Introduction to Bioinformatics

Continuation of GSEA

2021.

Gene Set Enrichment Analysis

- Gene set enrichment analysis (GSEA) is a statistical method to determine if predefined sets of genes are differentially expressed in different phenotypes

Gene Set Enrichment Analysis (GSEA) for Interpreting Gene Expression Profiles. Jing Shi et al.

Gene Set Enrichment Analysis

- Gene set enrichment analysis (GSEA) is a statistical method to determine if predefined sets of genes are differentially expressed in different phenotypes
- Predefined gene sets may be genes in a known metabolic pathway, located in the same cytogenetic band, sharing the same Gene Ontology category, or any user-defined set

Gene Set Enrichment Analysis (GSEA) for Interpreting Gene Expression Profiles. Jing Shi et al.

Gene Set Enrichment Analysis

- Gene set enrichment analysis (GSEA) is a statistical method to determine if predefined sets of genes are differentially expressed in different phenotypes
- Predefined gene sets may be genes in a known metabolic pathway, located in the same cytogenetic band, sharing the same Gene Ontology category, or any user-defined set
- In array experiments where no single gene shows statistically significant differential expression between phenotypes, GSEA has identified significant differentially expressed sets of genes

Gene Set Enrichment Analysis (GSEA) for Interpreting Gene Expression Profiles. Jing Shi et al.

Gene Set Enrichment Analysis

- Gene set enrichment analysis (GSEA) is a statistical method to determine if predefined sets of genes are differentially expressed in different phenotypes
- Predefined gene sets may be genes in a known metabolic pathway, located in the same cytogenetic band, sharing the same Gene Ontology category, or any user-defined set
- In array experiments where no single gene shows statistically significant differential expression between phenotypes, GSEA has identified significant differentially expressed sets of genes
- GSEA is likely to be more powerful than conventional single-gene methods for studying the large number of common diseases in which many genes each make subtle contributions

Gene Set Enrichment Analysis (GSEA) for Interpreting Gene Expression Profiles. Jing Shi et al.

Why GSEA?

- The conventional statistical analysis method for array experiments is to
  - examine one gene at a time,
  - determine a p-value that the gene is differentially expressed/methylated in different phenotypes
  - apply a correction (penalty) to the p-value for having tested multiple genes (described further below)

Gene Set Enrichment Analysis (GSEA) for Interpreting Gene Expression Profiles. Jing Shi et al.

Why GSEA?

- The conventional statistical analysis method for array experiments is to
  - examine one gene at a time,
  - determine a p-value that the gene is differentially expressed/methylated in different phenotypes
  - apply a correction (penalty) to the p-value for having tested multiple genes (described further below)

- This method had limitations
  - One such that; In common diseases, a large number of genes each make subtle contributions, and these genes are difficult to detect in single gene analyses.

Gene Set Enrichment Analysis (GSEA) for Interpreting Gene Expression Profiles. Jing Shi et al.

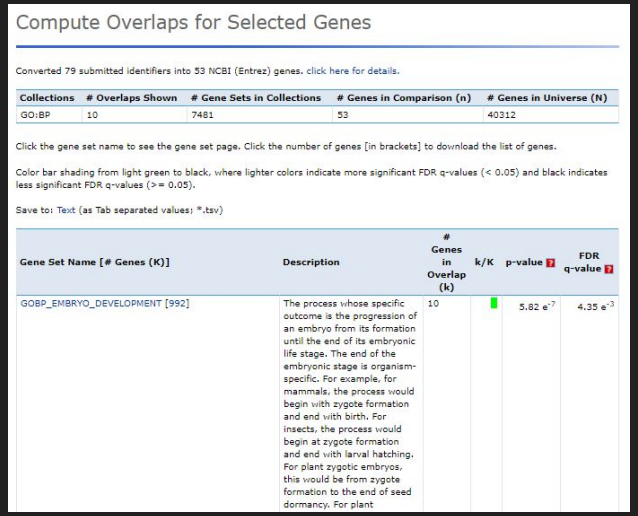
Why GSEA?

- The conventional statistical analysis method for array experiments is to
  - examine one gene at a time,
  - determine a p-value that the gene is differentially expressed/methylated in different phenotypes
  - apply a correction (penalty) to the p-value for having tested multiple genes (described further below)

- This method had limitations
  - One such that; In common diseases, a large number of genes each make subtle contributions, and these genes are difficult to detect in single gene analyses.

Gene Set Enrichment Analysis (GSEA) for Interpreting Gene Expression Profiles. Jing Shi et al.

We already did it previously, what now?



Compute Overlaps for Selected Genes

Converted 79 submitted identifiers into 53 NCBI (Entrez) genes. click here for details.

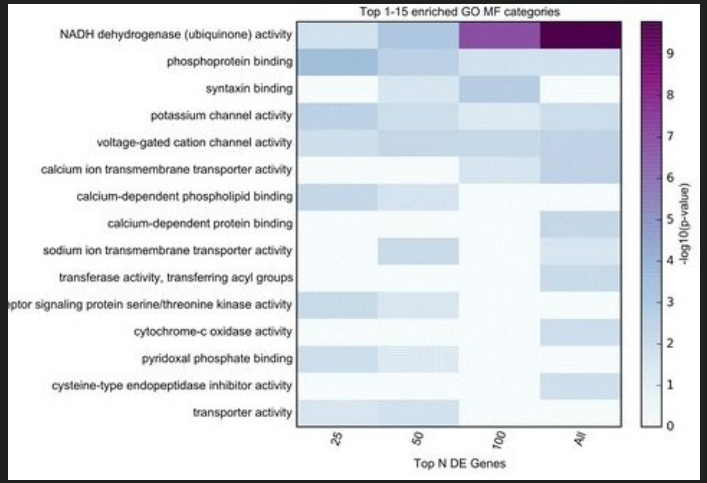
Collections	# Overlaps Shown	# Gene Sets in Collections	# Genes in Comparison (n)	# Genes in Universe (N)
GO:BP	10	7481	53	40312

Click the gene set name to see the gene set page. Click the number of genes [in brackets] to download the list of genes.

Color bar shading from light green to black, where lighter colors indicate more significant FDR q-values (< 0.05) and black indicates less significant FDR q-values (>= 0.05).

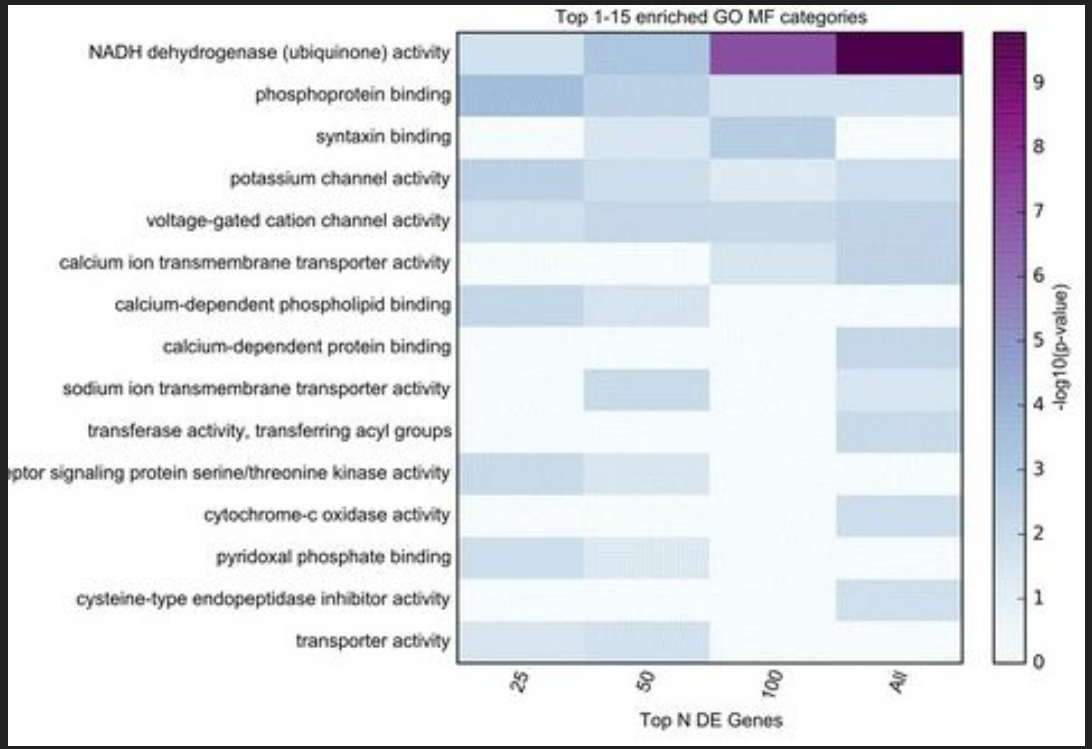
Save to: Text (as Tab separated values: *.tsv)

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p-value	FDR q-value
GOBP_EMBRYO_DEVELOPMENT [992]	The process whose specific outcome is the progression of an embryo from its formation until the end of its embryonic life stage. The end of the embryonic stage is organism-specific. For example, for mammals, the process would begin with zygote formation and end with birth. For insects, the process would begin at zygote formation and end with larval hatching. For plant zygotic embryos, this would be from zygote formation to the end of seed dormancy. For plant	10		5.82 e-7	4.35 e-3



How do we turn this one to this visualization?

Gene Set Enrichment Analysis (GSEA) for Interpreting Gene Expression Profiles. Jing Shi et al.


Top 1-15 enriched GO MF categories
NADH dehydrogenase (ubiquinone) activity
phosphoprotein binding
syntaxin binding
potassium channel activity
voltage-gated cation channel activity
calcium ion transmembrane transporter activity
calcium-dependent phospholipid binding
calcium-dependent protein binding
sodium ion transmembrane transporter activity
transferase activity, transferring acyl groups
ptor signaling protein serine/threonine kinase activity
cytochrome-c oxidase activity
pyridoxal phosphate binding
cysteine-type endopeptidase inhibitor activity
transporter activity
25
50
100
All
Top N DE Genes
-log10(p-value)
0
1
2
3
4
5
6
7
8
9