX.FORMAT=RAW
## 19   DISP.CORMATRIXMAP=TRUE
## 20       CORMATRIX.METHOD=spearman
## 21   CORMATRIX.COLOR.SCHEME=bluewhitered
## 22           DISP.PHYS.DIST=TRUE
## 23        DISP.COLOR.BAR=TRUE
## 24           DISP.TYPE=symbol
## 25           DISP.LEGEND=TRUE
## 26    LIST.TRACKS=transcriptENSEMBL
## 27                                     CGI
## 28                                   ChromHMM
## 29                                   DNase
## 30                                RegENSEMBL
## 31                                    SNP
## 32        DISP.MULT.LAB.X=FALSE
## 33            IMAGE.TYPE=pdf
## 34 IMAGE.TITLE="Example a-DMR in CYP1B1 in Adipose tissue"
## 35           IMAGE.NAME=cyp1b1_zoom_plus_name_expr
## 36             IMAGE.SIZE=3.5
## 37               GENOME=hg19
## 38   DATASET.GENE=hsapiens_gene_ensembl
## 39   DATASET.SNP=hsapiens_snp
## 40   VERSION.DBSP=snp138
## 41   DATASET.SNP.STOMA=hsapiens_snp_som
```

```
## 42          DATASET.REGULATION=hsapiens_feature_set
## 43          DATASET.STRU=hsapiens_structvar
## 44          DATASET.STRU.STOMA=hsapiens_structvar_som
## 45          PATTERN.REGULATION=GM12878
## 46          BROWSER.SESSION=UCSC
```

## 5 Creating a plot like webservice

---

User can draw coMET via the coMET website (<http://epigen.kcl.ac.uk/comet>). It is possible to reproduce the web service plotting defaults by using the function `comet.web`, for example see Figure 1.

```
extdata <- system.file("extdata", package="coMET", mustWork=TRUE)
myinfofile <- file.path(extdata, "cyp1b1_infofile.txt")
myexpressfile <- file.path(extdata, "cyp1b1_infofile_exprGene_region.txt")
mycorrelation <- file.path(extdata, "cyp1b1_res37_rawMatrix.txt")
configfile <- file.path(extdata, "config_cyp1b1_zoom_4webserver.txt")
comet.web(config.file=configfile, MYDATA.FILE=myinfofile,
          CORMATRIX.FILE=mycorrelation, MYDATA.LARGE.FILE=myexpressfile,
          PRINT.IMAGE=FALSE, VERBOSE=FALSE)
```

## 6 Creating a plot with the generic function: comet

---

It is possible to create the annotation tracks by Gviz, trackviewer or ggbio, for example see Figure 2. Currently, the Gviz option for annotation tracks, in combination with the heatmap of correlation values between genetics elements, provides the most informative and easy approach to visualize graphics.

### 6.1 coMET plot: pvalue plot, annotation tracks, and correlation matrix

```
extdata <- system.file("extdata", package="coMET", mustWork=TRUE)
configfile <- file.path(extdata, "config_cyp1b1_zoom_4comet.txt")
myinfofile <- file.path(extdata, "cyp1b1_infofile.txt")
myexpressfile <- file.path(extdata, "cyp1b1_infofile_exprGene_region.txt")
mycorrelation <- file.path(extdata, "cyp1b1_res37_rawMatrix.txt")

chrom <- "chr2"
start <- 38290160
end <- 38303219
gen <- "hg19"
strant <- "*"

BROWSER.SESSION="UCSC"
mySession <- browserSession(BROWSER.SESSION)
genome(mySession) <- gen
```



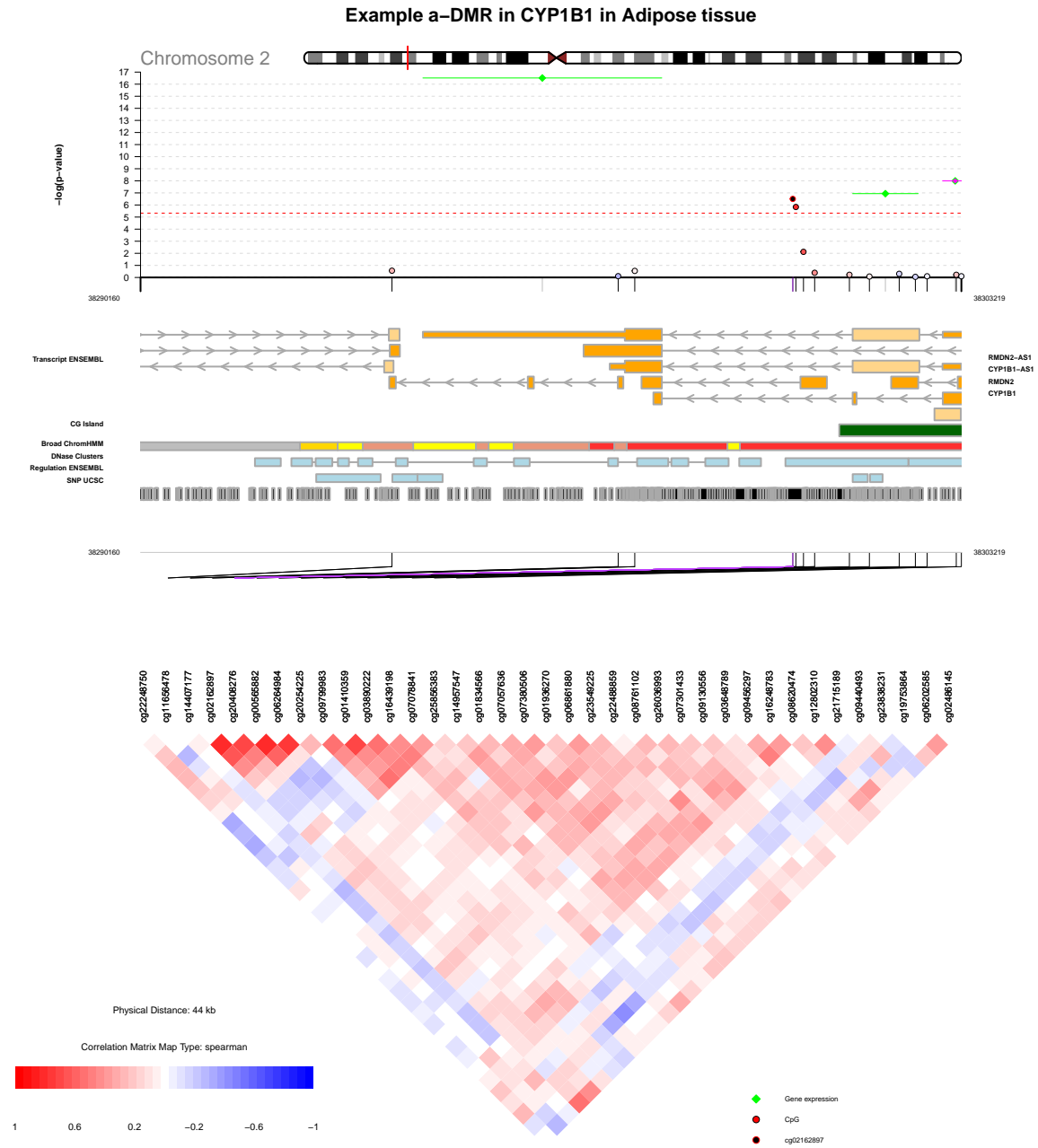


Figure 1: Plot with comet.web function.

```
genetrack <- genesENSEMBL(gen, chrom, start, end, showId=FALSE)
snptrack <- snpBiomart(chrom, start, end, dataset="hsapiens_snp_som", showId=FALSE)
strutrack <- structureBiomart(chrom, start, end, strand,
                             dataset="hsapiens_structvar_som", showId=FALSE)
iscatrack <- ISCATrack(gen, chrom, start, end, mySession, table="iscaPathogenic")

listgviz <- list(genetrack, snptrack, iscatrack)
```

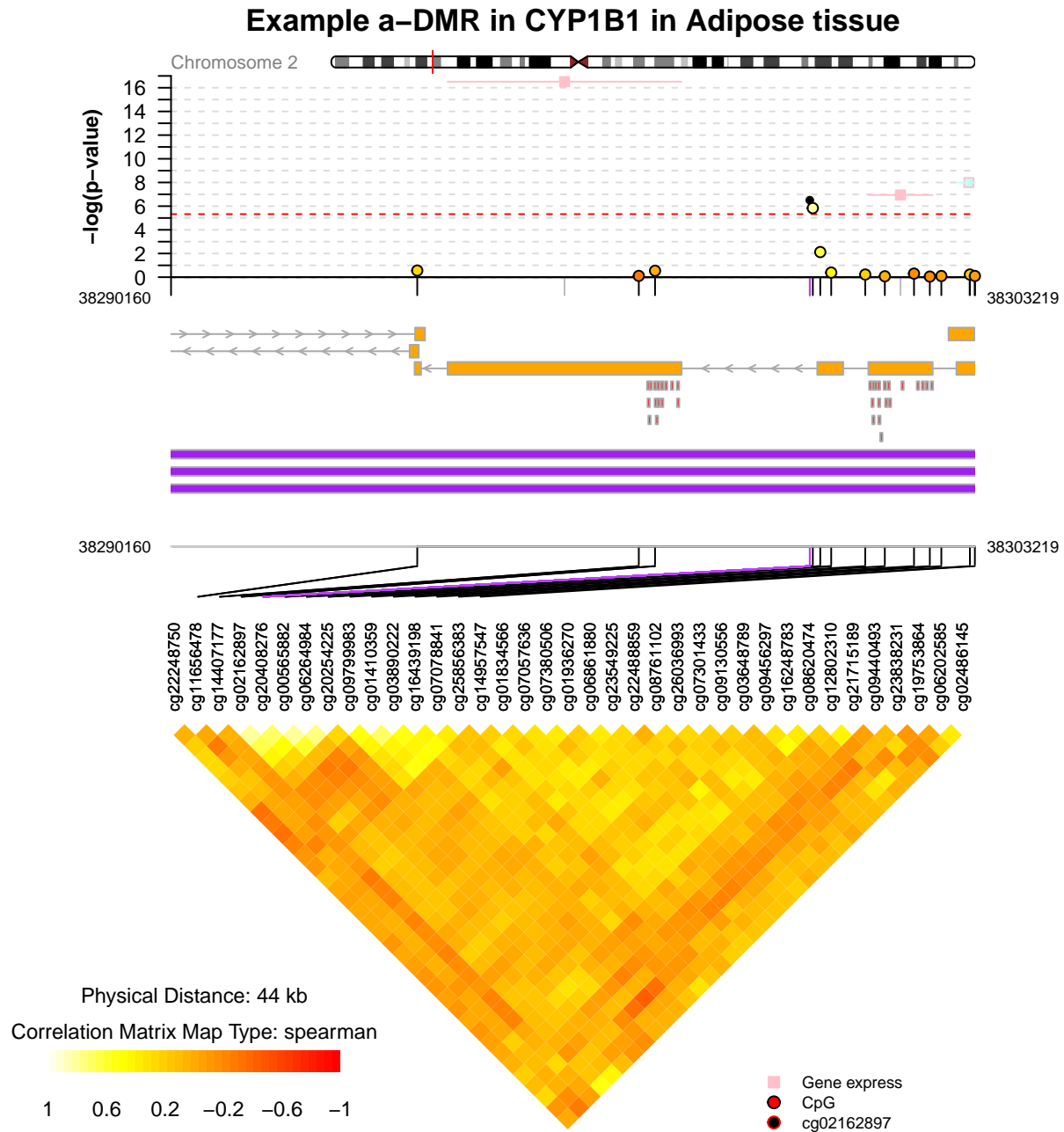


Figure 2: Plot with comet function.

```
comet(config.file=configfile, MYDATA.FILE=myinfofile, CORMATRIX.FILE=mycorrelation,
      MYDATA.LARGE.FILE=myexpressfile, TRACKS.GVIZ=listgviz,
      VERBOSE=FALSE, PRINT.IMAGE=FALSE)
```

## 6.2 coMET plot: annotation tracks and correlation matrix

It is possible to visualise only annotation tracks and the correlation between genetic elements. In this case, we need to use the option `DISP.PVALUEPLOT=FALSE`, for example see Figure 3.

```
extdata <- system.file("extdata", package="coMET", mustWork=TRUE)
configfile <- file.path(extdata, "config_cyp1b1_zoom_4cometnopval.txt")
myinfofile <- file.path(extdata, "cyp1b1_infofile.txt")
mycorrelation <- file.path(extdata, "cyp1b1_res37_rawMatrix.txt")
#configfile <- "../inst/extdata/config_cyp1b1_zoom_4comet.txt"

chrom <- "chr2"
start <- 38290160
end <- 38303219
gen <- "hg19"

genetrack <- genesENSEMBL(gen, chrom, start, end, showId=FALSE)
snptrack <- snpBiomart(chrom, start, end,
                      dataset="hsapiens_snp_som", showId=FALSE)
strutrack <- structureBiomart(chrom, start, end,
                              strand, dataset="hsapiens_structvar_som")
clinVariant <- ClinVarMainTrack(gen, chrom, start, end)
clinCNV <- ClinVarCnvTrack(gen, chrom, start, end)
gwastrack <- GWASTrack(gen, chrom, start, end)
geneRtrack <- GeneReviewsTrack(gen, chrom, start, end)

listgviz <- list(genetrack, snptrack, strutrack, clinVariant,
                 clinCNV, gwastrack, geneRtrack)
comet(config.file=configfile, MYDATA.FILE=myinfofile, CORMATRIX.FILE=mycorrelation,
      TRACKS.GVIZ=listgviz, VERBOSE=FALSE, PRINT.IMAGE=FALSE, DISP.PVALUEPLOT=FALSE)
```

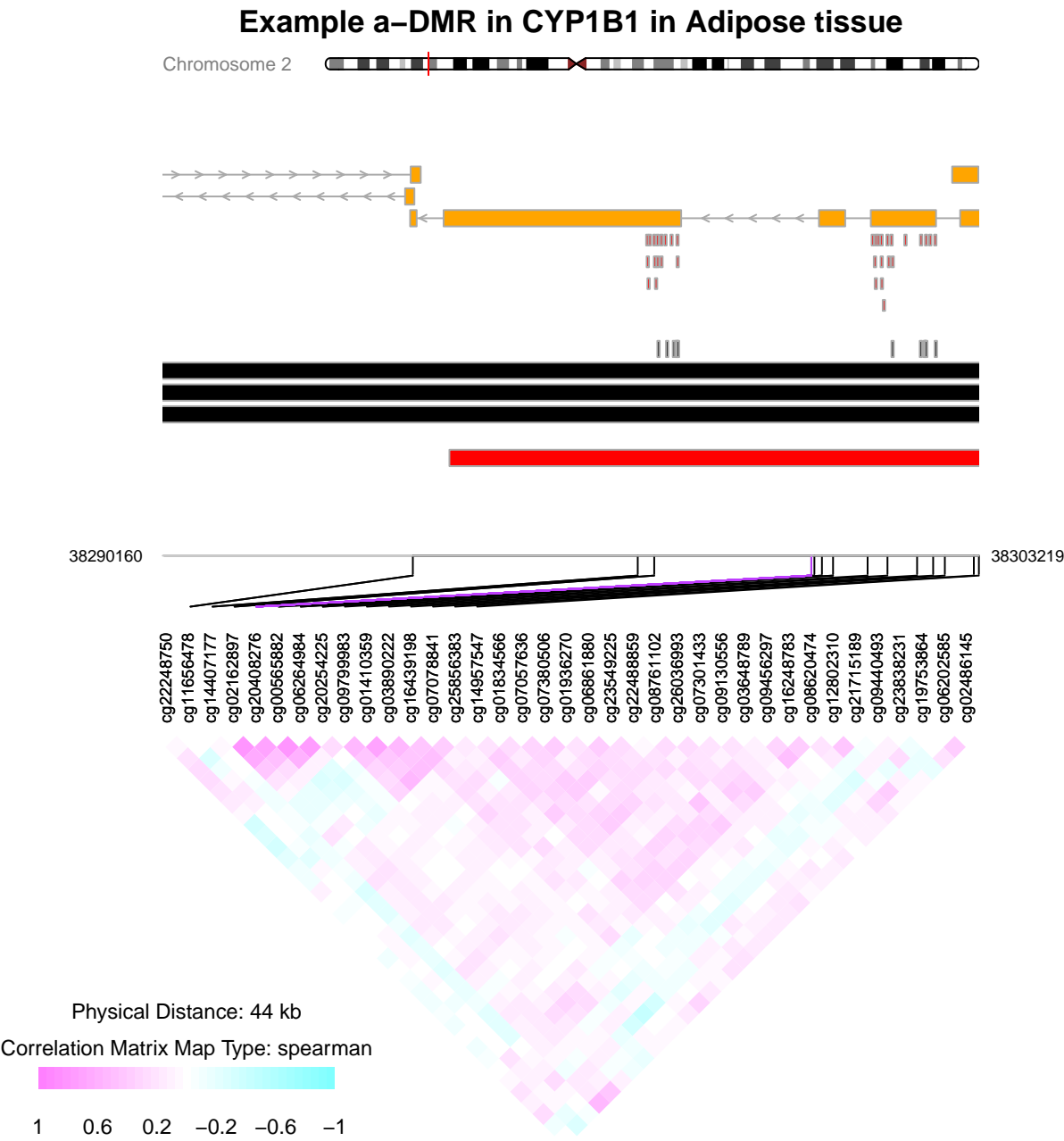


Figure 3: Plot with comet function without pvalue plot.

## SessionInfo

---

The following is the session info that generated this vignette:

```
toLatex(sessionInfo())
```

- R version 3.1.1 (2014-07-10), x86\_64-pc-linux-gnu
- Locale: LC\_CTYPE=en\_GB.UTF-8, LC\_NUMERIC=C, LC\_TIME=en\_US.UTF-8, LC\_COLLATE=en\_GB.UTF-8, LC\_MONETARY=en\_US.UTF-8, LC\_MESSAGES=en\_GB.UTF-8, LC\_PAPER=en\_US.UTF-8, LC\_NAME=C, LC\_ADDRESS=C, LC\_TELEPHONE=C, LC\_MEASUREMENT=en\_US.UTF-8, LC\_IDENTIFICATION=C
- Base packages: base, datasets, graphics, grDevices, grid, methods, parallel, stats, utils
- Other packages: BiocGenerics 0.10.0, biomaRt 2.20.0, coMET 0.99.3, GenomeInfoDb 1.0.2, GenomicRanges 1.16.4, Gviz 1.8.4, IRanges 1.22.10, knitr 1.6, rtracklayer 1.24.2, XVector 0.4.0
- Loaded via a namespace (and not attached): acepack 1.3-3.3, AnnotationDbi 1.26.1, base64enc 0.1-2, BatchJobs 1.4, BBmisc 1.7, Biobase 2.24.0, BiocParallel 0.6.1, BiocStyle 1.2.0, Biostrings 2.32.1, biovizBase 1.12.3, bitops 1.0-6, brew 1.0-6, BSgenome 1.32.0, checkmate 1.4, cluster 1.15.3, codetools 0.2-9, colorspace 1.2-4, colortools 0.1.5, DBI 0.3.1, dichromat 2.0-0, digest 0.6.4, evaluate 0.5.5, fail 1.2, foreach 1.4.2, foreign 0.8-61, formatR 1.0, Formula 1.1-2, GenomicAlignments 1.0.6, GenomicFeatures 1.16.3, ggbio 1.12.10, ggplot2 1.0.0, gridExtra 0.9.1, gtable 0.1.2, gWidgets 0.0-54, gWidgetstcltk 0.0-55, hash 2.2.6, highr 0.3, Hmisc 3.14-5, iterators 1.0.7, lattice 0.20-29, latticeExtra 0.6-26, MASS 7.3-34, matrixStats 0.10.0, munsell 0.4.2, nnet 7.3-8, pbapply 1.1-1, plyr 1.8.1, proto 0.3-10, RColorBrewer 1.0-5, Rcpp 0.11.3, RCurl 1.95-4.3, reshape2 1.4, R.methodsS3 1.6.1, rpart 4.1-8, Rsamtools 1.16.1, RSQLite 0.11.4, scales 0.2.4, sendmailR 1.2-1, splines 3.1.1, stats4 3.1.1, stringr 0.6.2, survival 2.37-7, tcltk 3.1.1, tools 3.1.1, trackViewer 1.0.2, VariantAnnotation 1.10.5, XML 3.98-1.1, zlibbioc 1.10.0