The coMET User Guide

Tiphaine C. Martin [1]***, Tom Hardiman, Idil Yet, Pei-Chien Tsai, Jordana T. Bell***

[1] tiphaine.martin@mssm.edu

Edited: December 2018; Compiled: November 1, 2022

1 Citation

```
citation(package='coMET')
```

```
##
## To cite 'coMET' in publications use:
##
##   Martin, T., Erte, I, Tsai, P-C, Bell, J.T. coMET: an R plotting package to
##   visualize regional plots of epigenome-wide association scan results QG14, 2014
##
##   Martin, T., Yet, I, Tsai, P-C, Bell, J.T. coMET: visualisation of regional
##   epigenome-wide association scan results and DNA co-methylation patterns BMC
##   Bioinformatics, 2015 (accepted)
##
## To see these entries in BibTeX format, use 'print(<citation>, bibtex=TRUE)',
## 'toBibtex(.)', or set 'options(citation.bibtex.max=999)'.
```

Contents

Abstract

The *coMET* package is a web-based plotting tool and R-based package to visualize omic-WAS results in a genomic region of interest, such as EWAS (epigenome-wide association scan). *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 in any species, 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*[1], with extensions to incorporate genomic annotation tracks and customized functions. *coMET* produces plots in PDF and Encapsulated Postscript (EPS) format.

The current version of *coMET* can visualise EWAS results and annotations from a genomic region up to an entire chromosome in the upper and middle panels of the coMET plot. However, the lower panel (co-methylation) is restricted to visualising a maximum of 120 single-CpG or region-based datapoints. This limitation is due to limitations in the size of a standard A4 plot, and may be updated in the near future. However, the user can use the function `comet.list` to extracts all significant correlations beyond a given threshold in the dataset from either a genomic region or from an entire chromosome if required.

2 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` to visualise omci-WAS results.

- The function `comet.web` generates output plot with the same settings of genomic annotation tracks as that of the webservice (http://epigen.kcl.ac.uk/comet or direcly http://comet.epigen.kcl.ac.uk:3838/coMET/).
- The function `comet` generates output plots with the customized annotation tracks defined by user.

```
if (!requireNamespace("BiocManager", quietly=TRUE))
  install.packages("BiocManager")
BiocManager::install("coMET")
```

coMET uses the packages called *psych*, *corrplot* and *colortools*, which are not available from bioconductor. This must be installed before the installation of *coMET*

```
install.packages("psych")
install.packages("corrplot")
install.packages("colortools")
```

coMET has a development version on gitHub, go to the section "Install the development version of *coMET* from *Bioconductor*".

You can install also on the version R 3.2.2 via the master version of package on gitHub. The same steps must be followed as described in the section "Install the development version of *coMET* from *Bioconductor*".

After downloading from *Bioconductor* or gitHUB, and installing on your computer, *coMET* can be loaded into a R session using this command:

```
library("coMET")
```

```
## Loading required package:  grid
```

```
## Loading required package:  biomaRt
```

```
## Loading required package:  Gviz
```

```
## Loading required package:  S4Vectors
```

```
## Loading required package:  stats4
```

```
## 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':
##
##     Filter, Find, Map, Position, Reduce, anyDuplicated, aperm, append,
##     as.data.frame, basename, cbind, colnames, dirname, do.call, duplicated,
##     eval, evalq, get, grep, grepl, intersect, is.unsorted, lapply, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int, rank, rbind,
##     rownames, sapply, setdiff, sort, table, tapply, union, unique, unsplit,
##     which.max, which.min

##
## Attaching package:  'S4Vectors'

## The following objects are masked from 'package:base':
##
##     I, expand.grid, unname

## Loading required package:  IRanges

## Loading required package:  GenomicRanges

## Loading required package:  GenomeInfoDb

## Loading required package:  psych

##
## Attaching package:  'psych'

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

## The following objects are masked from 'package:IRanges':
##
##     distance, reflect
```

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

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

2.1 Install the development version of coMET from Bioconductor

To install *coMET* from the development version of Bioconductor, the user must install the appropriate R version. See http://www.bioconductor.org/developers/how-to/useDevel/ for more details. Following this installation, use the standard Bioconductor command:

```
if (!requireNamespace("BiocManager", quietly=TRUE))
  install.packages("BiocManager")
BiocManager::install(version = "devel")
BiocManager::install("coMET")
```

2.2 Install the version of coMET from gitHub

Another way to install *coMET* is to download the master package from gitHUB https://github.com/TiphaineCMartin/coMET or the devel package https://github.com/TiphaineCMartin/coMET/tree/devel. Once downloaded use command line:

```
install.packages("YourPath/coMET_YourVersion.tar.gz",repos=NULL,type="source")
##This is an example
install.packages("YourPath/coMET_0.99.9.tar.gz",repos=NULL,type="source")
```

3 Functions in coMET

Currently, there are 3 main functions:

1. `comet.web` is the pre-customized function that allows us to visualise quickly EWAS (or other omic-WAS) results, annotation tracks, and correlations between features. This version is installed in the Shiny web-service. Currently, it is formated only to visualise human data.
2. `comet` is the generic function that allows us to visualise quickly EWAS results, annotation tracks, and correlations between features. Users can visualise more personalised annotation tracks and give multiple extra EWAS/omic-WAS results to plot.
3. `comet.list` is an additional function that allows us to extract the values of correlations, the pvalues, and estimates and confidence intervals for all datapoints that surpass a particular threshold.

The functions can read the data input files, but it is also possible to use data frames within R for all data input except for the configuration file. The latter can be achieved with the two functions `comet` and `comet.list`. The structure of the data frames (number of columns, type, format) follows the same rules as for the data input files (cf. section "File formats").

4 File formats

There are five types of files that can be given by the user to produce the plot:

1. Info file is defined in the option `mydata.file`. *Warning: This is mandatory and has to be in tabular format with a header*.
2. Correlation file is defined in the option `cormatrix.file`. *Warning: This is mandatory and has to be in tabular format with a header*.
3. Extra info files are defined in the option `mydata.file.large`. *Warning: This is optional, and if provided has to be in tabular format with a header*.
4. Annotation info file is defined in the option `biofeat.user.file`. This option exists only in the function `comet.web` and the user should inform also the format to visualise this data with the options `biofeat.user.type` and `biofeat.user.type.plot`.
5. Configuration file contains the values of these options instead of defining these by command line. *Warning: Each line in the file is one option. The name of the option is in capital letters and is separated by its value by "="*. If there are multiple values such as for the option `list.tracks` or the options for additional data, you need to separated them by a "comma".

4.1 Format of the info file (for option: `mydata.file`, mandatory)

Warning: This file is mandatory and has to be in tabular format with a header. The name of features has to start by a letter. Info files can be a list of CpG sites with/without Beta value (for example DNA methylation level) 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 can be either a numeric value (positive or negative values) or only direction sign ("+", "-"). The number of columns and their types are defined by the option `mydata.format`.

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

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

head(data_info)
```

```
##     TargetID CHR  MAPINFO         Pval
## 1 cg22248750   2 38294160 2.749858e-01
## 2 cg11656478   2 38297759 7.794549e-01
## 3 cg14407177   2 38298023 2.863869e-01
## 4 cg02162897   2 38300537 3.148201e-07
## 5 cg20408276   2 38300586 1.467739e-06
```

```
## 6 cg00565882   2 38300707 7.563132e-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 3.064357e-17    +
## 2 ENSG00000138061.7_38301489_38302532   2      38301489     38302532 1.145430e-07    +
## 3 ENSG00000138061.7_38302919_38303323   2      38302919     38303323 1.014050e-08    -
```

In summary, there are 4 possible formats for the info file:

1. ***site***: 4 columns with a header:
   (a) Name of omic feature
   (b) Name of chromosome
   (c) Position of omic feature
   (d) P-value of omic feature
2. ***region***: 5 columns with a header:
   (a) Name of omic feature
   (b) Name of chromosome
   (c) Start position of omic feature
   (d) End position of omic feature
   (e) P-value of omic feature
3. ***site_asso***: 5 columns with a header:
   (a) Name of omic feature
   (b) Name of chromosome
   (c) Position of omic feature
   (d) P-value of omic feature
   (e) Direction of association related to this omic feature. This can be the sign or an actual value of association effect size.

4. ***region_asso***: 6 columns with a header:
   (a) Name of omic feature
   (b) Name of chromosome
   (c) Start position of omic feature
   (d) End position of omic feature
   (e) P-value of omic feature
   (f) Direction of association related to this omic feature. This can be the sign or an actual value of association effect size.

4.2 Format of correlation matrix (for option: `cormatrix.file`, mandatory)

Warning: This file is mandatory and has to be in tabular format with an header. The data file used for the correlation matrix is described in the option `cormatrix.file`. This tab-delimited file can take 3 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. Need to put the CpG sites/regions in the ascending order of positions and to have a header with the name of CpG sites/regions;
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. Need to have a header with the name of CpG sites/regions ;
3. ***raw_rev***: Raw data format. Correlations of these can be computed by one of 3 methods Spearman, Pearson, Kendall (option `cormatrix.method`). Dimension of matrix : CpG_number X sample_size. Need to have the row names of CpG sites/regions and a header with the name of samples ;

```
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 -0.08636815 -0.4896557  1.6718967  0.52423342  0.1659252  0.224221521
## 2 -0.00107899 -0.6330666  0.3150612 -0.29820805 -0.4339332 -0.007794883
## 3  0.31656883 -0.2610083 -0.4942691  0.04657351  0.1840397  0.313967471
## 4 -0.40914999  0.6816058 -0.3251337 -0.58656175 -0.2069954  0.150719803
## 5  1.29953262  0.3985525  0.1119045  0.81181511  0.1833470  0.194928273
```

```
## 6 -1.11948826  0.3035820 -1.2794597 -0.49785237  0.1076348 -0.876011670
```

4.3 Format of extra info file (for option: `mydata.large.file`)

Warning: This file is optional file and if provided has to be in tabular format with an header. The name of features has to start by a letter. The extra info files can be described in the option `mydata.large.file` and their format in `mydata.large.format`. More than one extra info file can be used, each should be 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 (for option `biofeat.user.file`)

The 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

Warning: Each line in the file is one option. The name of the option is in lowercase letters and is separated by its value by "=" without space. If there are multiple values such as for the option `list.tracks` or options for additional data, these need to be separated them by a "comma" withou space. 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 on a regional Manhattan plot.
    - `pval.threshold`: Significance threshold to be displayed as a red dashed line
    - `pval.threshold2`: Another Significance threshold (optional)
    - `disp.pvalueplot`: Value can be TRUE or FALSE. Used to either display or hide Manhattan plot.
    - `disp.beta.association`: Value can be TRUE or FALSE. Used to show the effect size.
    - `disp.association`: This logical option works only if `mydata.file` contains the effect direction (`mydata.format`=***site_asso*** or ***region_asso***). The value can be TRUE or FALSE: if FALSE (default), for each point of data in the p-value plot, the colour of symbol is the colour of co-methylation pattern between the point and the reference site; if TRUE, the effect

direction is shown. If the association is positive, the colour is the one defined with the option `color.list`. On the other hand, if the association is negative, the colour is the inverse to that selected.

- `disp.region` : This logical option works only if the option `mydata.file` contains regions (`mydata.format`=***region*** or ***region_asso***). The value can be TRUE or FALSE (default). If TRUE, the genomic element will be shown as a continuous line with the colour of the element, in addition to the symbol at the center of the region. If FALSE, only the symbol is shown.

2. The ***middle panel*** provides customized annotation tracks;

   - `list.tracks` (for `comet.web` function): List of annotation tracks to be visualised. Tracks currently available: geneENSEMBL, CGI, ChromHMM, DNAse, RegENSEMBL, SNP, transcriptENSEMBL, SNPstoma, SNPstru, SNPstrustoma, GAD, ClinVar, GeneReviews, GWAS, ClinVarCNV, GC-content, genesUCSC, xenogenesUCSC, metQTL, eQTL, BindingMotifs-Biomart, chromHMM_RoadMap, miRNATargetRegionsBiomart, Other-RegulatoryRegionsBiomart, RegulatoryEvidenceBiomart and segmentalDups-sUCSC. The elements are separated by a comma.

   - `tracks.gviz` (for `comet` function): For each option, it is possible to give a list of annotation tracks that is created by the *Gviz* bioconductor packages. *Warning: It is noted that the new version of coMET does not support more the visualisation using `tracks.ggbio` and `tracks.trackviewer` from GGBio and TrackViewer because The integration of plots from ggbio and trackviewer can be sometimes not really perfect. So only now, it is possible to create plots from Gviz and use `tracks.gviz`*

3. The ***lower panel*** shows the correlation between selected CpG sites in the genomic region (heatmap).

   - `cormatrix.format` : Format of the input file ***cormatrix.file***: either raw data (option RAW if CpG sites are by column and samples by row or option RAW_REV if CpG site are by row and samples by column) 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 colour schemes (heat, bluewhitered, cm, topo, gray, bluetored)

   - `disp.cormatrixmap` : logical option TRUE or FALSE. TRUE (default), if FALSE correlation matrix is not shown

   - `cormatrix.conf.level` : Alpha level for the confidence interval. Default value= 0.05. CI will be the alpha/2 lower and upper values.

- `cormatrix.sig.level` : Significant level to visualise the correlation. If the correlation has a pvalue under the significant level, the correlation will be colored in "goshwhite", else the color is related to the correlation level and the color scheme choosen.Default value =1.
- `cormatrix.adjust` : Indicates which adjustment for multiple tests should be used. "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none".Default value="none".

```
extdata <- system.file("extdata", package="coMET",mustWork=TRUE)
configfile <- file.path(extdata, "config_cyp1b1_zoom_4webserver_Grch38.txt")

data_config <-read.csv(configfile, quote = "", sep="\t", header=FALSE)
data_config
```

```
##                                                    V1
## 1                                  disp.mydata=TRUE
## 2                                 mydata.format=site
## 3                                  sample.labels=CpG
## 4                                symbols=circle-fill
## 5                                          lab.Y=log
## 6                                disp.color.ref=TRUE
## 7                             mydata.ref=cg02162897
## 8                          pval.threshold=4.720623e-06
## 9                               disp.association=FALSE
## 10                                disp.region=FALSE
## 11                                   start=38066017
## 12                                     end=38108036
## 13                      mydata.large.format=region_asso
## 14                          disp.association.large=TRUE
## 15                               disp.region.large=TRUE
## 16                 sample.labels.large=Gene expression
## 17                               color.list.large=green
## 18                        symbols.large=diamond-fill
## 19                              cormatrix.format=raw
## 20                            disp.cormatrixmap=TRUE
## 21                          cormatrix.method=spearman
## 22                 cormatrix.color.scheme=bluewhitered
## 23                           cormatrix.conf.level=0.05
## 24                               cormatrix.sig.level=1
## 25                               cormatrix.adjust=none
## 26                                 disp.phys.dist=TRUE
## 27                                 disp.color.bar=TRUE
## 28                                    disp.legend=TRUE
## 29          list.tracks=geneENSEMBL,ChromHMM,RegENSEMBL
## 30                               disp.mult.lab.X=FALSE
## 31                                      image.type=pdf
```

```
## 32 image.title="Example a-DMR in CYP1B1 in Adipose tissue"
## 33                   image.name=cyp1b1_zoom_plus_name_expr
## 34                                          image.size=3.5
## 35                                             genome=hg38
## 36                       dataset.geneE=hsapiens_gene_ensembl
```

5 Creating a plot like the webservice: comet.web

5.1 *coMET* plot: usage and plot like in the webservice

The user can create a *coMET* plot 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_Grch38.txt")
myexpressfile <- file.path(extdata, "cyp1b1_infofile_exprGene_region_Grch38.txt")
mycorrelation <- file.path(extdata, "cyp1b1_res37_rawMatrix.txt")
configfile <- file.path(extdata, "config_cyp1b1_zoom_4webserver_Grch38.txt")
comet.web(config.file=configfile, mydata.file=myinfofile,
          cormatrix.file=mycorrelation ,mydata.large.file=myexpressfile,
          print.image=FALSE,verbose=FALSE)
```

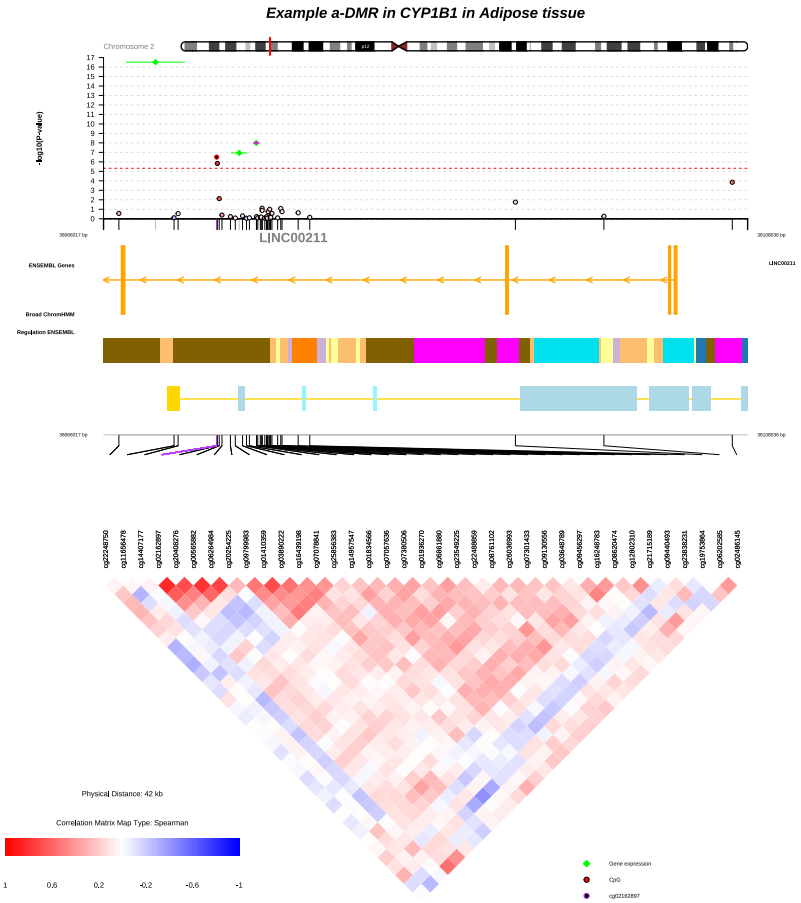

Figure 1: Plot with `comet.web` function.

5.2 Hidden values of `comet.web` function

Hidden values of `comet.web` function are shown in the section. If these values do not correspond to what you want to visualise, you need to use the function `comet`, as a more generic option.

Option	Value
mydata.type	FILE
mydata.large.type	LISTFILE
cormatrix.type	LISTFILE
disp.cormatrixmap	TRUE
disp.pvalueplot	TRUE
disp.mydata.names	TRUE
disp.connecting.lines	TRUE
disp.mydata	TRUE
disp.type	symbol
biofeat.user.type.plot	histogram
tracks.gviz	NULL
tracks.ggbio	NULL
tracks.trackviewer	NULL
biofeat.user.file	NULL
palette.file	NULL
disp.color.bar	TRUE
disp.phys.dist	TRUE
disp.legend	TRUE
disp.marker.lines	TRUE
disp.mult.lab.X	FALSE
connecting.lines.factor	1.5
connecting.lines.adj	0.01
connecting.lines.vert.adj	-1
connecting.lines.flex	0
color.list	red
font.factor	NULL
dataset.gene	hsapiens_gene_ensembl
DATASET.SNP	hsapiens_snp
VERSION.DBSNP	snp142Common
DATASET.SNP.STOMA	hsapiens_snp_som
DATASET.REGULATION	hsapiens_feature_set
DATASET.STRU	hsapiens_structvar
DATASET.STRU.STOMA	hsapiens_structvar_som
BROWSER.SESSION	UCSC

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 genomic elements, provides the most informative and easy approach to visualize graphics.

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

6.1.1 Input from data files

In this figure 2, we create different tracks outside to *coMET* with *Gviz*. The list of annotation tracks and different files are given to the function `coMET`.

```
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"
strand <- "*"

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

genetrack <-genes_ENSEMBL(gen,chrom,start,end,showId=TRUE)
snptrack <- snpBiomart_ENSEMBL(gen,chrom, start, end,
                               dataset="hsapiens_snp_som",showId=FALSE)
cpgIstrack <- cpgIslands_UCSC(gen,chrom,start,end)

prombedFilePath <- file.path(extdata, "/RoadMap/regions_prom_E063.bed")
promRMtrackE063<- DNaseI_RoadMap(gen,chrom,start, end, prombedFilePath,
                                 featureDisplay='promotor', type_stacking="squish")

bedFilePath <- file.path(extdata, "RoadMap/E063_15_coreMarks_mnemonics.bed")
chromHMM_RoadMapAllE063 <- chromHMM_RoadMap(gen,chrom,start, end,
                                            bedFilePath, featureDisplay = "all",
                                            colorcase='roadmap15' )
```

```
listgviz <- list(genetrack,snptrack,cpgIstrack,promRMtrackE063,chromHMM_RoadMapAllE063)

comet(config.file=configfile, mydata.file=myinfofile, mydata.type="file",
      cormatrix.file=mycorrelation,  cormatrix.type="listfile",
      mydata.large.file=myexpressfile, mydata.large.type="listfile",
      tracks.gviz=listgviz, verbose=FALSE, print.image=FALSE)
```

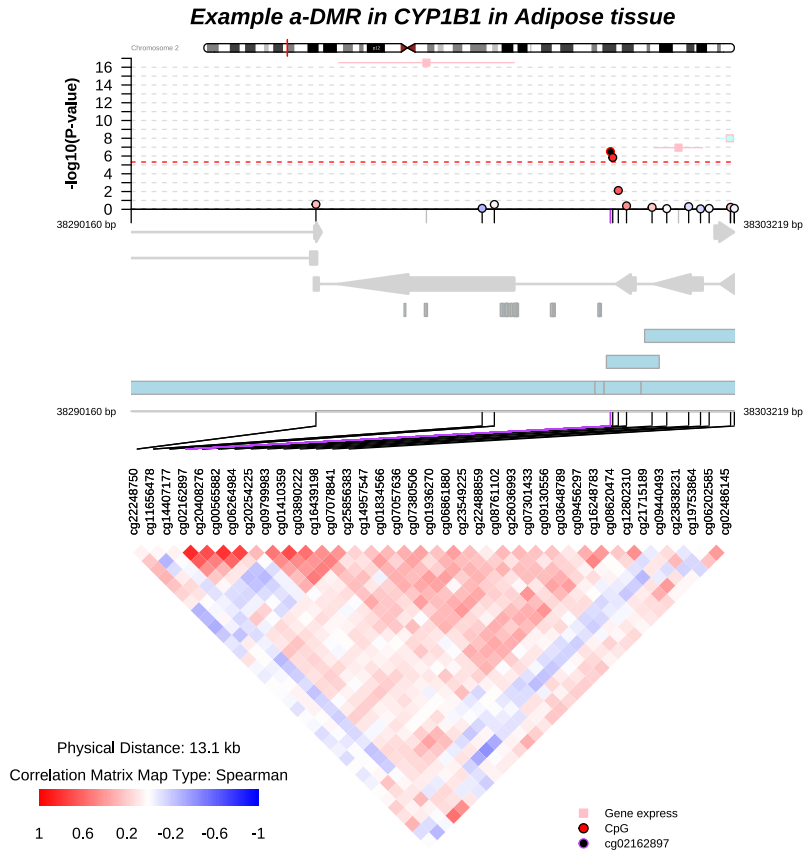

Figure 2: Plot with `comet` function from files.

6.1.2 *coMET* plot using input from a data frame

In this figure 3, we visualize the same data as in figure 2, but the data is in data frame format and not read in from an input file.

In addition, if the user would like to visualise only the correlations between CpG sites with P-value less than or equal to 0.05 in the upper plot, this option can be included. The correlations with a P-value greater than 0.05 can have the colour "goshwhite" whereas the other correlations will be displayed using a colour related to the correlation level. Conversely, in the P-value plot (upper plot), the points of each omic feature have their colours related to their correlations with the reference omic feature without taking into account the P-value associated with the correlation matrix.

Eventually, we increase the size of font using the option `fontsize.gviz`

```
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"
strand <- "*"

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

genetrack <-genes_ENSEMBL(gen,chrom,start,end,showId=TRUE)
snptrack <- snpBiomart_ENSEMBL(gen,chrom, start, end,
                               dataset="hsapiens_snp_som",showId=FALSE)

#Data no more available in UCSC (from September 2015)
iscatrack <-ISCA_UCSC(gen,chrom,start,end,mySession, table="iscaPathogenic")

listgviz <- list(genetrack,snptrack,iscatrack)

matrix.dnamethylation <- read.delim(myinfofile, header=TRUE, sep="\t", as.is=TRUE,
                                    blank.lines.skip = TRUE, fill=TRUE)
matrix.expression <- read.delim(myexpressfile, header=TRUE, sep="\t", as.is=TRUE,
                                blank.lines.skip = TRUE, fill=TRUE)
cormatrix.data.raw <- read.delim(mycorrelation, sep="\t", header=TRUE, as.is=TRUE,
                                 blank.lines.skip = TRUE, fill=TRUE)
```

```
listmatrix.expression <- list(matrix.expression)
listcormatrix.data.raw <- list(cormatrix.data.raw)
comet(config.file=configfile, mydata.file=matrix.dnamethylation,
      mydata.type="dataframe",cormatrix.file=listcormatrix.data.raw,
      cormatrix.type="listdataframe",cormatrix.sig.level=0.05,
      cormatrix.conf.level=0.05, cormatrix.adjust="BH",
      mydata.large.file=listmatrix.expression, mydata.large.type="listdataframe",
      fontsize.gviz =12,
      tracks.gviz=listgviz,verbose=FALSE, print.image=FALSE)
```

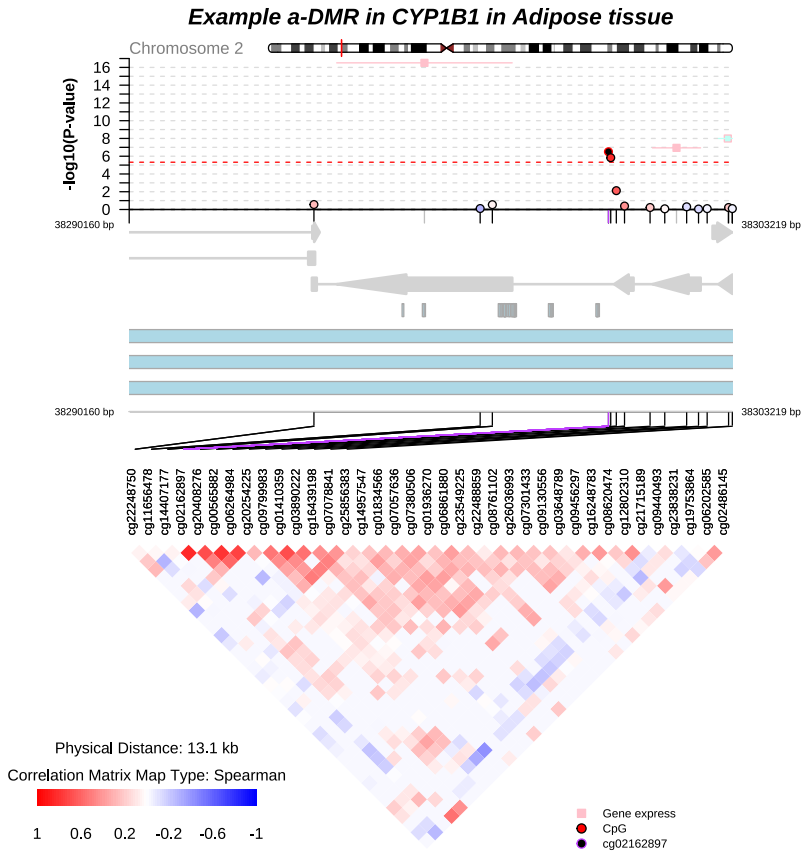

Figure 3: Plot with `comet` function from matrix data and with a pvalue threshold for the correlation between omics features (here CpG sites).

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 4.

```
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")

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

genetrack <-genes_ENSEMBL(gen,chrom,start,end,showId=FALSE)
snptrack <- snpBiomart_ENSEMBL(gen, chrom, start, end,
                               dataset="hsapiens_snp_som",showId=FALSE)
strutrack <- structureBiomart_ENSEMBL(chrom, start, end,
                                      strand, dataset="hsapiens_structvar_som")
clinVariant<-ClinVarMain_UCSC(gen,chrom,start,end)
clinCNV<-ClinVarCnv_UCSC(gen,chrom,start,end)
gwastrack <-GWAScatalog_UCSC(gen,chrom,start,end)
geneRtrack <-GeneReviews_UCSC(gen,chrom,start,end)

listgviz <- list(genetrack,snptrack,strutrack,clinVariant,
                 clinCNV,gwastrack,geneRtrack)
comet(config.file=configfile, mydata.file=myinfofile, mydata.type="file",
      cormatrix.file=mycorrelation, cormatrix.type="listfile",
      fontsize.gviz =12,
      tracks.gviz=listgviz, verbose=FALSE, print.image=FALSE,
      disp.pvalueplot=FALSE)
```

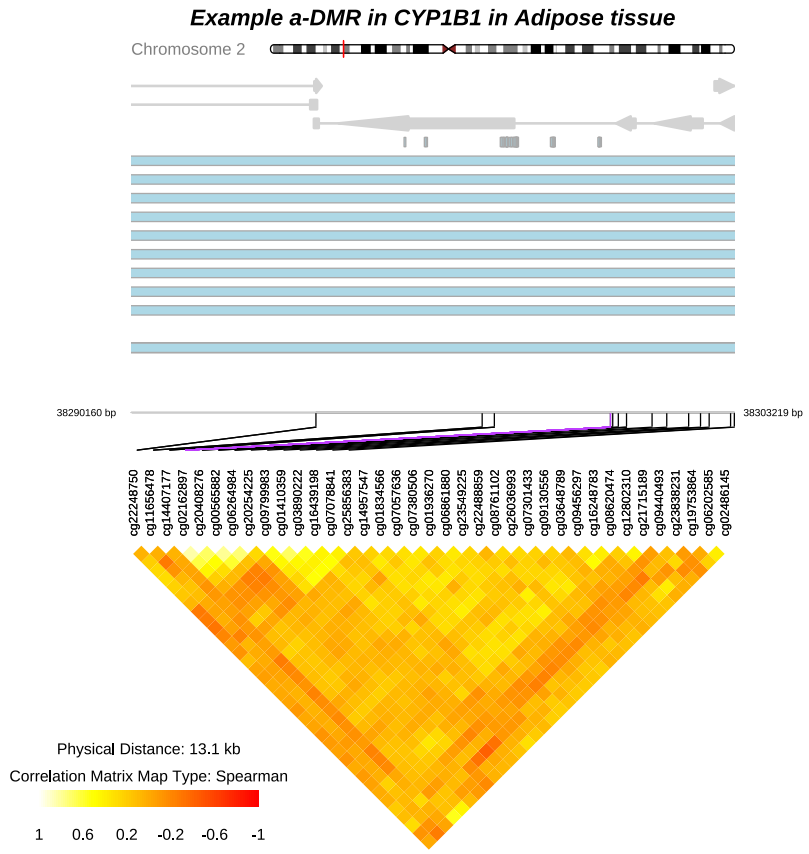

Figure 4: Plot with `comet` function without pvalue plot.

6.3 *coMET* plot: Manhattan plot and anonation track

It is possible to visualise only The Manhattan plot and the annotation tracks. In this case, we need to use the option `disp.cormatrixmap = FALSE`, for example see Figure 5.

```
extdata <- system.file("extdata", package="coMET",mustWork=TRUE)
configfile <- file.path(extdata, "config_cyp1b1_zoom_4nomatrix.txt")
myinfofile <- file.path(extdata, "cyp1b1_infofile.txt")
mycorrelation <- file.path(extdata, "cyp1b1_res37_rawMatrix.txt")

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

genetrack <-genes_ENSEMBL(gen,chrom,start,end,showId=FALSE)
snptrack <- snpBiomart_ENSEMBL(gen, chrom, start, end,
                               dataset="hsapiens_snp_som",showId=FALSE)
strutrack <- structureBiomart_ENSEMBL(chrom, start, end,
                                      strand, dataset="hsapiens_structvar_som")
clinVariant<-ClinVarMain_UCSC(gen,chrom,start,end)
clinCNV<-ClinVarCnv_UCSC(gen,chrom,start,end)
gwastrack <-GWAScatalog_UCSC(gen,chrom,start,end)
geneRtrack <-GeneReviews_UCSC(gen,chrom,start,end)

listgviz <- list(genetrack,snptrack,strutrack,clinVariant,
                 clinCNV,gwastrack,geneRtrack)
comet(config.file=configfile, mydata.file=myinfofile, mydata.type="file",
      cormatrix.file=mycorrelation, cormatrix.type="listfile",
      fontsize.gviz =12, font.factor=3,
      tracks.gviz=listgviz, verbose=FALSE, print.image=FALSE)
```

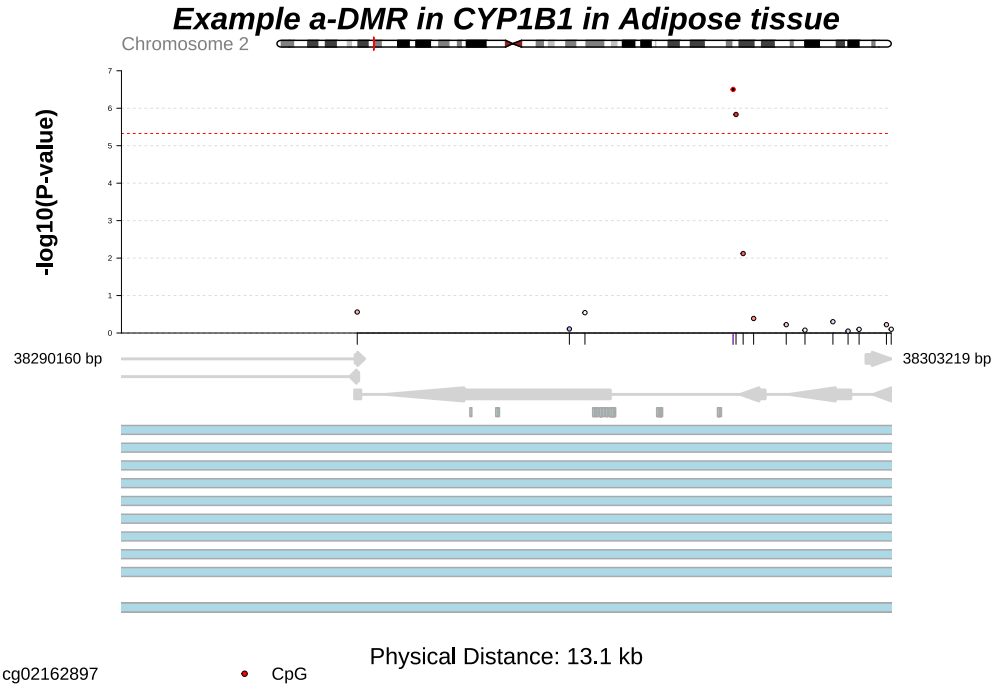

Figure 5: Plot with `comet` function without the correlation matrix.

7 Extract the significant correlations between omic features

coMET can help to visualise the correlations between omic features with EWAS results and other omic data. In addition, a function `comet.list` can extract the significant correlations according the method (options: `cormatrix.method`) and significance level (option: `cormatrix.sig.level`).

The output file has 7 columns:

1. the name of the first omic feature
2. the name of the second omic feature
3. the correlation between the omic features
4. the alpha/2 lower value (e.g. 0.05 (option `cormatrix.conf.level`))
5. the alpha/2 upper value (e.g. 0.05 (option `cormatrix.conf.level`))
6. the pvalue
7. the pvalue adjusted with the method selected (e.g. Benjamin and Hochberg) (option `cormatrix.adjust`)

```
extdata <- system.file("extdata", package="coMET",mustWork=TRUE)
mycorrelation <- file.path(extdata, "cyp1b1_res37_rawMatrix.txt")
myoutput <- file.path(extdata, "cyp1b1_res37_cormatrix_list_BH05.txt")


comet.list(cormatrix.file=mycorrelation,cormatrix.method = "spearman",
           cormatrix.format= "raw", cormatrix.conf.level=0.05,
           cormatrix.sig.level= 0.05, cormatrix.adjust="BH",
           cormatrix.type = "listfile", cormatrix.output=myoutput,
           verbose=FALSE)

listcorr <- read.csv(myoutput, header = TRUE,
                     sep = "\t", quote = "")
dim(listcorr)
```

```
## [1] 336   7
```

```
head(listcorr)
```

```
##   omicFeature1 omicFeature2 correlation     lowerCI    upperCI       pvalue
## 1   cg22248750   cg14407177   0.2153743  0.11294792  0.3132713 4.975020e-05
## 2   cg22248750   cg02162897   0.2761912  0.17632357  0.3704308 1.575519e-07
## 3   cg22248750   cg20408276   0.2807258  0.18108231  0.3746643 9.649818e-08
## 4   cg22248750   cg00565882   0.2345897  0.13288218  0.3314082 9.478992e-06
## 5   cg22248750   cg06264984   0.1793832  0.07583111  0.2791072 7.613440e-04
```

```
## 6   cg22248750   cg09799983  -0.2979454 -0.39070492 -0.1991959 1.382644e-08
##   pvalue.adjusted
## 1    2.029592e-04
## 2    1.178984e-06
## 3    7.472999e-07
## 4    4.477311e-05
## 5    2.414548e-03
## 6    1.261426e-07
```

8 Annotation tracks

Annotation tracks can be created with *Gviz* using four different functions:

1. `UcscTrack`. Different UCSC tracks can be selected for visualisation from the table Browser of UCSC http://genome-euro.ucsc.edu/cgi-bin/hgTables?hgsid=202842745_Dlvit14QO0G6ZPpLoEVABG8aqfrm&clade=mammal&org=Human&db=hg19&hgta_group=varRep&hgta_track=cpgIslandExt&hgta_table=0&hgta_regionType=genome&position=chr6%3A32726553\protect\discretionary{\char\hyphenchar\font}{}{}32727053&hgta_outputType=primaryTable&hgta_outFileName=
2. `BiomartGeneRegionTrack`. A connection should be established to the Biomart database to visualise the genetic elements.
3. `DataTrack`. This allows the visualisation of numerical data.
4. `AnnotationTrack`. This allows the visualisation of any annotation data.

For more information consult the user guide for *Gviz*.

8.1 Ensembl

The Ensembl project [2] produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online http://www.ensembl.org/index.html. A set of wrap R functions were created to extract data from Ensembl BioMart for human genome using Ensembl REST [3], but they can be extended to other genomes. You can ask help to `tiphaine.martin@mssm.edu`.

This is the list of R functions created in *coMET* to visualise ENSEMBL data. Below described the colors of tracks and specific characteristics of some annotation tracks.

- `bindingMotifsBiomart_ENSEBML` : Visualise the binding motifs in the genomic region of interest
- `genes_ENSEBML` : Visualise the genes from ENCODE in the genomic region of interest
- `genesName_ENSEBML` : Visualise the name of genes from ENCODE in the genomic region of interest
- `interestGenes_ENSEBML` : Visualise the genes from ENCODE in the genomic region of interest with a specific color for genes of interest
- `interestTranscript_ENSEBML` : Visualise the transcripts from ENCODE in the genomic region of interest with a specific color for exons of interest
- `miRNATargetRegionsBiomart_ENSEBML` : Visualise the miRNA target regions in the genomic region of interest
- `otherRegulatoryRegionsBiomart_ENSEBML` : Visualise the other regulatory regions in the genomic region of interest

- `regulationBiomart_ENSEBML` (obselet function): Visualise the other regulatory regions in the genomic region of interest
- `regulatoryEvidenceBiomart_ENSEBML` : Visualise the regulatory evidence regions in the genomic region of interest
- `regulatoryFeaturesBiomart_ENSEBML` : Visualise the regulatory features regions in the genomic region of interest
- `regulatorySegmentsBiomart_ENSEBML` : Visualise the regulatory segment regions in the genomic region of interest. *Warning: no more available*
- `snpBiomart_ENSEBML` : Visualise the SNPs in the genomic region of interest
- `structureBiomart_ENSEBML` : Visualise the structural variations in the genomic region of interest
- `transcript_ENSEBML` : Visualise the transcripts in the genomic region of interest

Below described the colors of tracks and specific characteristics of some annotation tracks.

8.1.1 Genes and transcripts from Ensembl

The color of the genetic elements is defined by the R package *Gviz*.

It is possible to chagne the colour of some exsons by using the function `interestGenesENSEMBL` or `interestTranscriptENSEMBL`. The elements and the colours to be displayed must be given as list. An example is given below:

```
gen <- "hg38"
chr <- "chr15"
start <- 75011669
end <- 75019876
interestfeatures <- rbind(c("75011883","75013394","bad"),
                          c("75013932","75014410","good"))
interestcolor <- list("bad"="red", "good"="green")

interestgenesENSMBLtrack<-interestGenes_ENSEMBL(gen,chr,start,end,interestfeatures,
                                                interestcolor,showId=TRUE)
plotTracks(interestgenesENSMBLtrack, from=start, to=end)
```

8.1.2 Regulatory elements from Ensembl

This function is now obselet in *coMET* as Ensembl have restructured their databases due to the new version of the genome GRCh38. The same data is now available by using the function `RegulatoryFeaturesBiomart`.

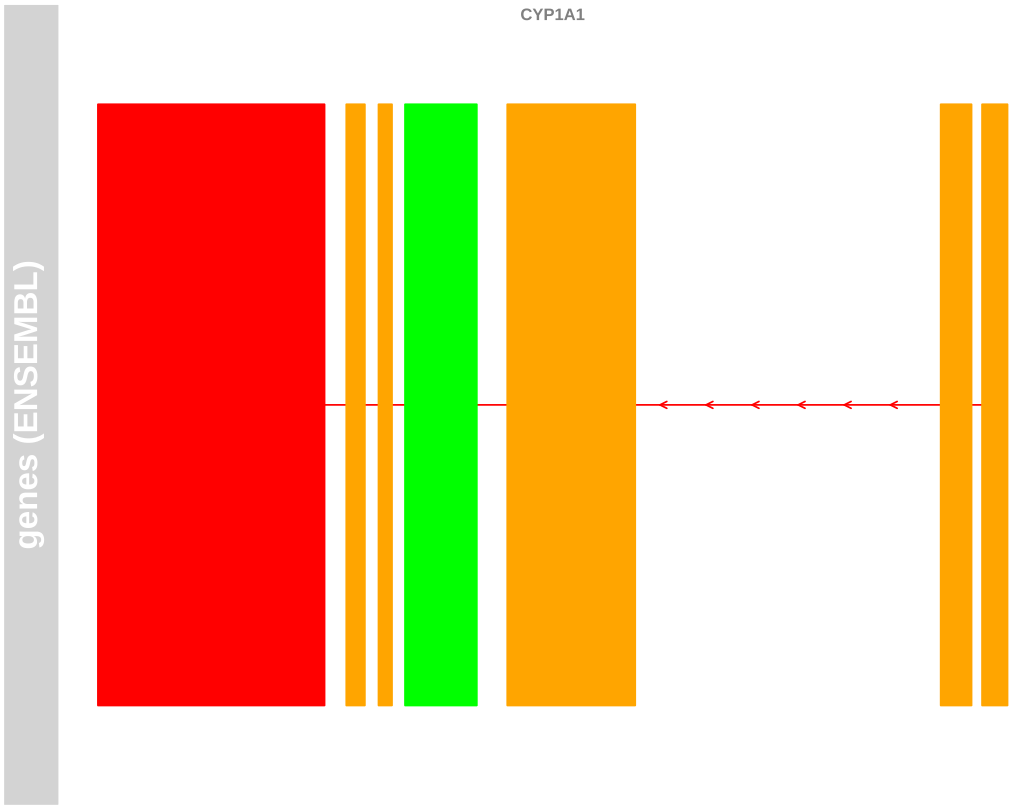

Figure 6: Plot genes with different colors according user's choice.

The colors were :

Omic Feature	Colour Name	Colour	Hex Code
Promoter Associated	darkolivegreen		#556B2F
CTCF Binding Site	cadetblue1		#98F5FF
Gene Associated	coral		#FF7256
Non-Gene Associated	darkgoldenrod1		#FFB90F
Predicted Transcribed Region	greenyellow		#ADFF2F
PolIII Transcription Associated	purple		#A020F0
Enhancer	gold		#FFD700
Transcription Factor Binding Site	darkorchid1		#BF3EFF
Predicted Weak enhancer/Cis-reg element	yellow		#FFFF00
Heterochromatin	wheat4		#8B7E66
Open Chromatin	snow3		#CDC9C9
Promoter Flank	tomato		#FF6347
Repressed/Low Activity	snow4		#8B8989
Unclassified	aquamarine		#7FFFD4

8.1.3 structureBiomart from Ensembl

Listed below are the colours for somatic structural variation and structural variation.

Omic Feature	Colour Name	Colour	Hex Code
copy_number_variation	royalblue1		#4876FF
inversion	green		#00FF00
translocation	cyan		#00FFFF
sequence_alteration	yellow		#FFFF00
snp	red		#FF0000
insertion	blueviolet		#8A2BE2
deletion	orange		#FFA500
indel	darkgoldenrod1		#FF3E96
substitution	green4		#008B00

8.1.4 miRNA Target Regions from Ensembl

The colour of the miRNA target regions is set to Plum4 (hex code: #8B668B)

8.1.5 Binding Motif Biomart from Ensembl

Listed on the next page are the colours used for the different types of binding motifs. The frequency shown is that found in GRCh38 (hg38). Motifs with red text are found only in GRCh37 (hg19), motifs with blue text are found only in GRCh38 (hg38)

Regulation Element	Frequency	Colour	Hex code
BHLHE40	147		#157BAB
Cfos	2824		#8E6363
Cjun	41483		#B2DF8A
Cmyc	2073		#FFF803
CTCF	62915		#1F78B4
CTCFL	1518		#FF5703
E2F1	**12668**		**#CCCC00**
E2F4	9425		#FFFF99
E2F6	3937		#FFD3D7
EBF1	3749		#FFCA08
EcR::usp	**176**		**#B51660**
Egr1	115261		#A6CEE3
ELF1	3130		#0BFCFF
ETS1	198		#57C716
FOSL1	2027		#8FB247
FOSL2	6004		#F05868
FOXA1	11409		#6A3D9A
FOXA2	1592		#03FFAC
Gabp	20823		#E31A1C
Gata1	204		#74C2D6
Gata2	719		#86941F
HNF4A	6477		#AFE6DC
HNF4G	820		#6F41F0
IRF4	762		#1F9433
JUN::FOS	**18932**		**#FDBF6F**
Junb	650		#8A0A66
Jund	15234		#FF7F00
Max	4160		#5BBAAE
MEF2A	3267		#3088F0
MEF2C	140		#5C728C
MYC::MAX	**5116**		**#8B2323**
NFKB	853		#C4FF00
Nr1h3::Rxra	**177**		**#2E16B5**
Nrf1	3372		#A0A0A0
Nrsf	6136		#3B3E19
Pax5	870		#36BB98
Pbx3	38		#E31D42
POU2F2	269		#8A520A
PPARG::RXRA	**654**		**#5E1A52**
PU1	27566		#FB9A99
RXR::RAR_DR5	**380**		**#0A328A**
RXRA	**821**		**#18044F**
RXRA::VDR	**213**		**#94DB48**
SP1	8115		#B15928
SP2	1126		#BB9C36
SRebp1	**21**		**#A33333**
SRebp2	**11**		**#B2B2F0**
Srf	7322		#383838
Tal1::Gata1	**332**		**#4A8A0A**
Tcf12	2000		#FC03FF
THAP1	168		#9BB516
Tr4	673		#1A5E45
USF1	31695		#33A02C
Yy1	7695		#62725B
ZBTB33	2383		#2F6568
ZEB1	898		#A036BB
Znf263	12457		#CAB2D6

8.1.6 Other Regulatory Regions Biomart from Ensembl

Listed below are the colours used for the different types of regulatory regions. The frequency shown is that found in GRCh38 (hg38).

Regions	Frequency	Colour	Hex Code
Enhancer	42926		#E41A1C
Transcription Start Site	378153		#4DAF4A

8.1.7 Regulatory Features Biomart from Ensembl

Listed below are the colours used for the different types of regulatory features The frequency shown is that found in GRCh38 (hg38).

Regulatory Features	Frequency	Colour	Hex Code
Enhancer	26702		#E6AB02
CTCF Binding Site	24935		#66A61E
Flanking Region	17976		#E7298A
Open chromatin Promoter	15213		#7570B3
Promoter	3351		#1B9E77
TF binding site	6031		#D95F02

8.1.8 Other Regulatory Segments Biomart from Ensembl

*Warning: (*No more available) Listed below are the colours used for the different types of regulatory segments. The frequency shown is that found in GRCh38 (hg38). Segments with red text are found only in GRCh37 (hg19)

Regulatory Segments	Frequency	Colour	Hex Code
CTCF enriched	198589		#1F78B4
Predicted Enhancer	1995055		#FB9A99
Predicted heterochomatin	8188068		#CAB2D6
Predicted low activity	3469054		#FDBF6F
Predicted Poised	**218260**		**#B2DF8A**
Predicted Promoter Flank	839976		#33A02C
Predicted Promoter with TSS	123878		#A6CEE3
Predicted Repressed	3507490		#FF7F00
Predicted Transcribed Region	3061736		#E31A1C

8.1.9 Regulatory Evidence Elements Biomart from Ensembl

Listed on the next 3 pages are the colours used for the different types of regulatory evidence elements. The frequency shown is that found in GRCh37 (hg19). At the current time this track has not been optimised for GRCh38 (hg38) meaning any elements found exclusively in GRCh38 do not have an assigned colour and will be displayed in the default track colour of *Gviz*.

Regulation Element	Frequency	Colour	Hex Code
Ap2alpha	5826		#D5A4DE
Ap2gamma	10006		#AC8B41
ATF3	3276		#AC5273
BAF155	548		#39394A
BAF170	368		#E5DE00
BATF	9021		#EE9C39
BCL11A	4126		#817E7A
BCL3	1402		#319C73
BCLAF1	1803		#BF7EFF
BHLHE40	249		#60A8BC
Brg1	2337		#C80096
Cfos	7398		#623918
Cjun	30052		#EEAC9C
Cmyc	32020		#FF93F0
CTCF	994692		#0432FF
CTCFL	1700		#6FE9FF
DNase1	963201		#FFFF00
E2F1	8387		#EEFF94
E2F4	5817		#AC3929
E2F6	13141		#186A88
EBF1	5337		#00FDFF
Egr1	22535		#83DEA4
ELF1	13642		#9C5A4A
ETS1	1293		#415A20
FOSL1	2216		#8B1608
FOSL2	3664		#DEC552
FOXA1	16278		#F6BD5A
FOXA2	1705		#FFC22C
Gabp	22129		#104131
Gata1	583		#FFCDBD
Gata2	4258		#A87E7F
GTF2B	550		#FFE699
H2AK5ac	888		#767582
H2AZ	98335		#CCEBC5
H2BK120ac	6655		#00FCC4
H2BK12ac	1351		#5DCF8B
H2BK15ac	380		#8B8BA4
H2BK20ac	2183		#EEB4CD
H2BK5ac	130		#CDB47B
H3K14ac	380		#98FFA2
H3K18ac	7345		#BDB46A
H3K23ac	256		#E3E186
H3K23me2	47		#41B4EE
H3K27ac	314361		#08ACD5

Regulation Element	Frequency	Colour	Hex Code
H3K27me3	3248516		#8EFF0E
H3K36me3	3524819		#FF0000
H3K4ac	6221		#C59700
H3K4me1	155338		#83B420
H3K4me2	428636		#FF603D
H3K4me3	495468		#9D49C7
H3K56ac	1409		#BD4A73
H3K79me1	376		#525262
H3K79me2	78701		#D54152
H3K9ac	276750		#52834A
H3K9me1	188		#D5DEF6
H3K9me3	10975		#552431
H4K20me1	30775		#085A73
H4K5ac	6886		#56CA4B
H4K8ac	4109		#7B6220
H4K91ac	3528		#83ACEE
HDAC2	4595		#6A6A52
HDAC8	258		#BD3162
HEY1	6626		#C55A6A
HNF4A	8524		#526229
HNF4G	814		#94ADFF
Ini1	3623		#A4CDE6
IRF4	3198		#41ACA4
Junb	360		#17B103
Jund	18964		#9F847F
Max	46624		#20206A
MEF2A	4964		#FF8158
MEF2C	202		#184173
Nanog	1366		#9C8B31
Nfe2	861		#4C31AF
NFKB	1344		#BDE673
NR4A1	256		#EE6241
Nrf1	769		#E66294
Nrsf	24417		#8BD5EE
p300	5423		#393997
Pax5	2391		#C54129
Pbx3	1093		#000000
PolII	216695		#0070C0
PolIII	215		#737B7B
POU2F2	861		#FF40FF
POU5F1	1446		#A83D4C
PU1	14641		#67B339
Rad21	173916		#7B9CB4
RXRA	1433		#628BBD

Regulation Element	Frequency	Colour	Hex Code
SETDB1	1823		#83C944
Sin3Ak20	6011		#A55593
SIX5	3823		#D0DAD3
SP1	9356		#5900FF
SP2	729		#C8C8FA
Srf	6065		#8B52AC
TAF1	39514		#C5C5C5
TAF7	6895		#4C4C4C
Tcf12	6110		#62317B
THAP1	627		#B35F41
Tr4	4124		#4A6A83
USF1	29377		#986664
XRCC4	242		#5A4120
Yy1	36710		#7B2941
ZBTB33	2522		#006A62
ZBTB7A	1407		#804000
ZEB1	532		#624A31
Znf263	2383		#DDFF00
ZNF274	338		#00FF00

8.2 UCSC

the UCSC Genome Browser [4] website http://genome-euro.ucsc.edu/ contains the reference sequence and working draft assemblies for a large collection of genomes.

This is the list of *R* wrapping functions of some tracks found in UCSC genome browser. Below described the colors of tracks and specific characteristics of some annotation tracks.

- `chromatinHMMAll_UCSC` : Visualise the chromHMM Broad found in UCSC genome browser of all tissues in the genomic region of interest.
- `chromatinHMMOne_UCSC` : Visualise the chromHMM Broad found in UCSC genome browser of the tissue of interest in the genomic region of interest.
- `ClinVarCnv_UCSC` : Visualise clinical CNVs found in ClinVar tracks of UCSC genome browser in the genomic region of interest.
- `ClinVarMain_UCSC` : Visualise clinical SNPs found in ClinVar tracks of UCSC genome browser in the genomic region of interest.
- `CoreillCNV_UCSC` : Visualise CNV found in Coreil tracks of UCSC genome browser in the genomic region of interest.
- `COSMIC_UCSC` : Visualise SNPs found in COSMIC tracks of UCSC genome browser in the genomic region of interest.*Warning: We could not more access to COSMIC data from UCSC genome browser, people needs to extract data from COSMIC directly*.
- `cpgIslands_UCSC` : Visualise CpG Island found in CpGIsland tracks of UCSC genome browser in the genomic region of interest.
- `DNAse_UCSC` (obselete): Visualise clinical CNV found in ClinVar tracks of UCSC genome browser in the genomic region of interest.
- `GAD_UCSC` : Visualise genes found in GAD tracks of UCSC genome browser in the genomic region of interest.
- `gcContent_UCSC` : Visualise GC content found in UCSC genome browser in the genomic region of interest.
- `GeneReviews_UCSC` : Visualise clinical genes found in GeneReviews tracks of UCSC genome browser in the genomic region of interest.
- `GWAScatalog_UCSC` : Visualise SNPS found in GWAS catalog tracks of UCSC genome browser in the genomic region of interest.
- `HistoneAll_UCSC` : Visualise histone patterns found in UCSC genome browser of all tissues in the genomic region of interest.
- `HistoneOne_UCSC` : Visualise histone patterns found in UCSC genome browser of one tissue of interest in the genomic region of interest.

- `ISCA_UCSC` (obselete) : Visualise clinical CNV found in UCSC genome browser in the genomic region of interest.
- `knownGenes_UCSC` : Visualise known genes found in UCSC genome browser in the genomic region of interest.
- `refGenes_UCSC` : Visualise reference genes found in UCSC genome browser in the genomic region of interest.
- `repeatMasker_UCSC` : Visualise repeat elements found in UCSC genome browser in the genomic region of interest.
- `segmentalDups_UCSC` : Visualise segmental duplcations found in UCSC genome browser in the genomic region of interest.
- `snpLocations_UCSC` : Visualise SNPs found in UCSC genome browser in the genomic region of interest.
- `xenorefGenes_UCSC` : Visualise xeno reference genes found in UCSC genome browser in the genomic region of interest.

8.2.1 ChromHMM from UCSC

For this function there are two possible colour schemes to choose from. The selection between schemes is made with the variable `colour`. The default scheme is *coMET*, the colours chosen have been selected so that different elements can be easily distinguished. The second scheme is *UCSC*, these are the set colours used by UCSC, in certain plots it may be difficult to distinguish elements apart. These UCSC colours are correct at the time this document was writtern however if these change in the future and this is not reflected here please contact us.

the colours used in both schemes are listed below:

coMET Colour Scheme		
Omic Feature	**Colour**	**Hex Code**
1_Active_Promoter		#E31A1C
2_Weak_Promoter		#FB9A99
3_Poised_Promoter		#6A3D9A
4_Strong_Enhancer		#FF7F00
5_Strong_Enhancer		#CAB2D6
6_Weak_Enhancer		#FFFF99
7_Weak_Enhancer		#FDBF6F
8_Insulator		#1F78B4
9_Txn_Transition		#B2DF8A
10_Txn_Elongation		#33A02C
11_Weak_Txn		#00E1EF
12_Repressed		#FF00FF
13_Heterochrom/lo		#806000
14_Repetitive/CNV		#808080
15_Repetitive/CNV		#BFBFBF

Omic Feature	Colour Name	Colour	Hex Code
1_Active_Promoter	firebrick1		#FF3030
2_Weak_Promoter	darksalmon		#FF967A
3_Poised_Promoter	blueviolet		#8A2BE2
4_Strong_Enhancer	orange		#FFA500
5_Strong_Enhancer	coral		#FF7F50
6_Weak_Enhancer	yellow		#FFFF00
7_Weak_Enhancer	gold		#FFD700
8_Insulator	cornflowerblue		#6495ED
9_Txn_Transition	darkolivegreen		#556B2F
10_Txn_Elongation	forestgreen		#228B22
11_Weak_Txn	darkseagreen1		#8BFFC1
12_Repressed	gainsboro		#DCDCDC
13_Heterochrom/lo	gray74		#BDBDBD
14_Repetitive/CNV	gray77		#C4C4C4
15_Repetitive/CNV	gray86		#DBDBDB

8.2.2 ISCA track (obselete database)

International Standards of Cytogenomic Arrays Consortium defined a set of phenotypes for CNVs. Different colours are defined to represent them. This database is not more accessible from UCSC (from September 2015).

Omic Feature	Colour Name	Colour	Hex Code
iscaPathogeni	purple		#9B30FF
iscaPathGainCum	red		#FF0000
iscaPathLossCum	blue		#0000FF
iscaCuratedPathogeni	purple		#9B30FF
iscaLikelyPathogeni	orchid1		#DF78E4
iscaUncertain	lightgrey		#D3D3D3
iscaBenign	black		#000000
iscaCuratedBenign	black		#000000
iscaLikelyBenign	black		#000000

8.2.3 Other potential data from UCSC

You can access to other data via UCSC track hub [5] :

- Other tracks and table accessible to UCSC genome browser https://genome.ucsc.edu/cgi-bin/hgTables?hgsid=444062899_lxuSrw4J9exVt1OafMuY4LDbVs1F&clade=mammal&org=Human&db=hg19&hgta_group=allTracks&hgta_track=knownGene&hgta_table=0&hgta_regionType=genome&position=chr21%3A33031597-33041570&hgta_outputType=primaryTable&hgta_outFileName=
- Track HUB of UCSC genome browser https://genome-euro.ucsc.edu/cgi-bin/hgHubConnect?hubUrl=http%3A%2F%2Ffantom.gsc.riken.jp%2F5%2Fdatahub%2Fhub.txt&hgHubConnect.remakeTrackHub=on&redirect=manual&source=genome.ucsc.edu

and use `DataTrack` or `AnnotationTrack` or `UCSCTrack` of *Gviz* to visuaslise them.

8.3 NIH Roadmap epigenomics project

NIH Roadmap epigenomics projects http://www.roadmapepigenomics.org/ [6] aims to produce a public resource of human epigenomic data to catalyze basic biology and disease-oriented research. The project has generated high-quality, genome-wide maps of several key histone modifications, chromatin accessibility, DNA methylation and mRNA expression across 100s of human cell types and tissues (111 consolidated epigenomes from the NIH Roadmap Epigenomics Project and 16 epigenomes from The Encyclopedia of DNA Elements (ENCODE) project).

Release 9 of the compendium contains uniformly pre-processed and mapped data from multiple profiling experiments (technical and biological replicates from multiple individuals and/or datasets from multiple centers) spanning 183 biological samples and 127 consolidated epigenomes.

More information on each type data are on the site of NIH Roadmap Epigenomics Program http://egg2.wustl.edu/roadmap/web_portal/index.html and the meta-data on different tissues (more for correspondance between Epigenome ID (EID) and the standartized epigenome name), you need to look at this spreadsheet https://docs.google.com/spreadsheets/d/1yikGx4MsO9Ei36b64yOy9Vb6oPC5IBGlFbYEt-N6gOM/edit#gid=15

The current data are done on Release 9. The data are mapped on the reference genome **hg19**. Below described the colors of tracks and specific characteristics of some annotation tracks.

- `chromHMM_RoadMap` : Visualisation of chromatin states defined in NIH Roadmap project
- `dgfootprints_RoadMap`: Visualisation of DNA motif positional bias in digital genomic Footprinting Sites
- `DNaseI_RoadMap` : Visualisation of promoter/enhancer regions

8.3.1 Chromatin state

There are 3 chromatin states defined in NIH Roadmap project (15 states, 18 states and 25 states). For 18 and 25 states, there are the choice between 2 set of colors. First, the colors defined by NIH Roadmap and second, the colors defined by us for a better differentiation between states.

you can use `chromHMM_RoadMap` to visualise chromatin state in :

- 15-states, go to http://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final/ and select the MNEMONICS BED FILES, where bins with the same state label are merged and a label is assigned to the entire merged regions, related to your tissue of interest.

- 18-states, go to http://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/core_K27ac/jointModel/final/ and select the MNEMONICS BED FILES, where bins with the same state label are merged and a label is assigned to the entire merged regions, related to your tissue of interest .
- 25-states, go to http://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/imputed12marks/jointModel/final/ and select your tissue of interest.

You can have more information about these data from NIH Roadmap website http://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state.

You can visualise this bed using the function `chromHMM_RoadMap` and you can choice the color between *roadmap15*, *roadmap18*, *comet18*, *roadmap25* and *comet25*.

Below you can find the color code for each state depending if 15-,18- or 25-state

Listed below are the colours used for the different elements contained in NIH Roadmap data with 15 states.

State & Acronym	Description	Colour	Hex Code
1_TssA	Active TSS		#FF0000
2_TssAFlnk	Flanking Active TSS		#FF6E00
3_TxFlnk	Transcr. at gene 5' and 3'		#32CD32
4_Tx	Strong transcription		#008000
5_TxWk	Weak transcription		#006400
6_EnhG	Genic enhancers		#C2E105
7_Enh	Enhancers		#FFFF00
8_ZNF/Rpts	ZNF genes & repeats		#66CDAA
9_Het	Heterochromatin		#8A91D0
10_TssBiv	Bivalent/Poised TSS		#CD5C5C
11_BivFlnk	Flanking Bivalent TSS/Enh		#E9967A
12_EnhBiv	Bivalent Enhancer		#BDB76B
13_ReprPC	Repressed PolyComb		#3A3838
14_ReprPCWk	Weak Repressed PolyComb		#808080
15_Quies	Quiescent/Low		#DCDCDC

Listed below are the colours used for the different elements contained in NIH Roadmap data with 18 states with NIH Roadmap colors.

RoadMap Colour Scheme				
State & Acronym	**Description**	**Colour Name**	**Colour**	**Hex Code**
1_TssA	Active TSS	red		#FF0000
2_TssFlnk	Flanking TSS	orangered		#FF4500
3_TssFlnkU	Flanking TSS Upstream	orangered		#FF4500
4_TssFlnkD	Flanking TSS Downstream	orangeRed		#FF4500
5_Tx	Strong transcription	green		#008000
6_TxWk	Weak transcription	darkgreen		#006400
7_EnhG1	Genic enhancer1	greenyellow		#C2FF05
8_EnhG2	Genic enhancer2	greenyellow		#C2FF05
9_EnhA1	Active Enhancer 1	orange		#FFC34D
10_EnhA2	Active Enhancer 2	orange		#FFC34D
11_EnhWk	Weak Enhancer	yellow		#FFFF00
12_ZNF/Rpts	ZNF genes & repeats	mediumaquamarine		#66CDAA
13_Het	Heterochromatin	paleturquoise		#8A91D0
14_TssBiv	Bivalent/Poised TSS	indianred		#CD5C5C
15_EnhBiv	Bivalent Enhancer	darkkhaki		#BDB76B
16_ReprPC	Repressed PolyComb	silver		#808080
17_ReprPC	Weak Repressed PolyComb	gainsboro		#C0C0C0
18_Quies	Quiescent/Low	white		#FFFFFF

Listed below are the colours used for the different elements contained in NIH Roadmap data with 18 states with *coMET* colors.

coMET Colour Scheme			
State & Acronym	**Description**	**Colour**	**Hex Code**
1_TssA	Active TSS		#FF0000
2_TssFlnk	Flanking TSS		#FF6E00
3_TssFlnkU	Flanking TSS Upstream		#FF9300
4_TssFlnkD	Flanking TSS Downstream		#DA7B08
5_Tx	Strong transcription		#008000
6_TxWk	Weak transcription		#006400
7_EnhG1	Genic enhancer1		#C2FF05
8_EnhG2	Genic enhancer2		#C2FFBD
9_EnhA1	Active Enhancer 1		#FE00DB
10_EnhA2	Active Enhancer 2		#FFA7D6
11_EnhWk	Weak Enhancer		#FFFF00
12_ZNF/Rpts	ZNF genes & repeats		#66CDAA
13_Het	Heterochromatin		#8A91D0
14_TssBiv	Bivalent/Poised TSS		#CD5C5C
15_EnhBiv	Bivalent Enhancer		#BDB76B
16_ReprPC	Repressed PolyComb		#323232
17_ReprPC	Weak Repressed PolyComb		#AFAFAF
18_Quies	Quiescent/Low		#DCDCDC

Listed below are the colours used for the different elements contained in NIH Roadmap data with 25 states with NIH Roadmap colors.

RoadMap Colour Scheme				
State & Acronym	**Description**	**Colour Name**	**Colour**	**Hex Code**
1_TssA	Active TSS	Red		#FF0000
2_PromU	Promoter Upstream TSS	Orange Red		#FF4500
3_PromD1	Promoter Downstream TSS 1	Orange Red		#FF4500
4_PromD2	Promoter Downstream TSS 2	Orange Red		#FF4500
5_Tx5'	Transcribed - 5' preferential	Green		#008000
6_Tx	Strong transcription	Green		#008000
7_Tx3'	Transcribed - 3' preferential	Green		#008000
8_TxWk	Weak transcription	Lighter Green		#009600
9_TxReg	Transcribed & regulatory (Prom/Enh)	Electric Lime		#C2FF05
10_TxEnh5'	Transcribed 5' preferential and Enh	Electric Lime		#C2FF05
11_TxEnh3'	Transcribed 3' preferential and Enh	Electric Lime		#C2FF05
12_TxEnhW	Transcribed and Weak Enhancer	Electric Lime		#C2FF05
13_EnhA1	Active Enhancer 1	Orange		#FFC34D
14_EnhA2	Active Enhancer 2	Orange		#FFC34D
15_EnhAF	Active Enhancer Flank	Orange		#FFC34D
16_EnhW1	Weak Enhancer 1	Yellow		#FFFF00
17_EnhW2	Weak Enhancer 2	Yellow		#FFFF00
18_EnhAc	Primary H3K27ac possible Enhancer	Yellow		#FFFF00
19_DNase	Primary DNase	Lemon		#FFFF66
20_ZNF/Rpts	ZNF genes & repeats	Aquamarine		#66CDAA
21_Het	Heterochromatin	Light Purple		#8A91D0
22_PromP	Poised Promoter	Pink		#E6B8B7
23_PromBiv	Bivalent Promoter	Dark Purple		#7030A0
24_ReprPC	Repressed Polycomb	Gray		#808080
25_Quies	Quiescent/Low	White		#FFFFFF

Listed below are the colours used for the different elements contained in NIH Roadmap data with 25 states with *coMET* colors.

coMET Colour Scheme			
State & Acronym	**Description**	**Colour**	**Hex Code**
1_TssA	Active TSS		#FF0000
2_PromU	Promoter Upstream TSS		#FC6D00
3_PromD1	Promoter Downstream TSS 1		#DD8100
4_PromD2	Promoter Downstream TSS 2		#AD7622
5_Tx5'	Transcribed - 5' preferential		#008000
6_Tx	Strong transcription		#004D00
7_Tx3'	Transcribed - 3' preferential		#009462
8_TxWk	Weak transcription		#00FE00
9_TxReg	Transcribed & regulatory (Prom/Enh)		#00FFFF
10_TxEnh5'	Transcribed 5' preferential and Enh		#009FFF
11_TxEnh3'	Transcribed 3' preferential and Enh		#0028FF
12_TxEnhW	Transcribed and Weak Enhancer		#0000AE
13_EnhA1	Active Enhancer 1		#FF00FF
14_EnhA2	Active Enhancer 2		#FFB2FF
15_EnhAF	Active Enhancer Flank		#FFD8FF
16_EnhW1	Weak Enhancer 1		#FFFF00
17_EnhW2	Weak Enhancer 2		#E3FF8C
18_EnhAc	Primary H3K27ac possible Enhancer		#FFD500
19_DNase	Primary DNase		#FFFFC2
20_ZNF/Rpts	ZNF genes & repeats		#66CDAA
21_Het	Heterochromatin		#8A91D0
22_PromP	Poised Promoter		#E6B8B7
23_PromBiv	Bivalent Promoter		#7030A0
24_ReprPC	Repressed Polycomb		#646464
25_Quies	Quiescent/Low		#DCDCDC

8.3.2 DNA Motif Positional Bias in Digital Genomic Footprinting Sites

The Digital Genomic Footprinting (DGF) sites in each cell type can be visualised using the function `dgfootprints_RoadMap` using the file of DNase/DGF Footprint calls http://egg2.wustl.edu/roadmap/data/byDataType/dgfootprints/

8.3.3 DNaseI-accessible regulatory regions

Using the core 15-state chromatin state model across any of the 111 NIH Roadmap reference epigenomes, and focusing on states TssA, TssAFlnk, and TssBiv for promoters, and EnhG, Enh, and EnhBiv for enhancers, and state BivFlnk (flanking bivalent Enh/Tss) for ambiguous regions, 3 set of data were constructed. The data can be visualised using the function `DNaseI_RoadMap` with the good name of data (variable `featureDisplay`) like in Fig. 2:

- for **promoter** regions the file of tissue of interest http://egg2.wustl.edu/roadmap/data/byDataType/dnase/BED_files_prom/ or RData files containing matrice of chromatin state call for promoter. Thus, user can select for different tissues.
- for **enhancer** regions the file of tissue of interest http://egg2.wustl.edu/roadmap/data/byDataType/dnase/BED_files_enh/
- for **dyadic** promoter/enhancer region the file of tissue of interest http://egg2.wustl.edu/roadmap/data/byDataType/dnase/BED_files_dyadic/

```
chr<-"chr2"
start <- 38290160
end <- 38303219
gen<-"hg19"

extdata <- system.file("extdata", package="coMET",mustWork=TRUE)
prombedFilePath <- file.path(extdata, "/RoadMap/regions_prom_E001.bed")

promRMtrack<- DNaseI_RoadMap(gen,chr,start, end, prombedFilePath,
                             featureDisplay='promotor', type_stacking="squish")

#lost file
#enhbedFilePath <- file.path(extdata, "/RoadMap/regions_enh_E001.bed")

#enhRMtrack<- DNaseI_RoadMap(gen,chr,start, end, enhbedFilePath,
#                             featureDisplay='enhancer', type_stacking="squish")

dyabedFilePath <- file.path(extdata, "/RoadMap/regions_dyadic_E001.bed")


dyaRMtrack<- DNaseI_RoadMap(gen,chr,start, end, dyabedFilePath,
                            featureDisplay='dyadic', type_stacking="squish")

genetrack <-genes_ENSEMBL(gen,chr,start,end,showId=TRUE)

#listRoadMap <- list(genetrack,promRMtrack,enhRMtrack,dyaRMtrack)
listRoadMap <- list(genetrack,promRMtrack,dyaRMtrack)
plotTracks(listRoadMap, chromosome=chr,from=start,to=end)
```

8.3.4 Processed data and Imputed data

BED and BigWIG file can be visualised with DataTrack objects from files of *Gviz* package. The data are in http://www.genboree.org/EdaccData/Release-9/sample-experiment/ and http://www.genboree.org/EdaccData/Release-9/experiment-sample/ or go to http://egg2.wustl.edu/roadmap/web_portal/processed_data.html for processed data or to http://egg2.wustl.edu/roadmap/web_portal/imputed.html#imp_sig for imputed data.

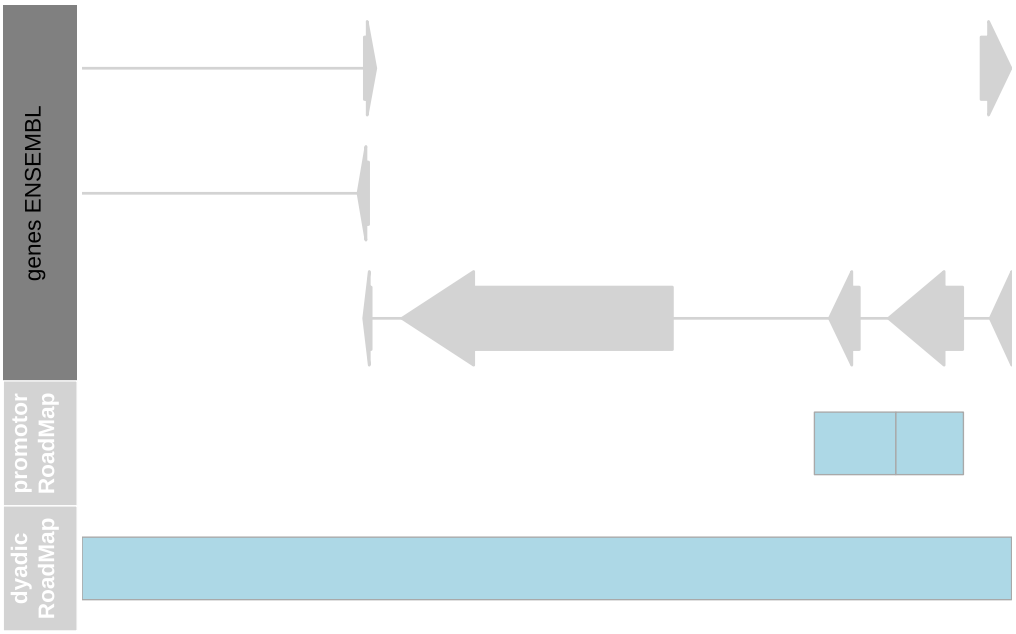

Figure 7: Plot of NIH Roadmap data.

8.4 ENCODE and GENCODE data

The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI) https://www.encodeproject.org/. The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

Genes and transcripts of GENCODE are accessible from ENSEMBL biomart or can be visualised wtith `GeneRegionTrack` of *Gviz*. Other data are in BED or BAM format that can be visualised with *Gviz* tracks.

```
#Genes from GENCODE
chr<-3
start <- 132239976
end <- 132541303
gen<-"hg19"

extdata <- system.file("extdata", package="coMET",mustWork=TRUE)
gtfFilePath <- file.path(extdata, "/GTEX/gencode.v19.genes.patched_contigs.gtf")
options(ucscChromosomeNames=FALSE)
grtrack <- GeneRegionTrack(range=gtfFilePath ,chromosome = chr, start= start,
                           end= end, name = "Gencode V19",
                           collapseTranscripts=TRUE, showId=TRUE,shape="arrow")
plotTracks(grtrack, chromosome=chr,from=start,to=end)
```

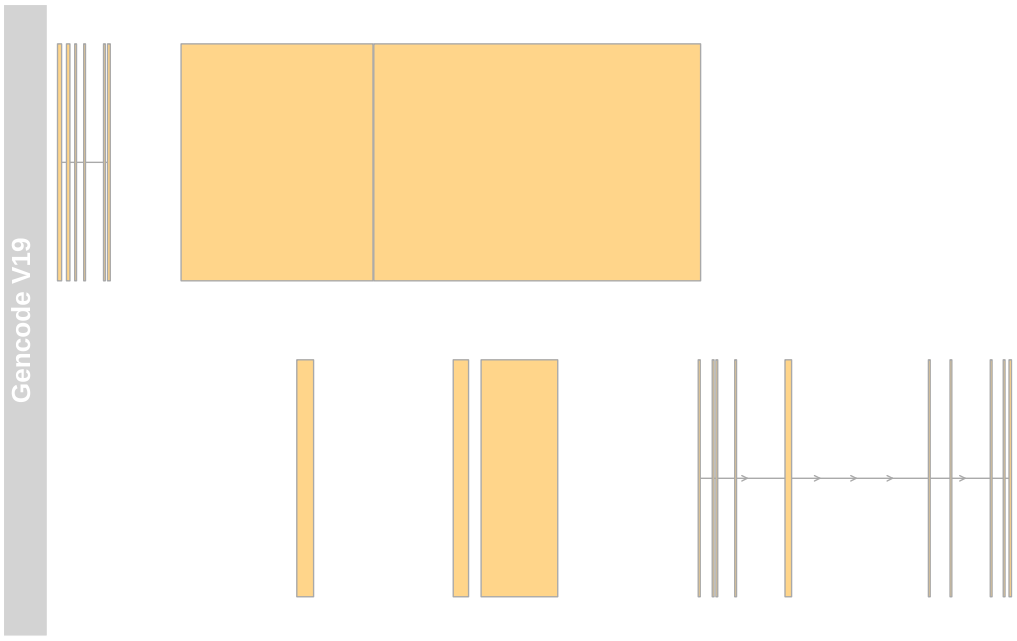

Figure 8: Plot of genes defined by GeneCode.

8.4.1 Predicting motifs and active regulators

You can browse known and discovered motifs for the ENCODE TF ChIP-seq datasets. The position of motifs can be visualised using the function `ChIPTF_ENCODE` using one of files from http://compbio.mit.edu/encode-motifs/ [7] such as http://compbio.mit.edu/encode-motifs/matches.txt.gz

```
#TF Chip-seq data
gen <- "hg19"
chr<-"chr1"
start <- 1000
end <- 329000
extdata <- system.file("extdata", package="coMET",mustWork=TRUE)
bedFilePath <- file.path(extdata, "ENCODE/motifs1000_matches_ENCODE.txt")
motif_color <- file.path(extdata, "ENCODE/TFmotifs_colors.csv")

chipTFtrack <- ChIPTF_ENCODE(gen,chr,start, end, bedFilePath,
                             featureDisplay=c("AHR::ARNT::HIF1A_1",
                                              "AIRE_1","AIRE_2","AHR::ARNT_1"),
                             motif_color,type_stacking="squish",showId=TRUE)

plotTracks(chipTFtrack, chromosome=chr,from=start,to=end)
```

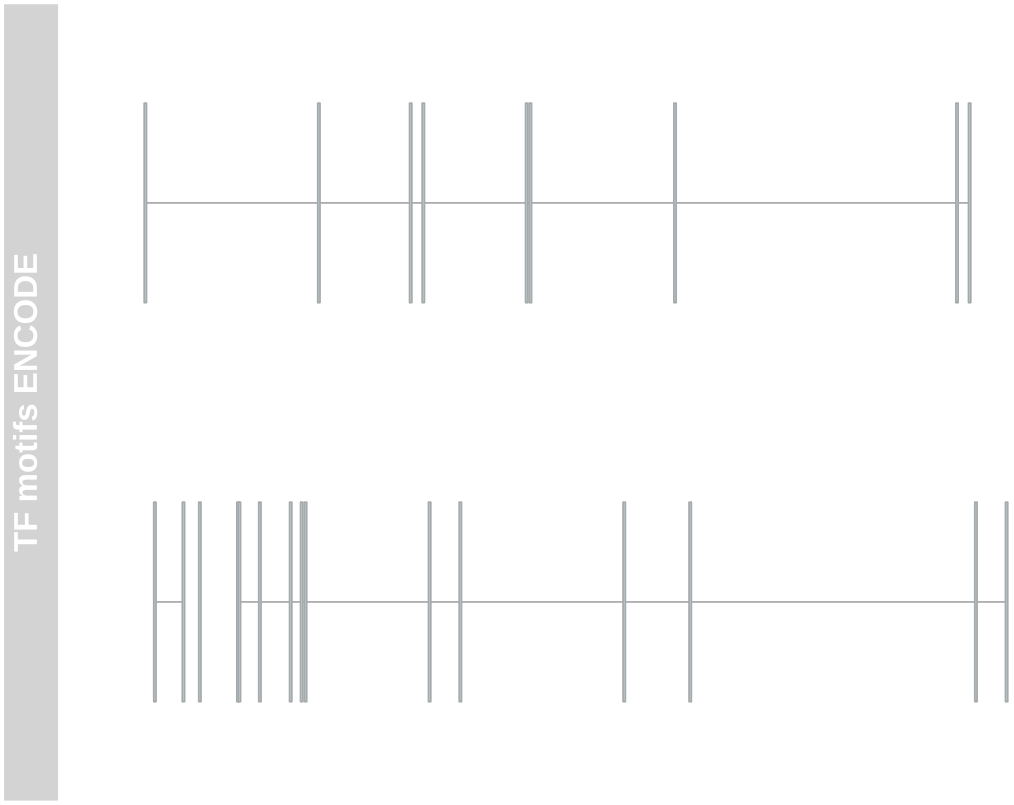

Figure 9: Plot ENCODE TF ChIP-seq datasets of ENCODE.

8.5 GTEx Portal

The Genotype-Tissue Expression (GTEx) [8] project aims to provide to the scientific community a resource with which to study human gene expression and regulation and its relationship to genetic variation. By analyzing global RNA expression within individual tissues and treating the expression levels of genes as quantitative traits, variations in gene expression that are highly correlated with genetic variation can be identified as expression quantitative trait loci, or eQTLs. The data are accessible via http://www.gtexportal.org/. A set of data are downloadable from http://www.gtexportal.org/home/datasets2 (need to have login).

The data were mapped on the reference genome **hg19**. Below described the colors of tracks and specific characteristics of some annotation tracks.

2 functions were created to visualise data from GTEx version 6:

1. `eQTL_GTEx` visualise eGene and significant snp-gene associations based on permutations in a tissue specific. The name of folder in GTEx version 6 is *GTEx_Analysis_V6_eQTLs.tar.gz*.
2. `geneExpression_GTEx` (need to update) visualise fully processed, normalized and filtered gene expression data, which was used as input into Matrix eQTL for eQTL discovery in a tissue specific. The name of folder in GTEx version 6 is *GTEx_Analysis_V6_eQTLInputFiles_geneLevelNormalizedExpression.tar.gz*

One function from *Gviz*:

1. `GeneRegionTrack` can visualise gene level model based on the GENCODE transcript model (cf. example below. Isoforms have been collapsed to single genes. The name of file in GTEx version 6 is *gencode.v19.genes.patched_contigs.gtf*.

```
## eQTL data
chr<-"chr3"
start <- 132239976
end <- 132541303
gen<-"hg19"

extdata <- system.file("extdata", package="coMET",mustWork=TRUE)
bedFilePath <- file.path(extdata, "/GTEX/eQTL_Uterus_Analysis_extract100.snpgenes")

eGTex<- eQTL_GTEx(gen,chr, start, end, bedFilePath, featureDisplay = 'all',
                  showId=TRUE, type_stacking="squish", just_group="left" )

eGTex_SNP<- eQTL_GTEx(gen,chr, start, end, bedFilePath,
                      featureDisplay = 'SNP', showId=FALSE,
                      type_stacking="dense", just_group="left")

#Genes from
gtfFilePath <- file.path(extdata, "/GTEX/gencode.v19.genes.patched_contigs.gtf")
```

```
options(ucscChromosomeNames=FALSE)
grtrack <- GeneRegionTrack(genome="hg19",range=gtfFilePath ,chromosome = chr,
                           start= start, end= end, name = "Gencode V19",
                           collapseTranscripts=TRUE, showId=TRUE,shape="arrow")
eGTexTracklist <- list(grtrack,eGTexTrackSNP)
plotTracks(eGTexTracklist, chromosome=chr,from=start,to=end)
```

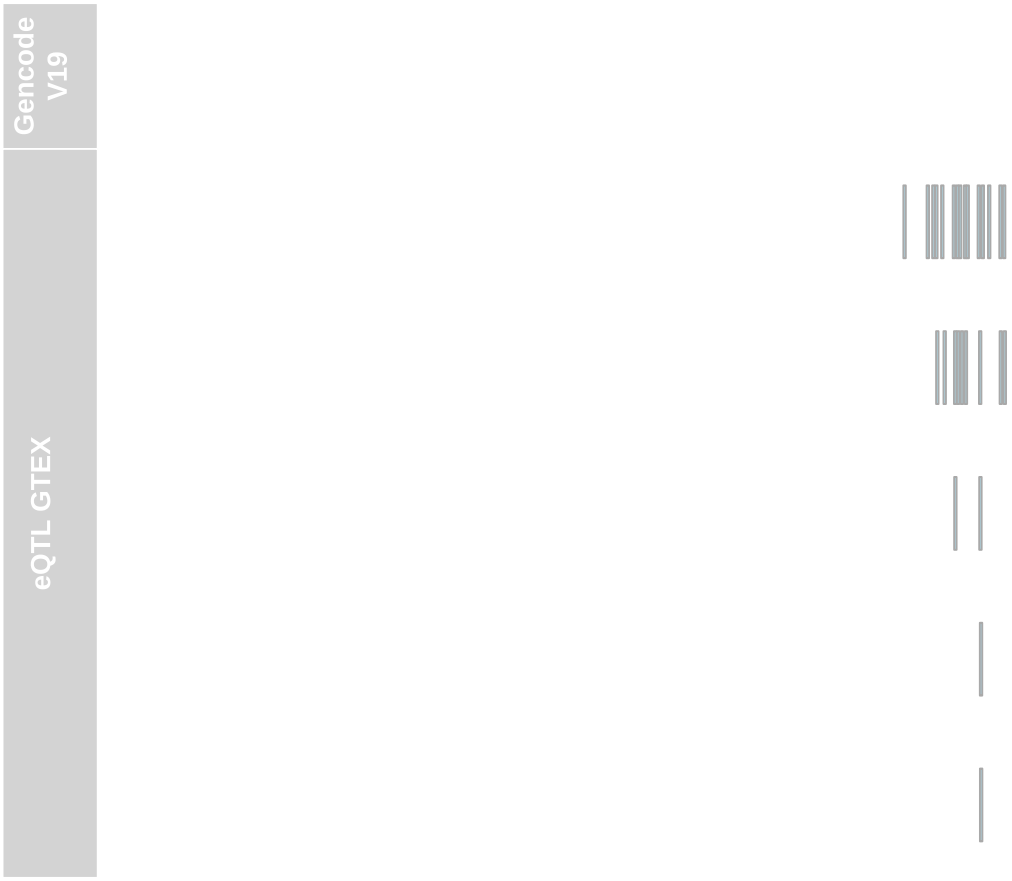

Figure 10: Plot eQTL from GTex.

2 other functions were created to visualise supplement data from GTEx version 3

1. `psiQTL_GTEx` visualise results from the protein truncating variants QTL (psiQTL) analysis for mine main tissues, plus brain, plus multi-tissue that averages the exons where data for three or more tissues is available. The name of file in GTEX version 3 is *gtex_psiqtls.zip*.
2. `imprintedGenes_GTEx` visuaslise gene imprinting genes in different tissues [9] via url http://www.gtexportal.org/home/imprintingPage. There are 33 tissues and 5 classification

```
### psiQTL
chr<-"chr13"
start <- 52713837
end <- 52715894
gen<-"hg19"

extdata <- system.file("extdata", package="coMET",mustWork=TRUE)
psiQTLFilePath <- file.path(extdata, "/GTEX/psiQTL_Assoc-total.AdiposeTissue.txt")

psiGTex<- psiQTL_GTEx(gen,chr,start, end, psiQTLFilePath, featureDisplay = 'all',
                      showId=TRUE, type_stacking="squish",just_group="above" )

genetrack <-genes_ENSEMBL(gen,chr,start,end,showId=TRUE)

psiTrack <- list(genetrack,psiGTex)
plotTracks(psiTrack, chromosome=chr,from=start,to=end)
```

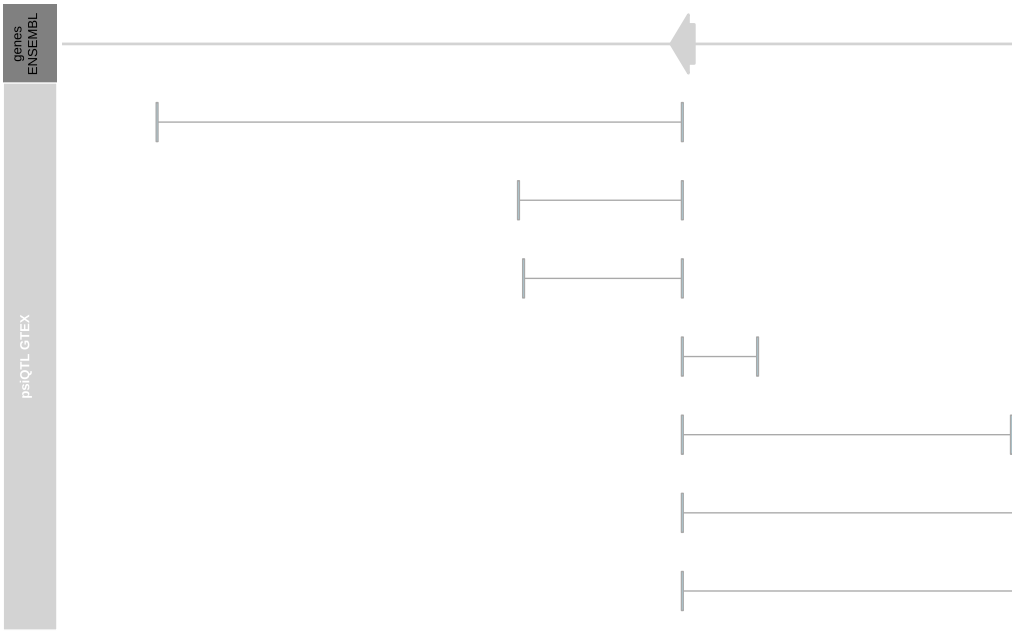

Figure 11: Plot psiQTL from GTex.

```
data(imprintedGenesGTEx)
as.character(unique(imprintedGenesGTEx$Tissue.Name))
```

```
##  [1] "Pancreas"                          "Whole_Blood"
##  [3] "Pituitary"                         "Lung"
##  [5] "Cells_EBV-transformed_lymphocytes" "Thyroid"
##  [7] "Adipose_Subcutaneous"              "Artery_Tibial"
##  [9] "Skin_Sun_Exposed_Lower_leg"        "Skin_Not_Sun_Exposed_Suprapubic"
```

```
## [11] "Brain"                                "Muscle_Skeletal"
## [13] "Breast_Mammary_Tissue"                "Nerve_Tibial"
## [15] "Adrenal_Gland"                        "Colon_Transverse"
## [17] "Prostate"                             "Artery_Coronary"
## [19] "Heart_Left_Ventricle"                 "Heart_Atrial_Appendage"
## [21] "Uterus"                               "Liver"
## [23] "Vagina"                               "Testis"
## [25] "Adipose_Visceral_Omentum"             "Fallopian_Tube"
## [27] "Esophagus_Muscularis"                 "Ovary"
## [29] "Cells_Transformed_fibroblasts"        "Esophagus_Mucosa"
## [31] "Kidney_Cortex"                        "Stomach"
## [33] "Artery_Aorta"
```

```
as.character(unique(imprintedGenesGTEx$Classification))
```

```
## [1] "consistent with biallelic"  "imprinted"                  "NC"
## [4] "consistent with imprinting" "biallelic"
```

```
### inprinted genes
chr<- "chr1"
start <- 7895752
end <- 7914572
gen<-"hg19"

genesTrack <- genes_ENSEMBL(gen,chr,start,end,showId=TRUE)

allIG <- imprintedGenes_GTEx(gen,chr,start, end, tissues="all",
                             classification="imprinted",showId=TRUE)

allimprintedIG <- imprintedGenes_GTEx(gen, chr,start, end, tissues="all",
                                      classification="imprinted",showId=TRUE)

StomachIG <-imprintedGenes_GTEx(gen,chr,start, end, tissues="Stomach",
                                classification="all",showId=TRUE)

PancreasIG <- imprintedGenes_GTEx(gen,chr,start, end,
                                  tissues="Pancreas",
                                  classification="all",showId=TRUE)
PancreasimprintedIG <- imprintedGenes_GTEx(gen,chr,start, end, tissues="Pancreas",
                                           classification="imprinted",showId=TRUE)

plotTracks(list(genesTrack, allIG, allimprintedIG,
                StomachIG,PancreasIG,PancreasimprintedIG),
           chromosome=chr, from=start, to=end)
```

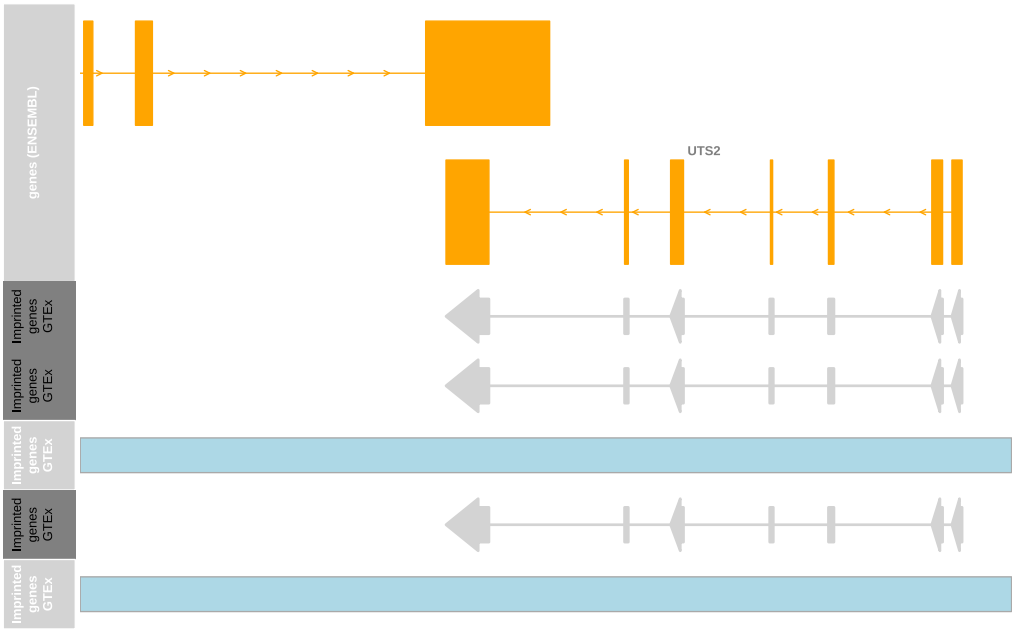

Figure 12: Plot imprinted genes from GTex.

8.6 Hi-C data

Below are examples of Hi-C data available for different tissues.

8.6.1 Hi-C data at 1kb resolution at Lieberman Aiden lab

They [10] used in situ Hi-C to probe the three-dimensional architecture of genomes, constructing haploid and diploid maps of nine cell types. The densest, in human lymphoblastoid cells, contains 4.9 billion contacts, achieving 1-kilobase resolution.The data were mapped on **hg19** reference genome.

You can download intrachromosomal matrice from http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63525 for the cell-type of interest.

```
library('corrplot')
#Hi-C data
gen <- "hg19"
chr<-"chr1"
start <- 5000000
end <- 9000000
extdata <- system.file("extdata", package="coMET",mustWork=TRUE)
bedFilePath <- file.path(extdata, "HiC/chr1_1mb.RAWobserved")

matrix_HiC <- HiCdata2matrix(chr,start, end, bedFilePath)
cor_matrix_HiC <- cor(matrix_HiC)
diag(cor_matrix_HiC)<-1
corrplot(cor_matrix_HiC, method = "circle")
```

You can quick visualise this data using this HiC-interaction tool http://promoter.bx.psu.edu/hi-c/view.php?species=human&assembly=hg19&source=inside&tissue=GM12878&resolution=1&c_url=&gene=CTXN1&sessionID=

8.6.2 Hi-C Data Browser

You can download heatmap of your region of interest from two cell-line GM06690 (immortalized lymphoblast) or K562 (leukemia) using their website http://hic.umassmed.edu/heatmap/heatmap.php. This data was produced by [11]. The region that you want to visualise with this data need to large more than either 100Kb or 1Mb as Heatmaps were generated by dividing the chromosome up into 100 Kb or 1 Mb windows. The data were mapped on **hg19** reference genome.

You need to create info file to define the position of each bin composing your interaction matrice in using the row name of matrice as the name of bin contain the start and end of bin.

8.6.3 Hi-C project at Ren Lab

Interaction matrices for each of the four cell types analysis (mouse ES cell, mouse cortex, human ES cell (H1), and IMR90 fibroblasts) by Ren Lab (to cite them, you need to select the publication for this url http://promoter.bx.psu.edu/hi-c/publications.html) are accessible via url http://chromosome.sdsc.edu/mouse/hi-c/download.html. The interaction matrices are created using either a 40kb bin size throughout the genome. So the region that you want to visualise with this data need to large more than 40Kb. The data were mapped on **hg19** reference genome.

You need to :

1. Extract from the BED file that contains the locations of each of the topological domains the region of interest
2. Extract in either raw or normalised matrice only the sub-matrice of interest

```
extdata <- system.file("extdata", package="coMET",mustWork=TRUE)
info_HiC <- file.path(extdata, "Human_IMR90_Fibroblast_topological_domains.txt")
data_info_HiC <-read.csv(info_HiC, header = FALSE, sep = "\t", quote = "")

intrachr_HiC <- file.path(extdata, "Human_IMR90_Fibroblast_Normalized_Matrices.txt")
data_intrachr_HiC <- read.csv(intrachr_HiC, header = TRUE, sep = "\t", quote = "")

chr_interest <- "chr2"
start_interest <- "1"
end_interest <- "160000"
list_bins <- which(data_info_HiC[,1] == chr_interest &
                     data_info_HiC[,2] >= start_interest &
                     data_info_HiC[,2] <= end_interest )

subdata_info_Hic <- data_info_HiC[list_bins,]
subdata_intrachr_HiC <- data_intrachr_HiC[list_bins,list_bins]
```

8.7 FANTOM5 database

FANTOM http://fantom.gsc.riken.jp/ established the FANTOM database (transcripts, transcription factors, promoters and enhancers active,TSS) and the FANTOM full-length cDNA clone bank, which are available worldwide for about 400 distinct cell types. Currently, FANTOM is in version FANTOM5 phase 2 where data were mapped on reference genome **hg19** for human or **mm9** for mouse [12].

To extract data

- from http://fantom.gsc.riken.jp/5/
- from http://fantom.gsc.riken.jp/data/ or http://fantom.gsc.riken.jp/views/
- from BED file used by UCSC HUB http://fantom.gsc.riken.jp/5/datahub/, more information here http://fantom.gsc.riken.jp/5/datahub/description.html

As the data are in classical format such as BED file, you can use easily *Gviz*'s `Data Track` function to visuaslise them. However, there are some comment lines that you need to remove in the top of files.

2 functions were created :

- `DNaseI_FANTOM` helps to visualise enhancer regions defined by FANTOM5
- `TFBS_FANTOM` helps to visualise TFBS regions defined by FANTOM5

```
gen <- "hg19"
chr<- "chr1"
start <- 6000000
end <- 6500000

extdata <- system.file("extdata", package="coMET",mustWork=TRUE)

##Enhancer
enhFantomFile <- file.path(extdata,
                           "/FANTOM/human_permissive_enhancers_phase_1_and_2.bed")
enhFANTOMtrack <-DNaseI_FANTOM(gen,chr,start, end, enhFantomFile,
                               featureDisplay='enhancer')

### TFBS motif
AP1FantomFile <- file.path(extdata, "/FANTOM/Fantom_hg19.AP1_MA0099.2.sites.txt")
tfbsFANTOMtrack <- TFBS_FANTOM(gen,chr,start, end, AP1FantomFile)
```

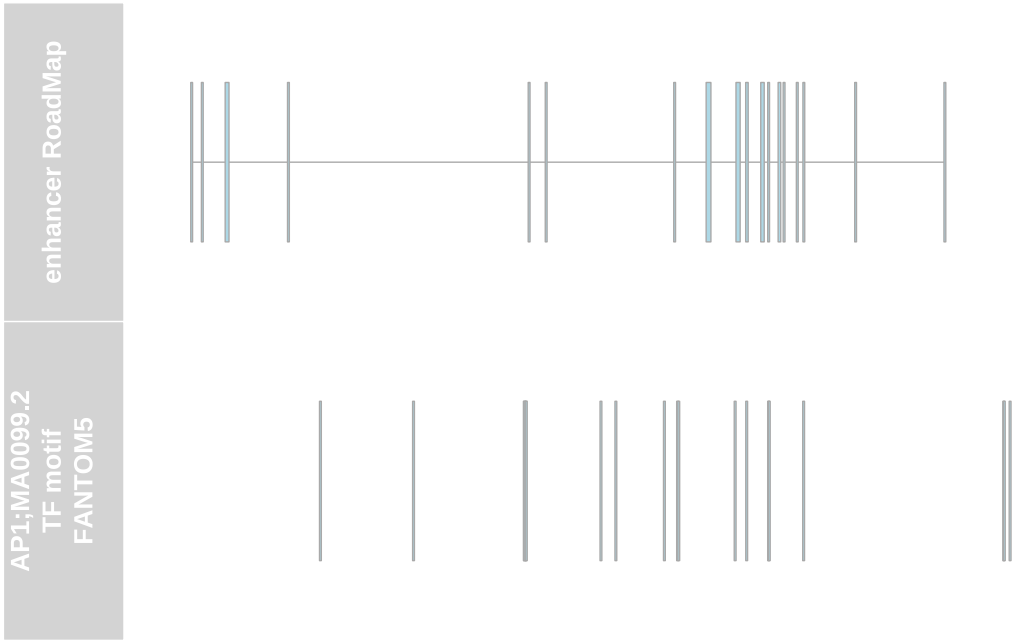

Figure 13: plot FANTOM5 data.

8.8 BLUEprint project

BLUEprint http://www.blueprint-epigenome.eu/ aims to further the understanding of how genes are activated or repressed in both healthy and diseased human cells. BLUEPRINT will focus on distinct types of haematopoietic cells from healthy individuals and on their malignant leukaemic counterparts.

the data were mapped on reference genome partially on **GRCh37** and all on **GRCh38**.

As the data are in classical format such as BED file, BigWig of GTF, you can use easily `DataTrack` or `AnnotationTrack` of *Gviz* to visuaslise them.

8.9 Our data

8.9.1 eQTL data

You can visualise our eQTL using `eQTL` function. Listed below are the colours used for the different elements contained in eQTL data.

Feature	Colour	Hex Code
SNP_pheno		#A6CEE3
SNP		#1F78B4
exon		#33A02C
exon_pheno		#B2DF8A
mRNA		#FB9A99
mRNA_pheno		#E31A1C
cis_local_eQTL		#FDBF6F
trans_local_eQTL		#FF7F00
distal_eQTL		#CAB2D6
cis_local_eQTL_pheno		#6A3D9A
trans_local_eQTL_pheno		#FFFF99
distal_eQTL_pheno		#B15928

8.9.2 metQTL data

You can visualise our eQTL using `metQTL` function.

Listed below are the colours used for the different elements contained in metQTL data.

Feature	Colour	Hex Code
SNP_pheno		#A6CEE3
SNP		#1F78B4
CpG_pheno		#33A02C
CpG		#B2DF8A
cis_local_metQTL		#FB9A99
trans_local_metQTL		#E31A1C
distal_metQTL		#FDBF6F
cis_local_metQTL_pheno		#FF7F00
trans_local_metQTL_pheno		#CAB2D6
distal_metQTL_pheno		#6A3D9A

9 *coMET*: *Shiny* web-service

9.1 How to use the *coMET* web-service

If you want to use *coMET* via its webservice, please go to http://epigen.kcl.ac.uk/comet and select one of different instances or direcly access one of the instances, for example http://comet.epigen.kcl.ac.uk:3838/coMET/. We have created different instances of *coMET* because we did not have access to the pro version of *Shiny*. All instances use the same version of *coMET*.

If you use *coMET* from a *Shiny* webservice, you do not need to install the *coMET* package on your computer. The web service is user friendly and requires input files and configuration of the plot. The creation of the *coMET* plot can take some time because it makes a live connection to UCSC or/and ENSEMBL for the annotation tracks. First, the plot is created on the webpage, and then it can be saved as an output file. For better quality plots please use the download option and the plot will be recreated in a file in pdf or eps format.

9.2 How to install the *coMET* web-service

These are different steps to install *coMET* on your *Shiny* web-service and you need to be root to install it.

1. You need to install *R*, *Bioconductor* and the *coMET* package under root.
2. You need first to install the *Shiny* and *rmarkdown* R package before *Shiny* Server.

   ```
   sudo su - -c "R -e \"install.packages('shiny', repos='http://cran.rstudio.com/')\""
   sudo su - -c "R -e \"install.packages('rmarkdown', repos='http://cran.rstudio.com/')\""
   ```

3. You can install *Shiny* Server http://shiny.rstudio.com/, go to https://www.rstudio.com/products/shiny/download-server/.

   ```
   sudo apt-get install gdebi-core
   wget \url{https://download3.rstudio.org/ubuntu-12.04/x86_64/shiny-server-1.4.2.786-amd64.deb}
   sudo gdebi shiny-server-1.4.2.786-amd64.deb
   ```

4. *Shiny* Server should now be installed and running on port 3838. You should be able to see a default welcome screen at http://your_server_ip:3838/. You can make sure your *Shiny* Server is working properly by going to http://your_server_ip:3838/sample-apps/hello/.
5. You now have a functioning *Shiny* Server that can host *Shiny* applications or interactive documents. The configuration file for *Shiny* Server is at /etc/shiny-server/shiny-server.conf. By default it is configured to serve applications in the /srv/shiny-server/ directory. This means that any *Shiny* application that is placed at /srv/shiny-server/app_name will be available to the public at http://your_server_ip:3838/app_name/.

6. In a *Shiny*'s folder (e.g. /var/shiny-server/www), you can create a folder called "COMET".
7. Following this, you can install the two *coMET* scripts in www of the *coMET* package, within this new folder.
8. You need to change owner and permissions to access this folder. Only the user called *Shiny* can access it.

```
mkdir -p /var/shiny-server/www/COMET

chmod -R 755 /var/shiny-server/www/COMET
chown -R shiny:shiny /var/shiny-server/www/COMET

mkdir -p /var/shiny-server/log

chmod -R 755 /var/shiny-server/log
chown -R shiny:shiny /var/shiny-server/log
```

9. You need now to update the configuration file of *Shiny* (e.g. /etc/shiny-server/shiny-server.conf).
10. You need to change owner and the permission to access this file

```
chmod 744 /etc/shiny-server/shiny-server.conf
chown shiny:shiny /etc/shiny-server/shiny-server.conf
```

11. At the end, you should restart the service *Shiny* via the command line:

```
###2.13.0.1 systemd (RedHat 7, Ubuntu 15.04+, SLES 12+)
#File to change:
/etc/systemd/system/shiny-server.service

#How to define the environment variable:
[Service]
Environment="SHINY\_LOG\_LEVEL=TRACE"

#Commands to run for the changes to take effect:
sudo systemctl stop shiny-server
sudo systemctl daemon-reload
sudo systemctl start shiny-server


###2.13.0.2 Upstart (Ubuntu 12.04 through 14.10 and RedHat 6)
#File to change:
/etc/init/shiny-server.conf

#How to define the environment variable:
env SHINY\_LOG\_LEVEL=TRACE
```

```
#Commands to run for the changes to take effect:
sudo restart shiny-server
```

Your Shiny's configuration file:

```
run_as shiny;
# Define a top-level server which will listen on a port
server {
  # Instruct this server to listen on port 3838
  listen 3838;
  # Define the location available at the base URL
  location / {
    # Run this location in 'site_dir' mode, which hosts the entire directory
    # tree at '/srv/shiny-server'
    site_dir /var/shiny-server/www;

    # Define where we should put the log files for this location
    log_dir /var/shiny-server/log;

    # Should we list the contents of a (non-Shiny-App) directory when the user
    # visits the corresponding URL?
    directory_index off;

    #   app_init_timeout 3600;
    #   app_idle_timeout 3600;
  }

}
```

10 FAQs

- **I cannot see my plot after running `comet` or `comet.web`. What should I do?**
  If the previous time `comet` or `comet.web` ran and error was produced it prevents the plot from being closed. to fix this use the command '*dev.off()*' as many times as necessary.

- **How do we know if my track has data? and what the data is?**
  Type the name of your track, visualise the track with `plotTrack` or read different parameters with `str` function.

```
genetrack <-genesENSEMBL(gen, chrom, start,
                         end, showId=TRUE)

plotTracks(genetrack)

str(genetrack)
```

- **How do you increase the size of the font of the name of an object**?
  To enlarge the name of gene, as the object is *Gviz* object, you can use the option from *Gviz*.
  You can see the value of different parameters via this command line:

```
genetrack <-genesENSEMBL(gen, chrom, start,
                         end, showId=TRUE)

displayPars(genetrack)
```

  So if you want to enlarge the name of gene, you need to do use the option `fontsize.gviz` in the `comet` function, an example is given below:

```
comet(config.file = configfile, mydata.file = myinfofile,
      mydata.format = "file",
      cormatrix.file = mycorrelation,
      cormatrix.type = "listfile",
      mydata.large.file = mylargedata,
      mydata.large.type = "listfile",
      tracks.gviz = listGviz, verbose = TRUE,
      print.image=TRUE, fontsize.gviz=10)
```

- **Can I make a selection of which genes or transcripts to display**?
  To make a selection of genes to display first create the track like you would if you were displaying all genes. From this track create another with only the genes you want to display like in the example below. Please note it is not possible to select genes based on their names unless the option to display gene names instead of gene reference is used, in other cases it is possible to make a selection based on the genes reference number.

```
geneTrack <- refGenesUCSC(gen, chr, start, end,
                          IdType ="name", showId = TRUE)

geneTrackShow <- geneTrack[gene(geneTrack) %in% c("AHRR")]
```

- **How can I better understand where the `comet` function stopped**?
  Use option *VERBOSE=TRUE* in the function `coMET` or `coMET.web`
  If this does not help resolve the issue, please to send your command line with *VERBOSE=TRUE* and its error message to `tiphaine.martin@mssm.edu`. Do not forget to give alsoinformation about the session by using `sessionInfo()`.

- **How do you visualise coMET plots working within a R *Markdown* or *knitr* framework**?
  When coMET writes to a PDF, it is writing out to a 7X7 square area. So, it turns out that one can 'force; the R *Markdown* block as well as *knitr* block to also write to a 7 x 7 square PDF using option for chunck `fig.height=7,fig.width=7` , as follows:

```
'''{r plot_ex1,fig.keep='last', fig.height=7, fig.width=7,dev='pdf'}
comet(config.gile=configfile,
      mydata.file=myinfofile, mydata_type="file",
      cormatrix.file-mycorrelation,
      cormatrix.type="listfile",
      mydata.large.file=myexpressfile,
      mydata.large.type="listfile",
      tracks.gviz=listgviz,
      verbose=FALSE, print.image=FALSE,
      disp.pvalueplot=FALSE)
'''
```

11 Acknowledgement

T.C.M would like to thank Bioconductor team for their help and advice in the development of a R Bioconductor package. Moreover, T.C.M would like to thank different users for their feedback that help to improve this present R package.

- Prof Daniel Weeks and Dr Annie Infancia Arockiaraj to share with us how to visualise correctly coMET plot in R Markdown code.

12 SessionInfo

The following is the session info that generated this vignette:

- R version 4.2.1 (2022-06-23), `x86_64-pc-linux-gnu`
- Locale: `LC_CTYPE=en_US.UTF-8`, `LC_NUMERIC=C`, `LC_TIME=en_GB`, `LC_COLLATE=C`, `LC_MONETARY=en_US.UTF-8`, `LC_MESSAGES=en_US.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`
- Running under: `Ubuntu 20.04.5 LTS`
- Matrix products: default
- BLAS: `/home/biocbuild/bbs-3.16-bioc/R/lib/libRblas.so`
- LAPACK: `/home/biocbuild/bbs-3.16-bioc/R/lib/libRlapack.so`
- Base packages: base, datasets, grDevices, graphics, grid, methods, stats, stats4, utils
- Other packages: BiocGenerics 0.44.0, GenomeInfoDb 1.34.0, GenomicRanges 1.50.0, Gviz 1.42.0, IRanges 2.32.0, S4Vectors 0.36.0, biomaRt 2.54.0, coMET 1.30.0, knitr 1.40, psych 2.2.9, showtext 0.9-5, showtextdb 3.0, sysfonts 0.8.8
- Loaded via a namespace (and not attached): AnnotationDbi 1.60.0, AnnotationFilter 1.22.0, BSgenome 1.66.0, Biobase 2.58.0, BiocFileCache 2.6.0, BiocIO 1.8.0, BiocManager 1.30.19, BiocParallel 1.32.0, BiocStyle 2.26.0, Biostrings 2.66.0, DBI 1.1.3, DelayedArray 0.24.0, Formula 1.2-4, GenomeInfoDbData 1.2.9, GenomicAlignments 1.34.0, GenomicFeatures 1.50.0, Hmisc 4.7-1, KEGGREST 1.38.0, Matrix 1.5-1, MatrixGenerics 1.10.0, ProtGenerics 1.30.0, R6 2.5.1, RColorBrewer 1.1-3, RCurl 1.98-1.9, RSQLite 2.2.18, Rcpp 1.0.9, Rsamtools 2.14.0, SummarizedExperiment 1.28.0, VariantAnnotation 1.44.0, XML 3.99-0.12, XVector 0.38.0, assertthat 0.2.1, backports 1.4.1, base64enc 0.1-3, biovizBase 1.46.0, bit 4.0.4, bit64 4.0.5, bitops 1.0-7, blob 1.2.3, cachem 1.0.6, checkmate 2.1.0, cli 3.4.1, cluster 2.1.4, codetools 0.2-18, colorspace 2.0-3, compiler 4.2.1, corrplot 0.92, crayon 1.5.2, curl 4.3.3, data.table 1.14.4, dbplyr 2.2.1, deldir 1.0-6, dichromat 2.0-0.1, digest 0.6.30, dplyr 1.0.10, ellipsis 0.3.2, ensembldb 2.22.0, evaluate 0.17, fansi 1.0.3, fastmap 1.1.0, filelock 1.0.2, foreign 0.8-83, generics 0.1.3, ggplot2 3.3.6, glue 1.6.2, gridExtra 2.3, gtable 0.3.1, hash 2.2.6.2, highr 0.9, hms 1.1.2, htmlTable 2.4.1, htmltools 0.5.3, htmlwidgets 1.5.4, httr 1.4.4, interp 1.1-3, jpeg 0.1-9, lattice 0.20-45, latticeExtra 0.6-30, lazyeval 0.2.2, lifecycle 1.0.3, magrittr 2.0.3, matrixStats 0.62.0, memoise 2.0.1, mnormt 2.1.1, munsell 0.5.0, nlme 3.1-160, nnet 7.3-18, parallel 4.2.1, pillar 1.8.1, pkgconfig 2.0.3, png 0.1-7, prettyunits 1.1.1, progress 1.2.2, purrr 0.3.5, rappdirs 0.3.3, restfulr 0.0.15, rjson 0.2.21, rlang 1.0.6, rmarkdown 2.17,

rpart 4.1.19, rstudioapi 0.14, rtracklayer 1.58.0, scales 1.2.1, splines 4.2.1, stringi 1.7.8, stringr 1.4.1, survival 3.4-0, tibble 3.1.8, tidyselect 1.2.0, tools 4.2.1, utf8 1.2.2, vctrs 0.5.0, withr 2.5.0, xfun 0.34, xml2 1.3.3, yaml 2.3.6, zlibbioc 1.44.0

References

[1] A. Luna and K.K. Nicodemus. snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. *Bioinformatics*, 23:774–6, 2007.

[2] Fiona Cunningham, M. Ridwan Amode, Daniel Barrell, Kathryn Beal, Simon Billis, Konstantinos ad Brent, Denise Carvalho-Silva, Peter Clapham, Guy Coates, Stephen Fitzgerald, Laurent Gil, Carlos Garcín Girón, Leo Gordon, Thibaut Hourlier, Sarah E. Hunt, Sophie H. Janacek, Nathan Johnson, Thomas Juettemann, Andreas K. Kähäri, Stephen Keenan, Fergal J. Martin, Thomas Maurel, William McLaren, Daniel N. Murphy, Rishi Nag, Bert Overduin, Anne Parker, Mateus Patricio, Emily Perry, Miguel Pignatelli, Harpreet Singh Riat, Daniel Sheppard, Kieron Taylor, Anja Thormann, Alessandro Vullo, Steven P. Wilder, Amonida Zadissa, Bronwen L. Aken, Ewan Birney, Jennifer Harrow, Rhoda Kinsella, Matthieu Muffato, Magali Ruffier, Stephen M.J. Searle, Giulietta Spudich, Stephen J. Trevanion, Andy Yates, Daniel R. Zerbino, and Paul Flicek. Ensembl 2015. *Nucleic Acids Research*, 43:D662–D669, 2015. doi:10.1093/nar/gku1010.

[3] Andrew Yates, Kathryn Beal, Stephen Keenan, William McLaren, Miguel Pignatelli, Graham R. S. Ritchie, Magali Ruffier, Kieron Taylor, Alessandro Vullo, and Paul Flicek. The Ensembl REST API: Ensembl Data for Any Language. *Bioinformatics*, 31:143–45, 2014. doi:10.1093/bioinformatics/btu613.

[4] W.J. Kent, C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler. The human genome browser at UCSC. *Genome Res.*, 12:996–1006, 2002.

[5] B.J. Raney, T.R. Dreszer, G.P. Barber, H. Clawson, P.A. Fujita, T. Wang, N. Nguyen, B. Paten, A.S. Zweig, D. Karolchik, and W.J. Kent. Track Data Hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Genome Res.*, 30:1003–5, 2013.

[6] Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, Pouya Kheradpour, Zhizhuo Zhang, Jianrong Wang, Michael J. Ziller, Viren Amin, John W. Whitaker, Matthew D. Schultz, Lucas D. Ward, Abhishek Sarkar, Gerald Quon, Richard S. Sandstrom, Matthew L. Eaton, Yi-Chieh Wu, Andreas R. Pfenning, Xinchen Wang, Melina Claussnitzer, Yaping Liu, Cristian Coarfa, R. Alan Harris, Noam Shoresh, Charles B. Epstein, Elizabeta Gjoneska, Danny Leung, Wei Xie, R. David Hawkins, Ryan Lister, Chibo Hong, Philippe Gascard, Andrew J. Mungall, Richard Moore, Eric Chuah, Angela Tam, Theresa K. Canfield, R. Scott Hansen, Rajinder Kaul, Peter J. Sabo, Mukul S. Bansal, Annaick Carles, Jesse R. Dixon, Kai-How Farh, Soheil Feizi, Rosa Karlic, Ah-Ram Kim, Ashwinikumar Kulkarni, Daofeng Li, Rebecca Lowdon, GiNell Elliott, Tim R. Mercer, Shane J. Neph, Vitor Onuchic, Paz Polak, Nisha

Rajagopal, Pradipta Ray, Richard C. Sallari, Kyle T. Siebenthall, Nicholas A. Sinnott-Armstrong, Michael Stevens, Robert E. Thurman, Jie Wu, Bo Zhang, Xin Zhou, Arthur E. Beaudet, Laurie A. Boyer, Philip L. De Jager, Peggy J. Farnham, Susan J. Fisher, David Haussler, Steven J. M. Jones, Wei Li, Marco A. Marra, Michael T. McManus, Shamil Sunyaev, James A. Thomson, Thea D. Tlsty, Li-Huei Tsai, Wei Wang, Robert A. Waterland, Michael Q. Zhang, Lisa H. Chadwick, Bradley E. Bernstein, Joseph F. Costello, Joseph R. Ecker, Martin Hirst, Alexander Meissner, Aleksandar Milosavljevic, Bing Ren, John A. Stamatoyannopoulos, Ting Wang, Manolis Kellis, Andreas Pfenning, Melina ClaussnitzerYaping Liu, R. Alan Harris, R. David Hawkins, R. Scott Hansen, Nezar Abdennur, Mazhar Adli, Martin Akerman, Luis Barrera, Jessica Antosiewicz-Bourget, Tracy Ballinger, Michael J. Barnes, Daniel Bates, Robert J. A. Bell, David A. Bennett, Katherine Bianco, Christoph Bock, Patrick Boyle, Jan Brinchmann, Pedro Caballero-Campo, Raymond Camahort, Marlene J. Carrasco-Alfonso, Timothy Charnecki, Huaming Chen, Zhao Chen, Jeffrey B. Cheng, Stephanie Cho, Andy Chu, Wen-Yu Chung, Chad Cowan, Qixia Athena Deng, Vikram Deshpande, Morgan Diegel, Bo Ding, Timothy Durham, Lorigail Echipare, Lee Edsall, David Flowers, Olga Genbacev-Krtolica, Casey Gifford, Shawn Gillespie, Erika Giste, Ian A. Glass, Andreas Gnirke, Matthew Gormley, Hongcang Gu, Junchen Gu, David A. Hafler, Matthew J. Hangauer, Manoj Hariharan, Meital Hatan, Eric Haugen, Yupeng He, Shelly Heimfeld, Sarah Herlofsen, Zhonggang Hou, Richard Humbert, Robbyn Issner, Andrew R. Jackson, Haiyang Jia, Peng Jiang, Audra K. Johnson, Theresa Kadlecek, Baljit Kamoh, Mirhan Kapidzic, Jim Kent, Audrey Kim, Markus Kleinewietfeld, Sarit Klugman, Jayanth Krishnan, Samantha Kuan, Tanya Kutyavin, Ah-Young Lee, Kristen Lee, Jian Li, Nan Li, Yan Li, Keith L. Ligon, Shin Lin, Yiing Lin, Jie Liu, Yuxuan Liu, C. John Luckey, Yussanne P. Ma, Cecile Maire, Alexander Marson, John S. Mattick, Michael Mayo, Michael McMaster, Hayden Metsky, Tarjei Mikkelsen, Diane Miller, Mohammad Miri, Eran Mukame, Raman P. Nagarajan, Fidencio Neri, Joseph Nery, Tung Nguyen, Henriette O'Geen, Sameer Paithankar, Thalia Papayannopoulou, Mattia Pelizzola, Patrick Plettner, Nicholas E. Propson, Sriram Raghuraman, Brian J. Raney, Anthony Raubitschek, Alex P. Reynolds, Hunter Richards, Kevin Riehle, Paolo Rinaudo, Joshua F. Robinson, Nicole B. Rockweiler, Evan Rosen, Eric Rynes, Jacqueline Schein, Renee Sears, Terrence Sejnowski, Anthony Shafer, Li Shen, Robert Shoemaker, Mahvash Sigaroudinia, Igor Slukvin, Sandra Stehling-Sun, Ron Stewart, Sai Lakshmi Subramanian, Kran Suknuntha, Scott Swanson, Shulan Tian, Hannah Tilden, Linus Tsai, Mark Urich, Ian Vaughn, Jeff Vierstra, Shinny Vong, Ulrich Wagner, Hao Wang, Tao Wang, Yunfei Wang, Arthur Weiss, Holly Whitton, Andre Wildberg, Heather Witt, Kyoung-Jae Won, Mingchao Xie, Xiaoyun Xing, Iris Xu, Zhenyu Xuan, Zhen Ye, Chia-an Yen, Pengzhi Yu, Xian Zhang, Xiaolan Zhang, Jianxin Zhao, Yan Zhou, Jiang Zhu, Yun Zhu, and Steven Ziegler. Integrative analysis of 111 reference human epigenomes. *Nature*, 518(7539):317–330, February 2015. URL: http://dx.doi.org/10.1038/nature14248, doi:10.1038/nature14248.

[7] Pouya Kheradpour and Manolis Kellis. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic acids research*, 42(5):2976–87, 2014. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3950668&tool=pmcentrez&rendertype=abstract, doi:10.1093/nar/gkt1249.

[8] The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.*, 45:580–5, 2013. doi:10.1038/ng.2653.

[9] Y Baran, M Subramaniam, A Biton, T Tukiainen, E.K. Tsang, M.A. Rivas, M. Pirinen, M. Gutierrez-Arcelus, K.S. Smith, K.R. Kukurba, R Zhang, C Eng, D.G. Torgerson, C Urbanek, GTEx Consortium, J.B. Li, J.R. Rodriguez-Santana, E.G. Burchard, M.A. Seibold, D.G. MacArthur, S.B. Montgomery, N.A. Zaitlen, and T Lappalainen. The landscape of genomic imprinting across diverse adult human tissues. *Genome Res.*, 25:927–36, 2015. doi:10.1101/gr.192278.115.

[10] Suhas S.P. Rao, Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, Ido Machol, Arina D. Omer, Eric S. Lander, and Erez Lieberman Aiden. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, 159(7):1665–1680, December 2014. URL: http://www.sciencedirect.com/science/article/pii/S0092867414014974, doi:10.1016/j.cell.2014.11.021.

[11] Erez Lieberman-Aiden, Nynke L van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, Richard Sandstrom, Bradley Bernstein, M A Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A Mirny, Eric S Lander, and Job Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, 326(5950):289–93, October 2009. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2858594&tool=pmcentrez&rendertype=abstract, doi:10.1126/science.1181369.

[12] Marina Lizio, Jayson Harshbarger, Hisashi Shimoji, Jessica Severin, Takeya Kasukawa, Serkan Sahin, Imad Abugessaisa, Shiro Fukuda, Fumi Hori, Sachi Ishikawa-Kato, Christopher J Mungall, Erik Arner, J Kenneth Baillie, Nicolas Bertin, Hidemasa Bono, Michiel de Hoon, Alexander D Diehl, Emmanuel Dimont, Tom C Freeman, Kaori Fujieda, Winston Hide, Rajaram Kaliyaperumal, Toshiaki Katayama, Timo Lassmann, Terrence F Meehan, Koro Nishikata, Hiromasa Ono, Michael Rehli, Albin Sandelin, Erik A. Schultes, Peter A.C. Hoen, Zuotian Tatum, Mark Thompson, Tetsuro Toyoda, Derek W Wright, Carsten O. Daub, Masayoshi Itoh, Piero Carninci, Yoshihide Hayashizaki, Alistair R R Forrest, and Hideya Kawaji. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome biology*, 16(1):22, January 2015. URL: http://genomebiology.com/2015/16/1/22, doi:10.1186/s13059-014-0560-6.