



Gene Ontology Analysis

...of a selected list of
(differentially expressed) genes



Gene Ontology

Introduction

<http://www.geneontology.org>

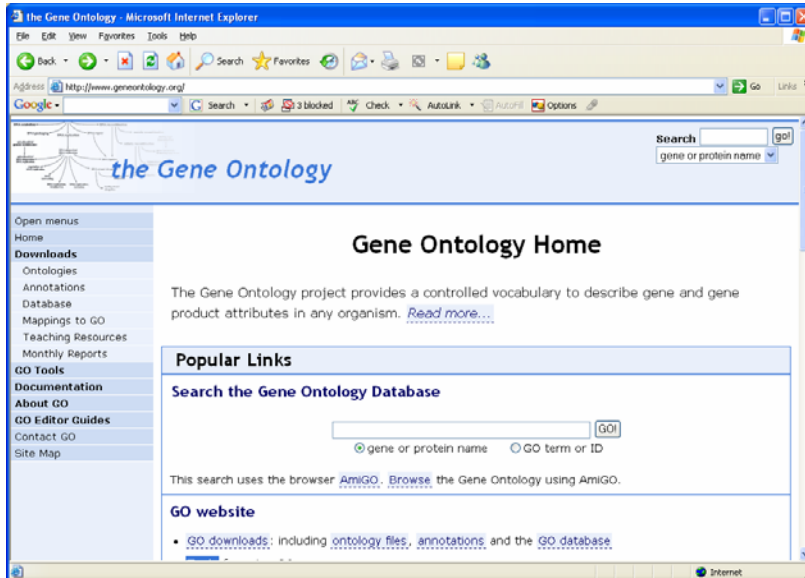
- Searching for all of the available information about each gene of interest is very time consuming. This is hampered further by the wide variations in terminology.
- The Gene Ontology (GO) project is a collaborative effort to address the need for consistent descriptions of gene products in terms of their 'biology' and in a species-independent manner.
- About 12,000 defined concepts, in a DAG (Direct Acyclic Graphs) with two link types (part-of, is-a) under three roots:
 - Cellular component
 - Biological process
 - Molecular function.
- Used as annotations for genes (19,408 terms as of 22/1/06)



Gene Ontology

Introduction

<http://www.geneontology.org>



Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Gene Ontology

Introduction

- Gene Ontology (GO) is a collection of controlled vocabularies describing the biology of a gene product in any organism
- There are 3 independent sets of vocabularies, or ontologies:
 - **Molecular Function (MF)**
 - e.g. "DNA binding" and "catalytic activity"
 - **Cellular Component (CC)**
 - e.g. "organelle membrane" and "cytoskeleton"
 - **Biological Process (BP)**
 - e.g. "DNA replication" and "response to stimulus"

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Introduction

MG1 - Gene Ontology Browser

Molecular Function Gene Ontology Browser
Term Detail

GO term: **L-malate dehydrogenase activity**
Synonym: **NAD-L-malate dehydrogenase**
Synonym: **malic dehydrogenase**
GO id: **GO:0030060**
Definition: **Catalysis of the reaction: (S)-malate + NAD+ = oxaloacetate + NADH + H+. Also oxidizes some other 2-hydroxydicarboxylic acids.**
Number of paths to Term: 2

denotes an 'is-a' relationship
denotes a 'part-of' relationship

Gene_Ontology
molecular_function
enzyme_activity
oxidoreductase_activity
oxidoreductase activity, acting on CH-OH group of donors
malate dehydrogenase activity

D-malate dehydrogenase (decarboxylating) activity
L-malate dehydrogenase activity [GO:0030060] (1 genes, 1 annotations)
malate dehydrogenase (acceptor) activity
malate dehydrogenase (NADP) activity
malic enzyme activity +

Chemical reaction diagram showing the conversion of L-malate to oxaloacetate, catalyzed by NAD+ to NADH + H+.



Introduction

Gene Ontology Browser
Term Detail

Biological Process

GO term: **tricarboxylic acid cycle**
Synonym: **Krebs cycle**
Synonym: **citric acid cycle**
GO id: **GO:0006099**
Number of paths to Term: 2

denotes an 'is-a' relationship
denotes a 'part-of' relationship

Gene_Ontology
biological_process
physiological_processes
metabolism
carbohydrate metabolism
main pathways of carbohydrate metabolism
Entner-Doudoroff pathway
gluconeogenesis +
glycolysis +
glyoxylate cycle
pentose-phosphate shunt +
TCA intermediate metabolism +
tricarboxylic acid cycle [GO:0006099] (21 genes, 21 annotations)
reductive citric acid pathway
reductive tricarboxylic acid cycle

Metabolic pathway diagram showing the tricarboxylic acid cycle (Krebs cycle) and its connections to other metabolic pathways like glycolysis and gluconeogenesis.



Gene Ontology

Introduction

Cellular Component Gene Ontology Browser
Term Detail

GO term: **mitochondrion**
GO id: **GO:0005739**
Definition: **A semiautonomous, self replicating organelle that occurs in varying numbers, shapes, and sizes in the cytoplasm of virtually all eukaryotic cells. It is notably the site of tissue respiration.**
Number of paths to term: **1**

denotes an 'is-a' relationship
 denotes a 'part-of' relationship

Gene Ontology

- cellular_component
 - cell
 - intracellular
 - cytoplasm
 - Golgi apparatus +
 - hydrogenosome +
 - lipid particle +
 - melanosome
 - membrane coat +
 - microbody +
 - microtubule organizing center +
 - mitochondrion [GO:0005739] (504 genes, 675 annotations)

- kinetoplast
- mitochondrial degradosome
- mitochondrial derivative +
- mitochondrial intermembrane space
- mitochondrial matrix +
- mitochondrial membrane +
- mitochondrial permeability transition pore
- mRNA editing complex
- Nebenkern

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Gene Ontology

References

BIOINFORMATICS

REVIEW

Vol. 21 no. 18 2005, pages 3587–3595
doi:10.1093/bioinformatics/bti565

Databases and ontologies

Ontological analysis of gene expression data: current tools, limitations, and open problems

Purvesh Khatri and Sorin Drăghici*

Department of Computer Science, Wayne State University, 431 State Hall, Detroit, MI 48202, USA

Methodology article

Open Access

Iterative Group Analysis (iGA): A simple tool to enhance sensitivity and facilitate interpretation of microarray experiments

Rainer Breitling*^{1,2}, Anna Amtmann¹ and Pawel Herzyk^{2,3}

Address: ¹Plant Science Group, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom.
²Bioinformatics Research Centre, Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, United Kingdom and ³Sir Henry Wellcome Functional Genomics Facility, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom

Email: Rainer Breitling* - r.breitling@bio.gla.ac.uk; Anna Amtmann - a.amtmann@bio.gla.ac.uk; Pawel Herzyk - p.herzyk@bio.gla.ac.uk

* Corresponding author

Published: 29 March 2004

Received: 15 November 2003

BMC Bioinformatics 2004, 5:34

Accepted: 29 March 2004

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Gene Ontology

References

Khatri and Draghici, 2005

Table 2. In the GO visualization column, 'flat' indicates that the tool does not represent the hierarchical structure of the GO when displaying the results, 'tree' indicates that the tool displays the GO hierarchy as a tree, whereas 'DAG' indicates that the tool displays the GO as a directed acyclic graph. The other columns are self-explanatory

<http://vortex.cs.wayne.edu/projects.htm#Onto-Express>

Tool	Statistical model	Correction for multiple experiments	GO Visualization	Microarrays supported	Time to process 200 genes (s)
Onto-Express	χ^2 , binomial, hypergeometric, Fisher's exact test	Šidák, Holm, Bonferroni, FDR	Flat, Tree	172 commercial arrays (Affymetrix, SuperArray, Sigma-Genosys, ClonTech, PerkinElmer, Operon, Takara, NIA); can also upload a user-defined list	7, 8, 16, 28
GoMiner	Fisher's exact test	Relative enrichment	Tree, DAG	uploads from user	77, 123, 223, 340
DAVID	None	None	Not available	Not applicable	15, 17, 27, 54
EASEonline	Fisher's exact test	Bonferroni	Not available	27 arrays (Affymetrix only); can also upload a user-defined list	15, 19, 34, 74
GeneMerge	Hypergeometric	Bonferroni	Flat, no hierarchical structure	Uploads from user	6, 6, 6, 8
FuncAssociate	Fisher's exact test	None	Not available	Uploads from user	22, 27, 29, 50
GOTM	Hypergeometric	None	Tree	37 arrays (Affymetrix only); uploads from user	59, 60, 157,
FatiGO	Percentage	Step-down minP, FDR (Benjamini and Hochberg, 1995), FDR (Benjamini)	Flat, Tree	Uploads from user	15, 49, 69, 105

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



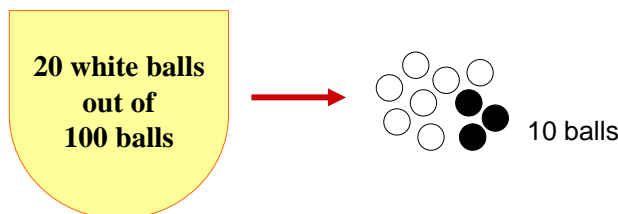
Gene Ontology

Analysis

The Hypergeometric Test

In a given list of genes of interest (eg. DE), is there a Gene Ontology term that is more represented than what it would be expected by chance only?

The hypergeometric distribution arises from sampling from a fixed population.



- We want to calculate the probability for drawing 7 or more white balls out of 10 balls given the distribution of balls in the urn

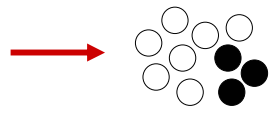
Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Analysis

The Hypergeometric Test

20 white balls
out of
100 balls



$$P(z, n, t, x) = \frac{\binom{t}{z} \binom{n-t}{x-z}}{\binom{n}{x}}$$

- t = N of balls draws from the ball \longrightarrow 10
- z = N of ball from the particular condition (ie. White) \longrightarrow 7
- n = N of ball in the urn \longrightarrow 100
- x = N of ball in the urn with the particular condition (ie. White) \longrightarrow 20

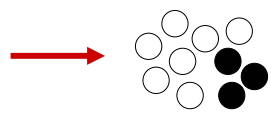
P = 0.000368



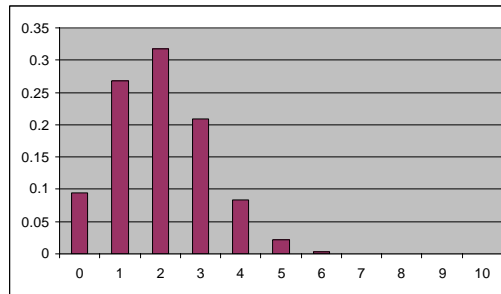
Analysis

The Hypergeometric Test

20 white balls
out of
100 balls



$$P(z, n, t, x) = \frac{\binom{t}{z} \binom{n-t}{x-z}}{\binom{n}{x}}$$





Gene Ontology

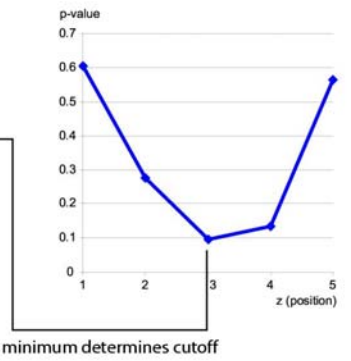
Analysis

The Hypergeometric Test

IGA: Iterative Group Analysis (Breitling et al. 2004)

- Gene 1
- Gene 2 = Group member 1 (t = 2; z = 1)
- Gene 3 = Group member 2 (t = 3; z = 2)
- Gene 4 = Group member 3 (t = 4; z = 3)
- Gene 5
- Gene 6
- Gene 7 = Groupmember 4 (t = 7; z = 4)
- Gene 8
- Gene 9
- Gene 10
- Gene 11
- Gene 12
- Gene 13 = Group member 5 (t = 13; z = 5)
- Gene 14

↑ Increasing fold-change



n = 14; x = 5



Gene Ontology

Example

Reverter et al. (2004) J. Anim Sci 82:3430

Genes = 7,638

DE Genes



450

387

Table 3. Description gene ontology (GO) classes with a probability of change by chance alone less than 1.0×10^{-4} and for each experiment (Exp. 1, 2, and 3), number of elements (No.), and differentially expressed (DE) elements for each GO class

GO ^a	Description	No.	DE
Exp. 1: Downregulated in high-quality diet			
0006355	Regulation of transcription	50	14
0003714	Transcription co-repressor	19	12
0006937	Muscle contraction	17	7
0006412	Protein biosynthesis	82	21
0005509	Calcium ion binding	133	7
Exp. 1: Upregulated in high-quality diet			
0005584	Collagen type I	22	21
0003754	Chaperone activity	23	22
0001503	Ossification	19	6
0004129	Cytochrome-c oxidase	67	59
0005200	Structural cytoskeleton	110	33
Exp. 2: Downregulated in Holstein breed			
0016491	Oxidoreductase activity	149	21
0016740	Transferase activity	45	18
0003754	Chaperone activity	23	8
0016529	Sarcoplasmic reticulum	24	20
0008151	Cell growth or maintenance	24	16
Exp. 2: Upregulated in Holstein breed			
0001501	Skeletal development	23	20
0003714	Transcription co-repressor	19	11
0017022	Myosin binding	7	4
0015992	Proton transport	35	10
0006412	Protein biosynthesis	82	26