

EMBO Practical Course on Analysis and Informatics of Microarray Data

Exercise sheet for BIIT tools 20.10.2010

MEM (<http://biit.cs.ut.ee/mem>)

- 1) Input Your favourite gene (e.g. Nanog).
- 2) Pick one probeset to start with.
- 3) Use organism filter to find Affymetrix platforms for Your specie (24 species listed)
- 4) Start with one platform, e.g U133 for human or 430 for mouse
- 5) Apply dataset filters to standard deviation (default is 0.29). Increasing the threshold results in lesser number of datasets used and faster output.
- 6) Try out using “*number of most variant datasets*” from dataset filters by using only 20 datasets where the query probe varies the most
- 7) Enable *show cell tooltips* from *Output* options to see more information on the output. Be patient for the information boxes to load.
- 8) Try out cloud tags of datasets' annotations (mouse-hovering on the dendrogram)
- 9) Using tag clouds select a subcluster of the datasets by clicking on a (tiny) gray square (after the page has fully loaded) on the dendrogram and go and check out expression profiles using *ExpressView* link. Be aware that selecting many datasets makes *ExpressView* calculations very slow so choose up to 30 datasets.
- 10) Cluster the genes by their profiles by enabling the option from *Output*
- 11) Input a set of genes to compare their similarity to Your one query gene using *Gene filters* option. Use for example Your other favourite genes or genes belonging to some protein complex or pathway
- 12) Send output to g:Profiler for functional profiling using *GO annotations* link
- 13) Make a query of a gene from mouse and human platforms. Compare the output gene lists and their respective g:Profiler results.

g:Profiler (<http://biit.cs.ut.ee/gprofiler>)

- 14) If You have Your favourite dataset in hand then pick most highly expressed genes and use them as input
- 15) Do the same as in previous exercise but this time also select *ordrered list* query option. Do You see a difference?
- 16) If You have certain chromosome location of interest then input it to g:Profiler to see what genes are located there and if some GO annotation, pathway or regulatory motif comes up as significant
- 17) If You have several experiments that result in a gene list (e.g. microarrays) then use the most highly/lowly expressed genes as input in g:cocoa that runs the same g:profiler analysis on several gene lists simultaneously. Find out what functions are constantly significant and which ones change from experiment to experiment.
- 18) Familiarise Yourself with ortholog mapping tool g:Orth for converting human genes to mouse for example and name conversion tool g:Convert to get from gene names to Affymetrix IDs to protein names.

KEGGAnim (<http://biit.cs.ut.ee/kegganim>)

- 19) Either upload Your own expression dataset by creating a folder or pick one of the existing datasets
- 20) Choose a pathway of interest by Your previous knowledge or select a pathway with many mapping probes. Start Animation. Look how much and which genes change their expression.
- 21) Find a gene that has more than one probe and look if all of the probes express at the same level.
- 22) Create a picture with most interesting datapoints for presentation using *CineFilm*, rename the Condition titles and make a figure for Your next presentation
- 23) Try out *Pathway components* selection menu to animate only a subsection of probesets/genes.

GraphWeb (<http://biit.cs.ut.ee/graphweb>)

- 24) If You known connections between Your genes then use this network together with public protein-protein interaction (PPI) data (*From a file in our server*) and find gene pairs being present in both networks by applying *Remove edges with less than N labels*
- 25) Filter PPI networks with Your genes of interest by using *Network neighbourhood*, keep the distance 1.
- 26) Look for modules in Your data using different network algorithms and filter out different module sizes. Check the functional annotations of Your modules.

VisHiC (<http://biit.cs.ut.ee/vishic>)

- 27) Use example datasets provided and limit the cluster sizes to minimum 20 genes and maximum 100 genes. With these options You get more reasonable output to play with.
- 28) Search for Your favourite gene(s) from the clusters using search box above the dendogram (look for red brackets to appear on dendogram)
- 29) Look for cluster having unique annotations (marked with asterix in the first column of the *Cluster statistics* table)
- 30) Adjust the *cluster size* limits and *additional threshold* to make the output even more compact

Those of You who work with human or mouse embryonic stem cells can play around in Embryonic Stem Cells Database (ESCDb at <http://biit.cs.ut.ee/escd>).