msarc : Graphical (mis)Representation of Mass-Spec Results

Gord Brown

gordon.brown@cruk.cam.ac.uk

Cancer Research UK

Cambridge Institute

2013 April 8

1 Introduction

This package generates circular plots showing the results of mass spectrometry experiments. GO categories are shown as arcs around the ouside; proteins that were identified in the experiment are drawn as radial lines, with the length of the line indicating the MASCOT score (longer lines are more certain results). See Figure 1 for an example.

Processing is done in several steps:

1. Load the mass spec data, either from a tab-separated spreadsheet generated by the mass spec software, or from a 3-column data frame: UniProt ID, gene symbol, and score (in that order).

2. Subtract the proteins that were also identified in the control sample.

3. Look up the GO terms associated with each protein, and their ancestors.

4. Choose the GO terms you want in the diagram. This step must be done manually, either by writing the list to disk, deleting the unwanted ones from the file, and loading it back, or by getting the terms in a data frame, making a list of the ones to keep, and passing that list to the `msarc` object.

5. Generate the plot. Because GO category names get longer as the category gets smaller (i.e. more specific), the names of sub-categories are not drawn on the plot; instead single letters are drawn, and "legend" is printed on the R console. You can manually edit the diagram to include whatever names you like.

6. If you have 3 or more experiments, you can also generate a heat map of their similarity, based on the sets of UniProt IDs, or on their MASCOT scores.

2 Example: Generating the Plot from Data Frames

```
> library(msarc)
> # load data
> data(sample_df,package="msarc")
> data(control_df,package="msarc")
> # create the msarc objects
> sample <- msarc(sample_df)
> control <- msarc(control_df)
> # subtract the control (i.e. remove UniProt IDs from sample that are also in
> # control)
> sample <- msarc.subtract(sample,control)
> # generate the list of GO terms
> sample <- msarc.findGOterms(sample)
> # get the list as a data frame
> term_df <- msarc.getTerms(sample)
> term_df <- term_df[c("GO:0017076","GO:0030554","GO:0032553","GO:0032555",
+                      "GO:0032559","GO:0035639","GO:0036094","GO:0097159",
+                      "GO:0017111","GO:0016462"),]
> sample <- msarc.filterTerms(sample,term_df)
> # generate the plot
> msarc.plotSVG(sample,file="thing.svg")

msarc:  proteins=40 (100 pre-subtraction)
  source=data frame
  control=data frame
  GO terms: 10 (pre-filtering: 103 direct, 114 by ancestral links)
```

3 Acknowledgements

This package was developed at Cancer Research UK's Cambridge Institute, using data generated by Hisham Mohammed.

This package was inspired in part by CircOS, though the implementation is not in any way dependent on or derived from it.?

4 Setup

This vignette was built on:
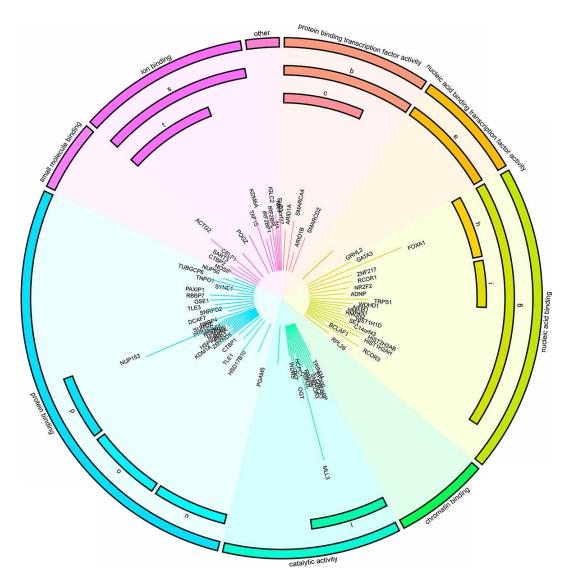
```
> sessionInfo()
```

2

Figure 1: A sample plot generated by `msarc`.

```
R version 3.0.0 (2013-04-03)
Platform: x86_64-unknown-linux-gnu (64-bit)

locale:
[1] C

attached base packages:
[1] parallel  stats     graphics  grDevices utils     datasets  methods
[8] base

other attached packages:
[1] msarc_1.2.0        AnnotationDbi_1.22.1 Biobase_2.20.0
[4] BiocGenerics_0.6.0  RSQLite_0.11.2       DBI_0.2-5

loaded via a namespace (and not attached):
[1] GO.db_2.9.0        IRanges_1.18.0      gdata_2.12.0.2      gplots_2.11.0.1
[5] gtools_2.7.1       org.Hs.eg.db_2.9.0 org.Mm.eg.db_2.9.0 stats4_3.0.0
[9] tools_3.0.0
```