Interpretation

Dick de Ridder

6/10/2018

Before you begin

Yesterday, you saved one of the WGCNA modules to a file. The most widely used package to visualise and inspect biological networks is Cytoscape. Download the latest version from http://www.cytoscape.org/. While you are waiting for the download to finish, please continue with the sections below.

Enrichment

In this first exercise, you will continue to work with the ST vs. HT dataset you clustered yesterday, in particular with the clusters you stored (if you did not manage, you can find sample files `hc.txt` and `km.txt` at http://www.bioinformatics.nl/courses/RNAseq"). R has some functions available for calculating Gene Ontology enrichments, but these mostly only carry data for humans and related model organisms. For these exercises, you will therefore use web services to interpret your cluster(s).

AgriGO

- Visit the AgriGO website at http://bioinfo.cau.edu.cn/agriGO/ and select the "Analysis tool" tab (note: for model organisms such as Arabidopsis thaliana, GOrilla (http://cbl-gorilla.cs.technion.ac.il/) is a good alternative).
- Under 1., select "Singular enrichment analysis"; under 2. the appropriate organism; and keep 3. at "Suggested backgrounds".
- Have a look at the options under 4. (but don't change anything).
- Open one of your saved clusters (`hc.txt` or `km.txt`) in a text editor, copy all gene names and paste them into the Query list box under 2.
- Press "Submit" at the bottom and wait.
- On the output page, inspect the results and generate some images. How many significant annotations do you find? Do they correspond to what you expected based on the biology?

PlantGSEA

- Visit the PlantGSEA website at http://structuralbiology.cau.edu.cn/PlantGSEA/ and select "Analysis".
- Select a few gene sets (e.g. BP, CC and PlantCyc) and choose the appropriate species.

- Open one of your saved clusters (`hc.txt` or `km.txt`) in a text editor, copy all gene names and paste them into the Query box.
- Press "Start analysis" at the bottom and wait.
- On the output page, inspect the results. Again, does what you find correspond to what you expected based on the biology? And to what you found earlier?

Network-based analysis

For this part, we will focus on a single gene (AT5G10140). Both STRING and GeneMania also allow you to upload lists of genes, but the resulting networks quickly become rather complex.

STRING

- Visit http://string-db.org/, select "Protein by name" and enter "AT5G10140" as the gene to search for. Then press "Search" and on the next screen press "Continue" at the top.
- Click on the FLC node to learn more about the protein. What does it do?
- Inspect the network around FLC. How many functional interaction partners are predicted? Do they make sense if you look at their descriptions below? What evidence was used to find these partners? Click on "Legend" below the figure to learn about the colors.
- Try a few different edge representations under the "Settings" tab, e.g. "confidence" and "molecular action", press "Update" and try to interpret the results.
- Go back to the "evidence" view. From the "Viewers" tab, select the different sources of evidence. Can you explain what you see?
- Go back to the "Network" view under "Viewers" and press the "+ More" button below the network. What happens? Does that make sense?
- *Optional*: see whether the network you found is enriched for some GO functions or KEGG pathways, under the "Analysis" tab.

GeneMania

- In a new browser window, visit http://genemania.org/ and search for "FLC" (in Arabidopsis thaliana, click on the icon next to the search box at the top left to select it).
- Inspect the resulting network. Is it the same as the one you found using STRING? Why do you think this is the case? Select the "Functions" icon (bottom left). Does the list make sense given what you know about FLC?

Optional

Try out AraNet (http://www.functionalnet.org/aranet/) and TF2Network (http://bioinformatics.psb.ugent.be/webtools/TF2Network/) for additional tools to interpret gene lists. The first is a network of functional interactions in Arabidopsis; the

second looks for transcription factors whose binding sites are overrepresented in the promoters of the gene list you specify.

Visualising networks

Cytoscape should have downloaded by now. Install it on the `D:` drive. When running it for the first time, don't allow network access. Note that Cytoscape is a complex program with many options and can be hard to get used to; if you get stuck, always make sure you work with the correct network (you can select networks in the `Network` tab in the `Control Panel`) or, if all else fails, restart Cytoscape.

- First, from the menu, select `Apps->App Manager`, search for "bingo" and install the BiNGO app. Likewise, install the ClusterONE app.
- Import the WGCNA module you saved yesterday (or the version on http://www.bioinformatics.nl/courses/RNAseq): select `File->Import->Network->File...` and open the file `wgcna_edges.txt`. You will need to click in the column headings (`fromNode`, `toNode`) to select which columns contain the start and end nodes of each edge. For `fromNode`, select the green circle; for `toNode`, the red target. Then press `OK`.
- Choose `Layout->Perfuse Force Directed Layout->weight` to get a different visualisation of the network.
- Select all nodes in the network (`Ctrl-A`), then choose `Apps->BiNGO` to run the BiNGO app.
- Enter a short description under "Cluster name" and under "Select organism/annotation:" choose Arabidopsis.
- Press `BiNGO` and have a look at the results. What functions/processes are enriched in this module? Note that the enriched parts of the Gene Ontology are displayed as a network itself; to return to the original WGCNA module, select it under `Network` in the `Control Panel`.

Inferring regulation networks

You have already used WGCNA to find modules in correlation networks, but R has some more advanced packages to infer gene regulation networks. In this final exercise, you will construct a network on the ST vs. HT data using the `minet` package and mine it in Cytoscape.

First, install the package:

```
source("https://bioconductor.org/biocLite.R")
biocLite("minet")

library(minet)
library(WGCNA)
```

Then load the ST vs. HT data:

```
counts =
read.table("http://www.bioinformatics.nl/courses/RNAseq/ST_vs_HT.csv",
sep=",", row.names=1, header=TRUE, stringsAsFactors=FALSE)

# Select 500 interesting genes
ngenes=500
m = apply(counts,1,mad)
ind = order(m,decreasing = TRUE,na.last = NA)
data = counts[ind[1:ngenes],]

# Infer a network using the CLR method
net <- minet(t(data),method="clr")

# Calculate log2(differential expression)
expr_st <- as.matrix(log2(rowMeans(data[,seq(1,3)])));
expr_ht <- as.matrix(log2(rowMeans(data[,seq(4,6)])));
diff_expr <- expr_st - expr_ht;

# Save the network to a file
exportNetworkToCytoscape(net, edgeFile='clr_edges.txt',
                         nodeFile='clr_nodes.txt',
                         weighted=TRUE, threshold = 0.02,
                         altNodeNames=NULL,
                         nodeNames=rownames(net),
                         nodeAttr=diff_expr);
```

Now go to Cytoscape and import the network:

- Use `File->Import->Network->File...` to open the file `clr_edges.txt` and select the start and end nodes as above. Once loaded, how many edges does the network contain? Given that there are 500 genes, would you consider that a lot?
- Select `Layout->Perfuse Force Directed Layout->weights`.
- Add the gene expression values to the nodes: use `File->Import->Table->File...` to read in the file `clr_nodes.txt`. This will add the differential expression as node annotations.
- Color the nodes by this differential gene expression: under `Control Panel`, find the `Style` button, click on `Fill color`, select `nodeAttr.nodesPresent...` as `Column` and `Continuous Mapping` as `Mapping Type`.
- Double click the color map, and click on the triangles pointing down to select two contrasting colors (e.g. red and green) for low and high expression values, respectively. Finally, press `OK`. What do you notice in the network?
- Now you can use Cytoscape to find clusters in the network: run `Apps->ClusterONE`, set `Minimum size` to 10 and press `Start`.
- In the results panel (on the right, containing small red cluster representations), select one of the clusters. This should select the corresponding nodes in the main network.

Next, choose `Apps->BINGO` (as above) to learn about the annotations enriched in this module. Repeat this for some other clusters. Do your findings make sense?

Optional

Repeat the analysis above, using the ARACNE inference method:

```
net <- minet(t(data),method="aracne")
```

Do you find a similar network, and similar modules?

Optional

Infer a network on a set of selected genes in the Arabidopsis apical meristem timeseries data (see the Clustering practical) and inspect it in Cytoscape.