# Introduction to FunTFPair package

December 11, 2014

# Contents

# 1 Introduction

Transcription factors (TFs) are fundamental regulators of gene expression and generally function in a complex and coopera-tive manner. Identifying context-dependent cooperative TFs is es-sential for understanding how cells respond to environmental change. The huge amount of various omics data currently available, providing genome-wide physical binding and functional effect information about transcription, holds a great opportunity to study TF cooperativity. Here we developed an R package, FunTFPair, which provides an easy and powerful way to identify condition-specific TF pairs by integrating transcription factor binding sites from ENCODE and gene expression profiles from GEO. Users only need to provide the GEO ID that they are interested in. FunTFPair will automatically retrieve the expression profiles of the input GEO ID, get TF targets from ENCODE, and identify TF pairs whose common targets show statistically significant differences under changed experimental conditions or exhibit coordinated transcription in a specific condition. The functional TF pairs and their relative importance will be reported in a TF cooperation network. Two datasets from GEO are used as examples to demonstrate the usage and reliability of the package.

# 2 Functions in the package

- *prepareGeoData*: A function to download GDS or GSE data and convert ID if needed;
- *differentialAnalysis*: A function to perform differential analysis and then find the enrichment of target genes;
- *correlationAnalysis*: A function to perform correlation analysis for expression data, to see if the shared targets of two TF were significantly changed between two conditions;
- *networkVis*: A function to perform functional TF network visualization;

# 3 Transcription factor pairs dataset from ENCODE project

FunTFPair package uses the transcription factor pairs dataset from ENCODE project, which contains 115 transcription factors and 4655 pairs in 12 cell lines. You can find them in the variable *pairs2TargetRemDup*. FunTFpair package will use it to find functional TF pairs in a specified condition.

```
# dataset in pairs2TargetRemDup variable
library(FunTFPair)
str(pairs2TargetRemDup[, 1:3])
```

```
## 'data.frame': 4655 obs. of  3 variables:
##  $ HGNC.TF1 : chr  "MAFK" "RAD21" "RAD21" "RAD21" ...
##  $ HGNC.TF2 : chr  "MAFF" "CTCF" "CTCF" "CTCF" ...
##  $ Cell.line: chr  "HEPG2" "GM12878" "HEPG2" "K562" ...
```

```
unique(pairs2TargetRemDup$Cell.line)
```

```
##  [1] "HEPG2"    "GM12878" "K562"    "HELAS3"  "H1HESC"
##  [6] "SKNSHRA" "HUVEC"   "NB4"     "HCT116"  "A549"
## [11] "MCF7"    "NHEK"
```

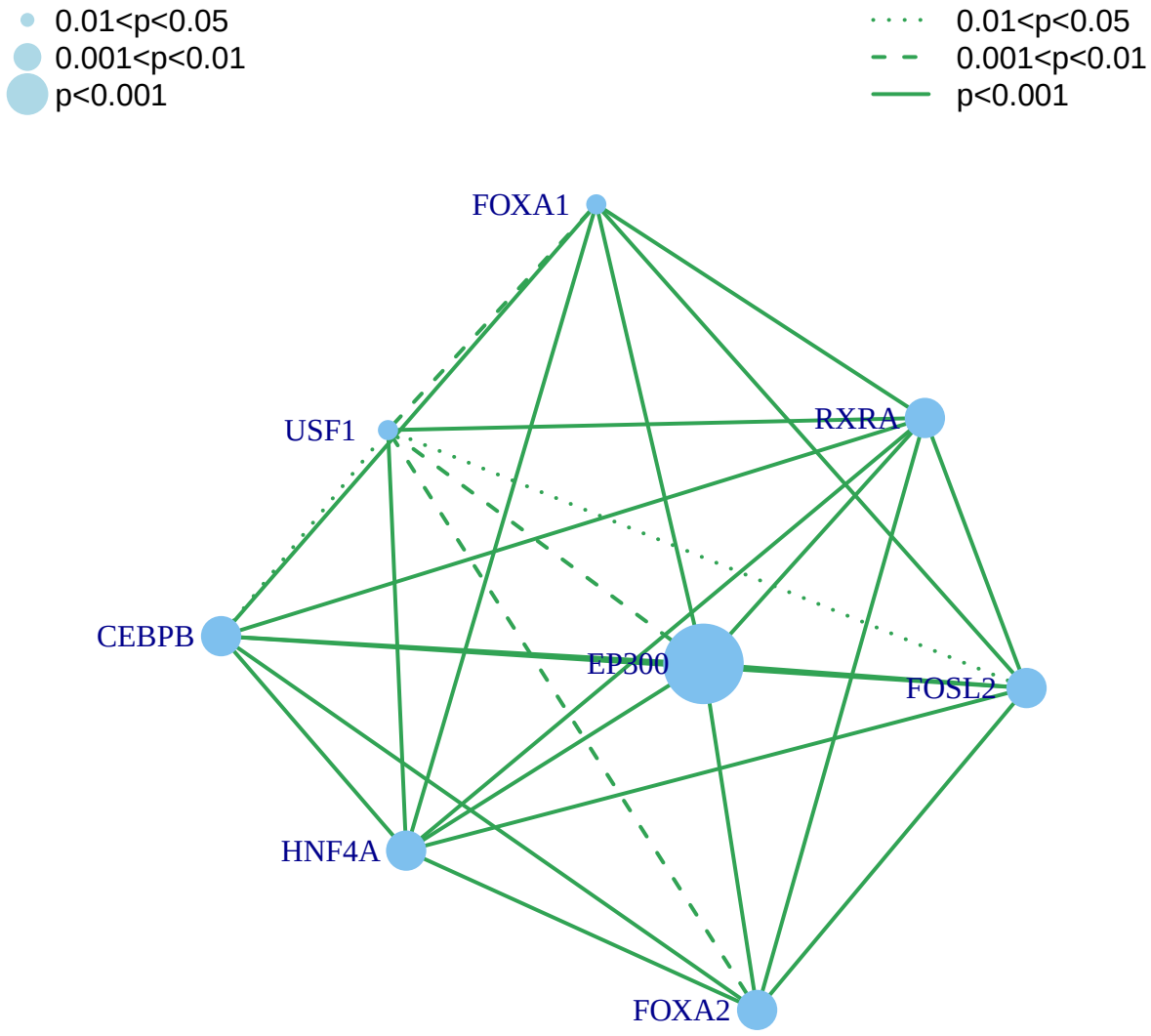
If you are only interesed in TF pairs in one specified cell line, you can select part of the *pairs2TargetRemDup* variable and use it in the analysis functions.

```
# select TF pairs in HEPG2 cell line. Then you can use
# tfPairsUsed variable in analysis function.
tfPairsUsed <- pairs2TargetRemDup[which(pairs2TargetRemDup$Cell.line ==
    "HEPG2"), ]
```

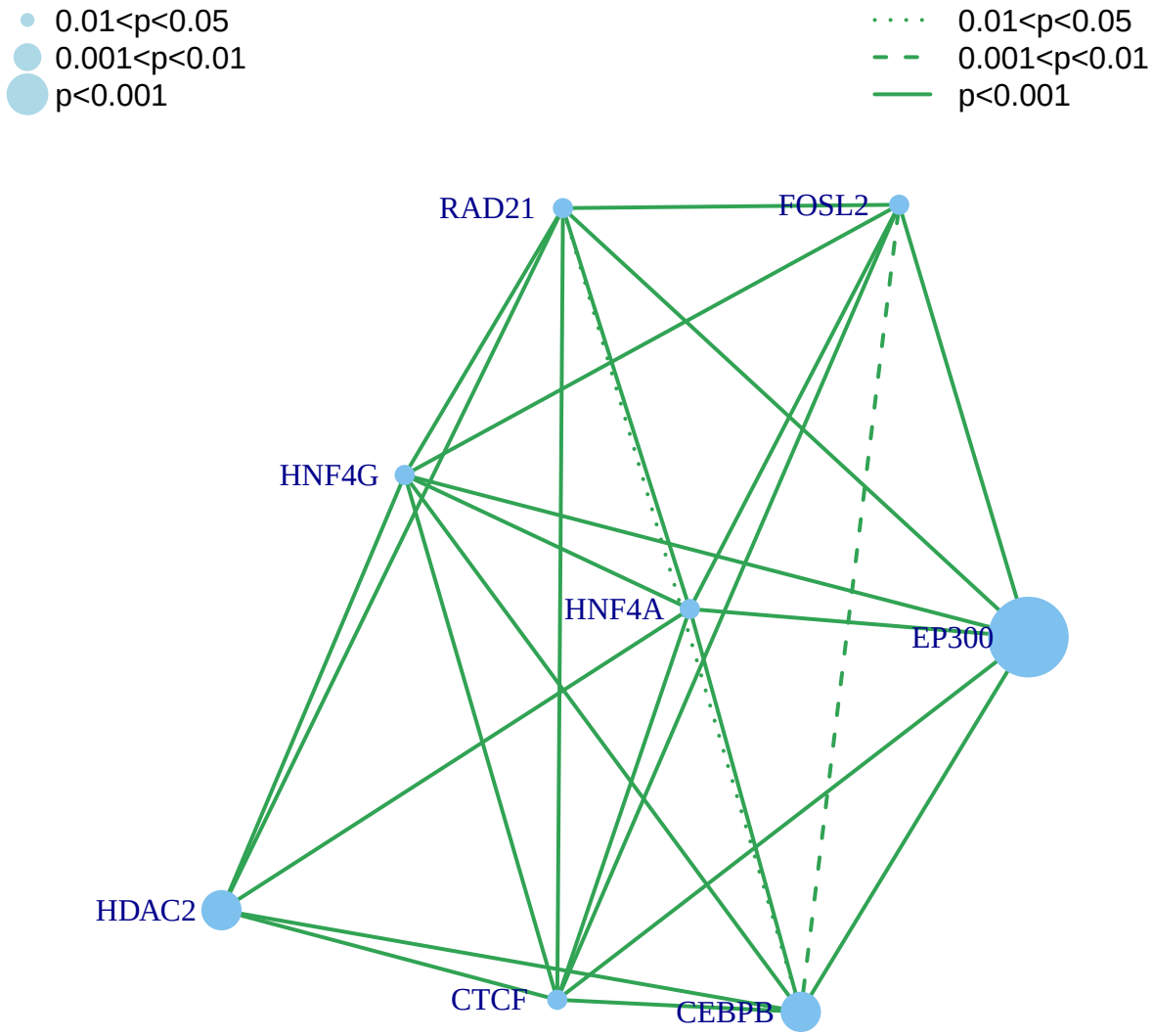
# 4 Example of differential analysis-based functional TF pair identification

Here we used one dataset (GDS2213) for gene expression of HepG2 cells treated by 5-aza-2'-deoxycytidine (5-aza-dC), trichostatin A (TSA), or both, as an example of differential analysis-based functional TF pair identification. The gene expression profiles under three kinds of treatments were compared with those in the control group, respectively. As a result, 24 TF pairs, involving 8 TFs, were identified to be functionally important under the treatment of 5-aza-dC, which inhibits DNA methylation. Consistently, most TFs have been reported to be influenced by DNA methylation. More interestingly, HDAC2, an important histone deacetylase, joined the functional TF network when TSA, an inhibitor of histone decetylation, was added.

```
# Download GEO dataset. The experiment design will be
# printed.
dataMatrix <- prepareGeoData(GEO = "GDS2213")
# Choose the TF pairs in HEPG2 cells
tfPairsUsed <- pairs2TargetRemDup[which(pairs2TargetRemDup[,
    3] == "HEPG2"), ]
# Differential analysis based functional TF pairs
# identification
enrichTargetResult <- differentialAnalysis(dataMatrix, groups = c(rep("untreated",
    4), rep("aza", 4), rep("TSA", 4), rep("azaAndTSA", 4)),
    contrasts = c("aza-untreated", "TSA-untreated", "azaAndTSA-untreated"),
    pairs2Target = tfPairsUsed)
# Network visualization for aza vs untreated
enrichTargetResultNetwork1 <- networkVis(enrichTargetResult[[1]])
```




```
# Network visualization for azaAndTSA vs untreated
enrichTargetResultNetwork3 <- networkVis(enrichTargetResult[[3]])
```




Here the size of nodes indicated the importance of the TF in the cooprative network, while the width of the edges between pairs indicated the association between two TFs, which is the negative log 10 p value for the enrichement of shared targets in significantly changed genes.

# 5 Example of correlation analysis-based functional TF pairs identification

You can also use the *correlationAnalysis* function to perform correlation analysis-based functional TF pair identification. The example codes can be viewed with the following R codes.

```
example(correlationAnalysis, run.dontrun = TRUE)
```