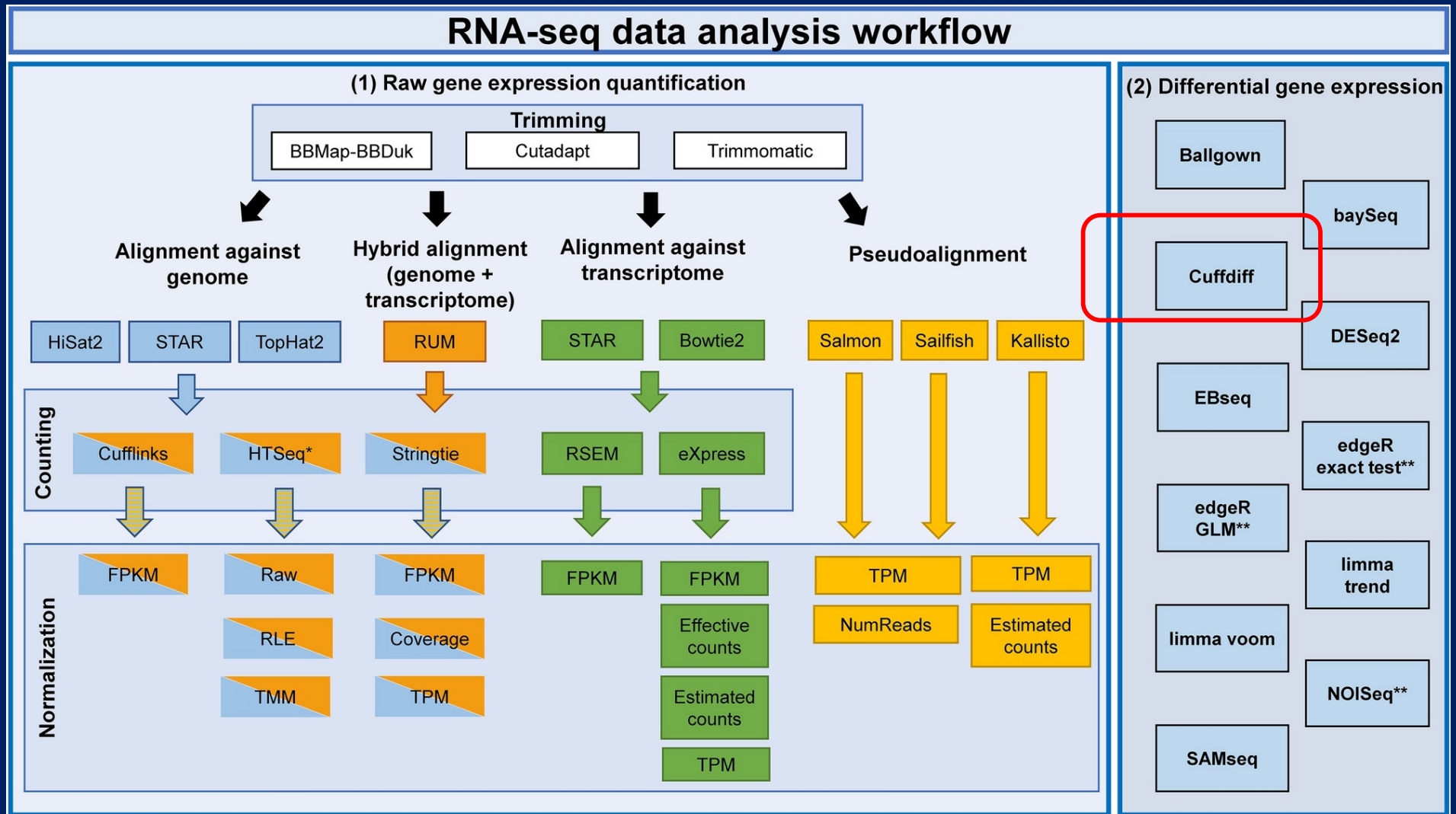# RNA-Seq Overview

Four major steps, semi-independent of each other.

I. Mapping → produce SAM/BAM or counts data.

II. Quantification → produce RPKM for each gene/transcript.

III. Identifying DEG (Differentially expressed genes) → gene list.

IV. Identifying affected biological processes/pathways.

RNA-seq: Identifying DEGs

RNA-seq data analysis workflow

(1) Raw gene expression quantification

Trimming

| BBMap-BBDuk | Cutadapt | Trimmomatic |

- Alignment against genome
- Hybrid alignment (genome + transcriptome)
- Alignment against transcriptome
- Pseudoalignment

HiSat2 | STAR | TopHat2 | RUM | STAR | Bowtie2 | Salmon | Sailfish | Kallisto

Counting

Cufflinks | HTSeq* | Stringtie | RSEM | eXpress

Normalization

FPKM | Raw | FPKM | FPKM | FPKM | TPM | TPM
RLE | Coverage | Effective counts | NumReads | Estimated counts
TMM | TPM | Estimated counts
TPM

(2) Differential gene expression

Ballgown
baySeq
Cuffdiff
DESeq2
EBseq
edgeR exact test**
edgeR GLM**
limma trend
limma voom
NOISeq**
SAMseq

Generate gene list from CuffDiff output

Practice

1. Download the "Wholegut_aging" folder to your laptop.

2. Open the "gene_exp.diff" file in MS Excel.

3. Sort the whole table according to the p-value column. Save it as a .xlxs file.

4. Make tabs for increased vs. decreased genes.

5. Generate lists in your text editor and save those in your project folder.

Functional Analysis of HTS data

- ➤ Gene Ontology – http://www.geneontology.org/

- ➤ Regulatory pathways.

- ➤ Modeling & Systems Biology.

Gene Ontology – hierarchical framework of terms / concepts

Top Docs Gene Ontology GO Links GO Summary

GO:0003673 : Gene_Ontology (46199)
- GO:0008150 : biological_process (30188)
  - GO:0016265 : death (525)
    - GO:0008219 : cell death (484)
      - GO:0012501 : programmed cell death (447)
        - GO:0006915 : apoptosis (419)
          - GO:0006916 : anti-apoptosis (111)
          - GO:0008632 : apoptotic program (51)
            - GO:0008637 : apoptotic mitochondrial changes (11)
            - GO:0030262 : apoptotic nuclear changes (10)
              - GO:0030263 : apoptotic chromosome condensation (1)
              - GO:0006309 : DNA fragmentation (9)
              - GO:0030264 : nuclear fragmentation (0)
            - GO:0006919 : caspase activation (16)
            - GO:0006921 : disassembly of cell structures (10)
            - GO:0008633 : induction of proapoptotic gene products (0)
            - GO:0045884 : regulation of survival gene products (7)
          - GO:0006917 : induction of apoptosis (148)
            - GO:0008624 : induction of apoptosis by extracellular signals (46)
            - GO:0008629 : induction of apoptosis by intracellular signals (23)
            - GO:0019051 : induction of apoptosis by virus (0)
          - GO:0006925 : killing of inflammatory cells (0)
          - GO:0006927 : killing transformed cells (3)
          - GO:0006926 : killing virus-infected cells (1)
          - GO:0045476 : nurse cell apoptosis (1)
          - GO:0006924 : peripheral killing of activated T-cells (0)
        - GO:0012502 : induction of programmed cell death (148)
          - GO:0006917 : induction of apoptosis (148)
            - GO:0008624 : induction of apoptosis by extracellular signals (46)
            - GO:0008629 : induction of apoptosis by intracellular signals (23)
            - GO:0019051 : induction of apoptosis by virus (0)
- GO:0005575 : cellular_component (22371)
- GO:0003674 : molecular_function (37018)

DAG view

Gene Ontology

Goal – "produce a dynamic controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells is accumulating and changing" – GO consortium (~2001)
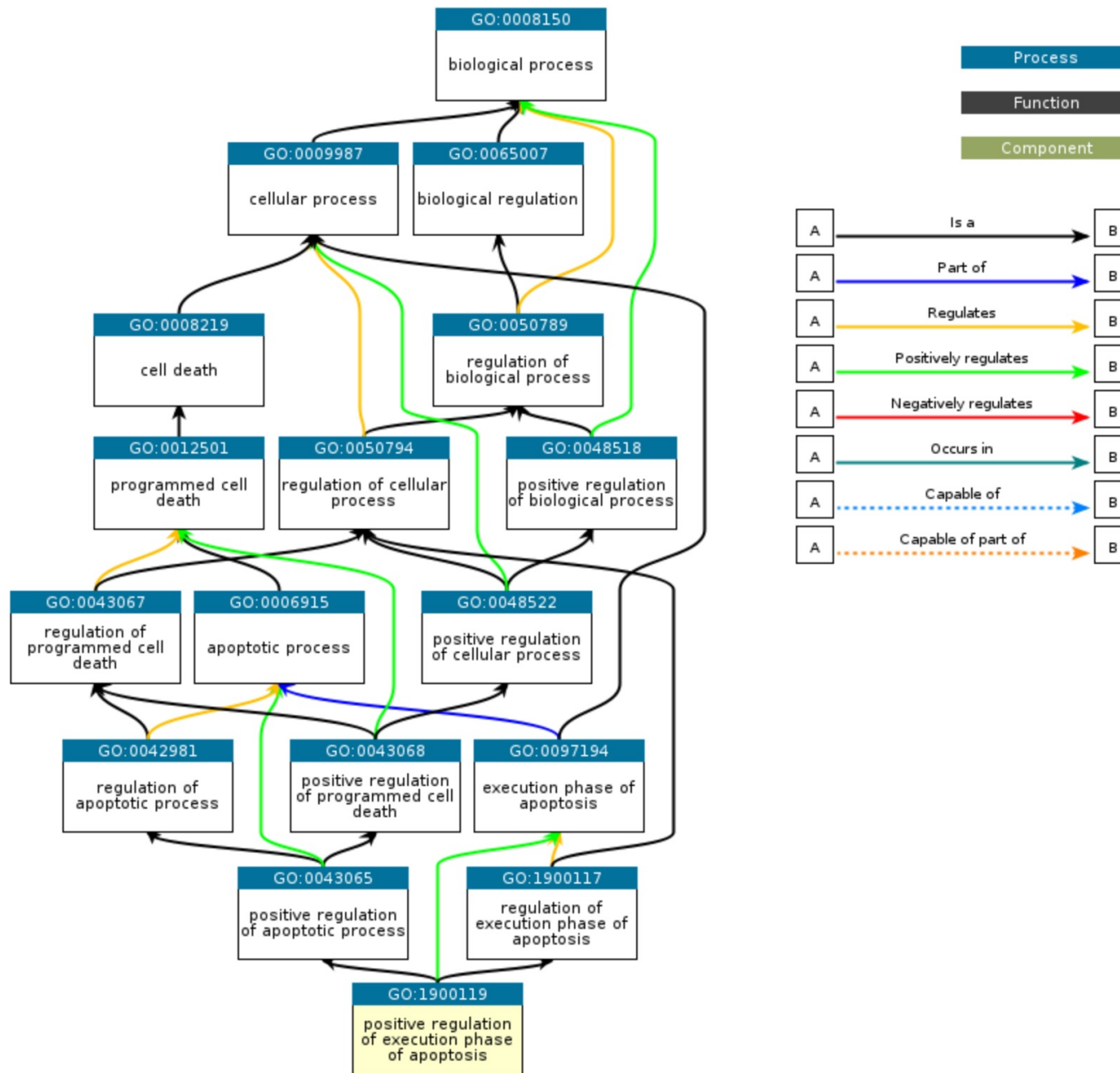
Ontology:

" The branch of metaphysics that deals with the nature of being" – The American Heritage Dictionary

Gene Ontology – hierarchical framework of terms / concepts

Gene Ontology – hierarchical framework of terms / concepts

QuickGO - https://www.ebi.ac.uk/QuickGO
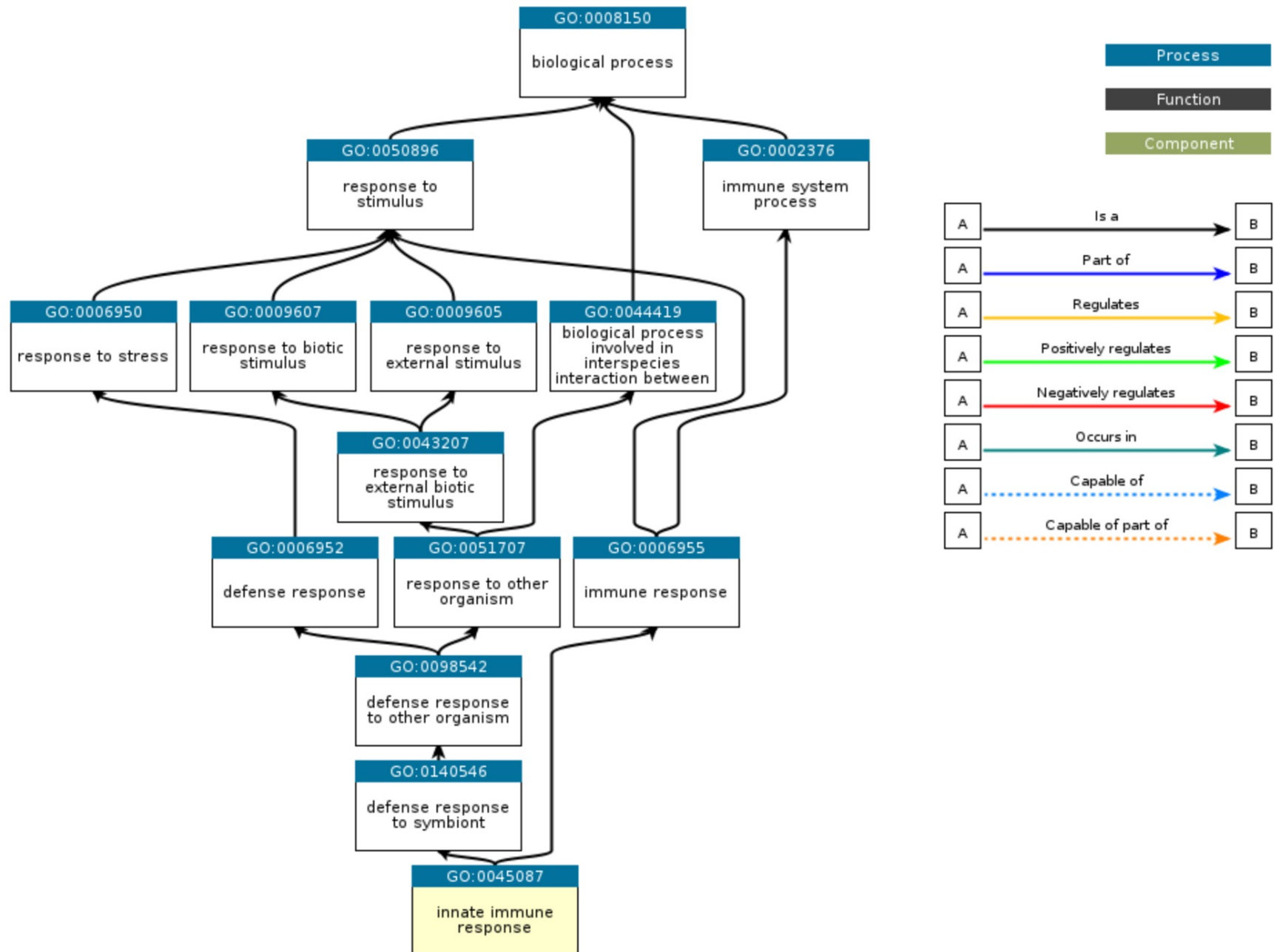
Implications of Gene Ontology (I)

Monitoring biological processes or molecular functions beyond individual gene.

Example:

1.) Which biological process (mol. Function) is activated/suppressed following a treatment?

Gene Expression Profile Differences between the two lung cancer cell lines A549 and H23

extracellular (GO:0005576)	1.91E-08	169
Cell Communication	1.32E-07	690
plasma membrane (GO:0005886)	1.34E-07	511
Complement and coagulation cascades - Homo sapiens	1.73E-07	20
Metabolism	2.10E-06	174
carbohydrate metabolism (GO:0005975)	2.45E-06	207
cell adhesion molecule activity (GO:0005194)	0.000102	113
Structural Protein	0.000231	271
extracellular matrix (GO:0005578)	0.000235	53
Cell Growth and Maintenance	0.000569	590
Cell Adhesion	0.000917	100

development	1.40E-07	596
cell differentiation (GO:0030154)	6.60E-05	186
regulation of gene expression, epigenetic (GO:0040029)	7.71E-05	442
cell growth (GO:0016049)	8.37E-05	307
transcription regulator activity (GO:0030528)	0.000307	319
extracellular (GO:0005576)	0.000515	153

Implications of Gene Ontology (II)

Basis for cross genome comparison and integrating knowledge from different model systems.

Term	Fly Genes	Worm Genes	Mouse Genes	Human Genes	Sacc. Yeast Genes	Pombe Yeast Genes	Weed Genes
cell cycle	265	182	294	717	424	622	181
cell cycle dependent actin filament reorganization	2	0	0	0	4	0	0
DNA replication and chromosome cycle	134	128	67	175	172	73	146
endomitotic cell cycle	0	0	1	1	0	3	0
M phase	171	39	69	181	213	253	3
mitotic cell cycle	133	140	102	314	239	202	141
nuclear migration	1	0	0	0	13	0	0
regulation of cell cycle	42	4	136	383	87	65	3
schizogony	0	0	0	0	0	0	0
second mitotic wave (sensu Drosophila)	1	0	0	0	0	0	0

Tools associated with GO

A comprehensive list at GO web site.

Tools for browsing, AmiGO, QuickGO at EBI, etc.

Tools for identifying over represented GOs/pathways, **etc.**

Using GO to gain a comprehensive picture of what your RNA-Seq data reveal

Practice: Load a gene list to identify over-represented GO