SpidermiR:Application Examples

Claudia Cava, Antonio Colaprico, Alex Graudenzi, Gloria Bertoli,Tiago C. Silva,Catharina Olsen,Houtan Noushmehr, Gianluca Bontempi, Giancarlo Mauri, Isabella Castiglioni

2016-11-03

Introduction

In this vignette, we demonstrate some applications of `SpidermiR` as tool for the study of miRNA network. For basic use of the `SpidermiR` package, please refer to the vignette `Working with SpidermiR package`.

SpidermiR Downstream Analysis: Case Studies

Case Study n.1: Role of miRNAs in shared protein domains network in Prostate Cancer

In this case study, we downloaded shared protein domains network in Homo Sapiens, using SpidermiRquery, SpidermiRprepare, and SpidermiRdownload with the function `Case_Study1_loading_1_network`.
This function downloads the shared proteind network in HomoSapiens as provided by GeneMania. Then preprocessing of the network give us the measures of the network.

```
Case_Study1_loading_1_network<-function(species){
org<-SpidermiRquery_species(species)
net_shar_prot<-SpidermiRquery_spec_networks(organismID = org[6,],
                                            network = "SHpd")
out_net<-SpidermiRdownload_net(net_shar_prot)
geneSymb_net<-SpidermiRprepare_NET(organismID = org[6,],data = out_net)
ds<-do.call("rbind", geneSymb_net)
data2<-as.data.frame(ds[!duplicated(ds), ])
m<-c(data2$gene_symbolA)
m2<-c(data2$gene_symbolB)
s<-c(m,m2)
fr<- unique(s)
network = "SHpd"
print(paste("Downloading of 1 ",network, " network ",
            "in ",org[6,]," with number of nodes: ",
            length(fr)," and number of edges: ",nrow(data2),
            sep = ""))
return(geneSymb_net)
}
```

Then, we focused on role of miRNAs in this network. We integrated miRNA information using SpidermiRanalyze in the fucntion `Case_Study1_loading_2_network`.

```
Case_Study1_loading_2_network<-function(data){
miRNA_complNET<-SpidermiRanalyze_mirna_gene_complnet(data,
                                                     disease="prostate cancer",
                                                     miR_trg="val")
m2<-c(miRNA_complNET$V1)
m3<-c(miRNA_complNET$V2)
```

```
s2<-c(m2,m3)
fr2<- as.data.frame(unique(s2))
print(paste("Downloading of 2 network with the
            integration of miRNA-gene-gene interaction with number of nodes ",
            nrow(fr2)," and number of edges ", nrow(miRNA_complNET), sep = ""))
return(miRNA_complNET)
}
```

In order to understand the underlying biological process of a set of biomarkers of interest (e.g. from differentially expressed genes, DEGs) we performed an analysis to identify the DEGs connected between them in the shared protein domains network.

```
Case_Study1_loading_3_network<-function(data,dataFilt,dataClin){
highstage <- dataClin[grep("7|8|9|10", dataClin$gleason_score), ]
highstage<-highstage[,c("bcr_patient_barcode","gleason_score")]
highstage<-t(highstage)
samples_hight<-highstage[1,2:ncol(highstage)]
dataSmTP <- TCGAquery_SampleTypes(barcode = colnames(dataFilt),
                                  typesample = "TP")
dataSmNT <- TCGAquery_SampleTypes(barcode = colnames(dataFilt),
                                  typesample ="NT")
colnames(dataFilt)<-substr(colnames(dataFilt),1,12)
se<-substr(dataSmTP, 1, 12)
common<-intersect(colnames(dataFilt),samples_hight)
dataSmNT<-substr(dataSmNT, 1, 12)
sub_net2<-SpidermiRanalyze_DEnetworkTCGA(data,
                                         TCGAmatrix=dataFilt,
                                         tumour=common,normal=dataSmNT)
ft<-sub_net2$V1
ft1<-sub_net2$V2
fgt<-c(ft,ft1)
miRNA_NET<-SpidermiRanalyze_mirna_network(sub_net2,
                                          disease="prostate cancer",miR_trg="val")
TERZA_NET<-rbind(miRNA_NET,sub_net2)
print(paste("In the 3 network we found",length(unique(miRNA_NET$V1)),
            " miRNAs and ",
            length(unique(fgt)), " genes with ", nrow(TERZA_NET),
            " edges " ))
return(TERZA_NET)
}
```

The function `Case_Study1_loading_4_network` is able to reveal the communites based on density metrics. We focused on the community with the higher number of elements.

```
Case_Study1_loading_4_network<-function(TERZA_NET){
comm<-  SpidermiRanalyze_Community_detection(data=TERZA_NET,type="FC")
#SpidermiRvisualize_mirnanet(TERZA_NET)
cd_net<-SpidermiRanalyze_Community_detection_net(data=TERZA_NET,
                                                 comm_det=comm,size=5)
ft<-cd_net$V1
ft1<-cd_net$V2
fgt<-c(ft,ft1)
print(paste("In the 4 network we found",length(unique(fgt)),
            " nodes and ", nrow(cd_net), " edges " ))
return(cd_net)
```

```
}
```

Case Study n.2: miRNAs regulating degree centrality genes in physical interactions network in breast cancer

In this case study, we downloaded physical interactions network in Homo Sapiens, using SpidermiRquery, SpidermiRprepare, and SpidermiRdownload with the function `Case_Study2_loading_1_network`.
This function downloads the physical interactions network in HomoSapiens as provided by GeneMania. Then preprocessing the network give us the measures of the network.

```
Case_Study2_loading_1_network<-function(species){
org<-SpidermiRquery_species(species)
net_PHint<-SpidermiRquery_spec_networks(organismID = org[6,],
                                        network = "PHint")
out_net<-SpidermiRdownload_net(net_PHint)
geneSymb_net<-SpidermiRprepare_NET(organismID = org[6,],data = out_net)
ds<-do.call("rbind", geneSymb_net)
data1<-as.data.frame(ds[!duplicated(ds), ])
sdas<-cbind(data1$gene_symbolA,data1$gene_symbolB)
sdas<-as.data.frame(sdas[!duplicated(sdas), ])
m<-c(data1$gene_symbolA)
m2<-c(data1$gene_symbolB)
s<-c(m,m2)
fr<- unique(s)
network="PHint"
print(paste("Downloading of 1 ",network,
            " network ","in ",org[6,],
            " with number of nodes: ",length(fr),
            " and number of edges: ",nrow(sdas), sep = ""))
return(geneSymb_net)
}
```

A network of miRNA-protein PI was found using `Case_Study2_loading_2_network`.

```
Case_Study2_loading_2_network<-function(data){
miRNA_NET<-SpidermiRanalyze_mirna_network(data,
                                          disease="breast cancer",miR_trg="val")
m2<-c(miRNA_NET$V1)
m3<-c(miRNA_NET$V2)
s2<-c(m2,m3)
fr2<- as.data.frame(unique(s2))
print(paste("Downloading of 2 network with the integration of
            miRNA-gene interaction with number of nodes ", nrow(fr2),"
            and number of edges ", nrow(miRNA_NET), sep = ""))
return(miRNA_NET)
}
```

Statistical results showed that proteins with higher centrality are effectively targets of miRNAs with higher centrality.

```
Case_Study2_loading_3_network<-function(sdas,miRNA_NET){
ds<-do.call("rbind", sdas)
  data1<-as.data.frame(ds[!duplicated(ds), ])
  sdas<-cbind(data1$gene_symbolA,data1$gene_symbolB)
```

```
  sdas<-as.data.frame(sdas[!duplicated(sdas), ])
topwhol<-SpidermiRanalyze_degree_centrality(sdas)
topwhol_mirna<-SpidermiRanalyze_degree_centrality(miRNA_NET)

miRNA_degree<-topwhol_mirna[grep("hsa",topwhol_mirna$dfer),]
seq_gd<-as.data.frame(seq(1, 15400, by = 50))
even<-seq_gd[c(F,T),]
even2<-even
odd<-seq_gd[c(T,F),]
odd2<-odd[-1]
odd2[154]<-15400
f<-cbind(even2,odd2-1)

SQ<-cbind(odd,even-1)

h<-as.data.frame(rbind(f,SQ))
SQ <- as.data.frame(h[order(h$even2,decreasing=FALSE),])

table_pathway_enriched <- matrix(1, nrow(SQ),4)
colnames(table_pathway_enriched) <- c("interval min",
                                      "interval max","gene","miRNA")
table_pathway_enriched <- as.data.frame(table_pathway_enriched)

j=1
for (j in 1:nrow(SQ)){
  a<-SQ$even2[j]
  b<-SQ$V2[j]
  d<-c(a,b)
gene_degree10<-topwhol[a:b,]
vfg<-rbind(miRNA_degree[1:10,],gene_degree10)
subnet<-SpidermiRanalyze_direct_subnetwork(data=miRNA_NET,BI=vfg$dfer)

table_pathway_enriched[j,"interval min"] <- d[1]
table_pathway_enriched[j,"interval max"] <- d[2]
s<-unique(subnet$V1)
x<-unique(subnet$V2)
table_pathway_enriched[j,"miRNA"]<-length(s)
table_pathway_enriched[j,"gene"]<-length(x)
}

df<-cbind(table_pathway_enriched$gene,table_pathway_enriched$miRNA)
rownames(df)<-table_pathway_enriched$`interval max`
categories <- c("protein", "miRNA")
colors <- c("green", "magenta")
op <- par(mar = c(5, 5, 4, 2) + 0.1)
matplot(df, type="l",col=colors,xlab = "N of Clusters",
        main = "",ylab = "Interactions",cex.axis=2,cex.lab=2,cex.main=2)
legend("topright", col=colors, categories, bg="white", lwd=1,cex=2)
j=1
a<-SQ$even2[j]
b<-SQ$V2[j]
d<-c(a,b)
gene_degree10<-topwhol[a:b,]
```

```
vfg<-rbind(miRNA_degree[1:10,],gene_degree10)
subnet<-SpidermiRanalyze_direct_subnetwork(data=miRNA_NET,BI=vfg$dfer)
m2<-c(subnet$V1)
m3<-c(subnet$V2)
s2<-c(m2,m3)
fr2<- as.data.frame(unique(s2))
print(paste("Downloading of 3 network with proteins and miRNAs
with highest degree
            centrality with  ", nrow(fr2)," nodes and number of
edges ", nrow(subnet), sep = ""))
return(subnet)
}
```

References

1. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems. 2006;1695(5), 1-9.
2. Cohen-Gihon I, Nussinov R, Sharan R. Comprehensive analysis of co-occurring domain sets in yeast proteins. BMC Genomics. 2007 Jun 11;8:161.
3. Hegyi H, Gerstein M: Annotation transfer for genomics: measuring functional divergence in multi-domain proteins. Genome Res 2001, 11:1632-1640.