



Short Application Session 6:

Hands-on GSEA training for gene expression and proteomic data

Yana Stackpole, PhD (Application Specialist)

If you want to teach people a new way of thinking... give them a tool, the use of which will lead to new ways of thinking (Richard Buckminster Fuller)

Big Data –a lot of data, faster and at a lower cost

Biological interpretation – the biggest challenge:

Looking for a Needle in a stack of similar Needles



DEMO

What it feels like to do GSEA in Qlucore

GSEA – built to address limitations of single-gene approach

- A method to infer biological pathway activity from gene expression data
- GSEA evaluates data at the level of gene sets / pathways
- The gene sets are defined based on prior biological knowledge, e.g., published information about biochemical pathways or coexpression in previous experiments
- The goal of GSEA is to determine whether members of a gene set (pathway) are over- or underrepresented in your data

Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles

Aravind Subramanian^{a,b}, Pablo Tamayo^{a,b}, Vamsi K. Mootha^{a,c}, Sayan Mukherjee^d, Benjamin L. Ebert^{a,e}, Michael A. Gillette^{a,f}, Amanda Paulovich^g, Scott L. Pomeroy^h, Todd R. Golub^{a,e}, Eric S. Lander^{a,c,i,j,k}, and Jill P. Mesirov^{a,k}

^aBroad Institute of Massachusetts Institute of Technology and Harvard, 320 Charles Street, Cambridge, MA 02141; ^cDepartment of Systems Biology, Alpert 536, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02446; ^dInstitute for Genome Sciences and Policy, Center for Interdisciplinary Engineering, Medicine, and Applied Sciences, Duke University, 101 Science Drive, Durham, NC 27708; ^eDepartment of Medical Oncology, Dana–Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; ^fDivision of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114; ^gFred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, C2-023, P.O. Box 19024, Seattle, WA 98109-1024; ^hDepartment of Neurology, Enders 260, Children’s Hospital, Harvard Medical School, 300 Longwood Avenue, Boston, MA 02115; ⁱDepartment of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142; and ^jWhitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, MA 02142

Contributed by Eric S. Lander, August 2, 2005

Although genome-wide RNA expression analysis has become a standard tool, it does not evaluate microarray data at the level of gene sets. The gene sets are

GSEA – applications

- **Gene expression** data (arrays, RNAseq)
- **Proteomics** data with gene symbol annotation, or looking for a protein set enrichment with your defined protein sets
- **Metabolomics** data – it will function as MSEA with your metabolite sets

GSEA – overview

- GSEA considers profiles from samples belonging to two (or more) classes so a group comparison can be made
- All genes are ranked based on the correlation between their expression and specified phenotype(=your sample annotation) Top to Bottom
- The goal of GSEA is to determine whether the genes of the gene set are randomly distributed or enriched on the Top or Bottom of the list

GSEA implementation in QOE

Hands-on experience

GSEA hands-on – run the analysis

1. Define your dataset input. Make sure:

- You are using a complete data set and not only significant genes (clear variable cut off if any applied – **p, q, fold-change, variance, gene lists as filters**);
- Use the same gene annotation for dataset as in your gene sets Go to **Qlucore - Data – Variable ID** – drop down menu;
- Your gene symbols (if using sets in gene symbols) must be unique, if not - collapse them on Gene Symbol in **Qlucore - Data – Variable ID – Collapse** – drop down

GSEA hands-on – Gene sets

Find and point GSEA to your gene sets Folder

1. You can use **single pathways** (like two in the shared folder: GSEA_HALLMARK_IL6_JAK_STAT3_SIGNALING and HALLMARK_P53_PATHWAY (confirmatory or biomarker analysis))
2. Use **curated public gene sets** in gmt or gmt.gz: [MSigDB](#) for human (Broad Institute), [GSKB](#) (mouse analog of human MSigDB) – discovery analysis. My shared folder has sets from both.
3. **Build your own sets** (you can open gmt files with excel or change them into .txt then open to see the layout, can be in txt also)
4. Point to the Folder containing your sets from Qlucore GSEA or move them into Qlucore folder with GSEA sets

GSEA hands-on – run the analysis

1. Open GSEA on Qlucore (GSEA button under Data tab in the central panel)
2. Select single gene set(s) or a whole gene set collection
3. Select appropriate metric/test and the sample annotation with comparison groups
4. Click RUN

GSEA hands-on – Understanding your results

Review the List of gene sets found to be enriched

Review the annotations – number of genes, number of matches, ES, NES, p , q

Apply a q cut off and sort by q

OR

Explore Leading edge subsets

Export the results – highlight, and save (all folders, and Lists)

Vizualize and work with the most interesting in main interface (after using "Export Lists" they appear in Variable space)

GSEA hands-on – Vizualize your results

Vizualize and work with the most interesting in main interface (after using "Export Lists" they appear in Variable space)

Synchronize plots like PCA - Heatmap side by side to view pathway genes in relation to your groups

Use lists as filters

Use lists to color plots (PCA, heatmap)

Copy gene lists and keep only Leading edge genes – see your core enrichment genes

GSEA – what to expect

- **Shift from single gene thinking to functional** units (processes, pathways etc)
- **Detect gene sets correlated with the phenotypes**
- **High resolution** - detect important gene sets in a dataset with weak signals
- **High consistency or results between independently derived data sets** (even when single-gene approach gives small overlap in DE genes)
- **Generate stronger more compelling hypotheses** for further

Make your quantum leap!

Supported TRIAL access @ Qlucore.com - download, contact us!

Yana.Stackpole@glucore.com (Boston-based)

support@glucore.com (global team in Lund, Sweden)