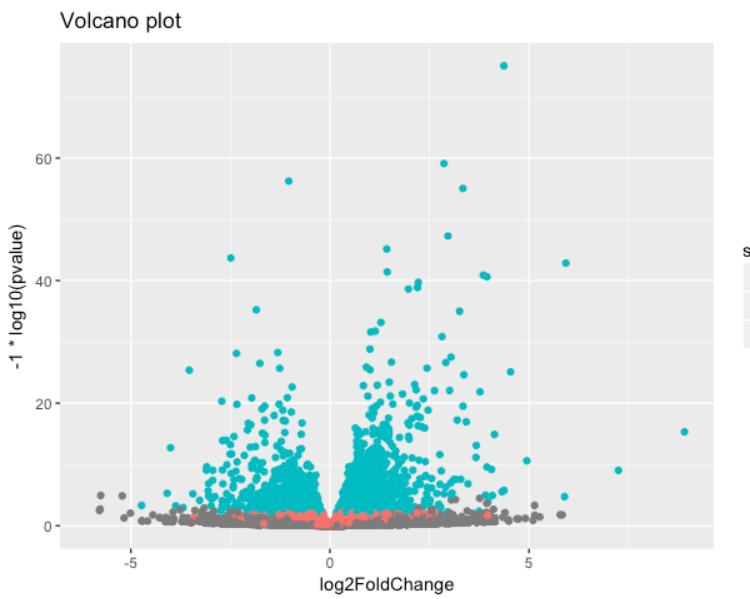


X	baseMean	log2FoldChange	IfcSE	stat	pvalue	padj	symbol
ENSG00000152583	954.77093	4.3683590	0.23713648	18.421286	8.867079e-76	1.342919e-71	SPARCL1
ENSG00000179094	743.25269	2.8638885	0.17555825	16.313039	7.972621e-60	6.037267e-56	PER1
ENSG00000116584	2277.91345	-1.0347000	0.06505273	-15.905557	5.798513e-57	2.927283e-53	ARHGEF2
ENSG00000189221	2383.75371	3.3415441	0.21241508	15.731200	9.244206e-56	3.500088e-52	MAOA
ENSG00000120129	3440.70375	2.9652108	0.20370277	14.556557	5.306416e-48	1.607313e-44	DUSP1
ENSG00000148175	13493.92037	1.4271683	0.10036663	14.219550	6.929711e-46	1.749175e-42	STOM
ENSG00000178695	2685.40974	-2.4890689	0.17806407	-13.978501	2.108817e-44	4.562576e-41	KCTD12
ENSG00000109906	439.54152	5.9275950	0.42819442	13.843233	1.397758e-43	2.646131e-40	ZBTB16
ENSG00000134686	2933.64246	1.4394898	0.10582729	13.602255	3.882769e-42	6.533838e-39	PHC2
ENSG00000101347	14134.99177	3.8504143	0.28490701	13.514635	1.281894e-41	1.941428e-38	SAMHD1
ENSG00000096060	2630.23049	3.9450524	0.29291821	13.468102	2.409807e-41	3.317866e-38	FKBP5
ENSG00000166741	7542.25287	2.2195906	0.16673544	13.312050	1.970000e-40	2.486304e-37	NNMT
ENSG00000125148	3695.87946	2.1985636	0.16700546	13.164621	1.402400e-39	1.633797e-36	MT2A
ENSG00000162614	5646.18314	1.9711402	0.15020631	13.122885	2.434854e-39	2.633990e-36	NEXN
ENSG00000106976	989.04683	-1.8501713	0.14778657	-12.519211	5.861471e-36	5.918132e-33	DNM1
ENSG00000187193	199.07694	3.2551424	0.26090711	12.476250	1.006146e-35	9.523804e-33	MT1X
ENSG00000256235	1123.47954	1.2801193	0.10547438	12.136779	6.742862e-34	6.007096e-31	SMIM3
ENSG00000177666	2639.57020	1.1399947	0.09606884	11.866436	1.768422e-32	1.487930e-29	NPPLA2
ENSG00000164125	7257.00808	1.0248523	0.08657600	11.837603	2.494830e-32	1.988642e-29	FAM198B
ENSG00000198624	2020.04495	2.8141014	0.24063429	11.694515	1.359615e-31	1.029569e-28	CCDC69
ENSG00000123562	5008.55294	1.0045453	0.08901501	11.285123	1.554241e-29	1.120904e-26	MORF4L2
ENSG00000144369	1283.77980	-1.3090041	0.11714863	-11.173875	5.473974e-29	3.768333e-26	FAM171B
ENSG00000196517	241.91536	-2.3456877	0.21047366	-11.144804	7.591120e-29	4.998588e-26	SLC6A9
ENSG00000135821	19973.40000	3.0413943	0.27601796	11.018828	3.100706e-28	1.956675e-25	GLUL



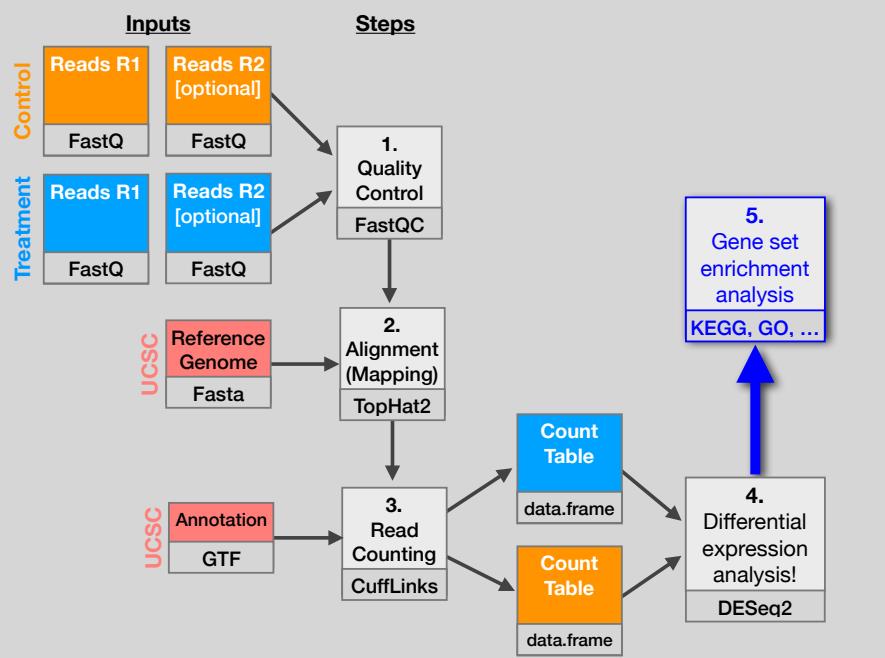
My high-throughput experiment generated a long list of genes/proteins...

What do I do now?

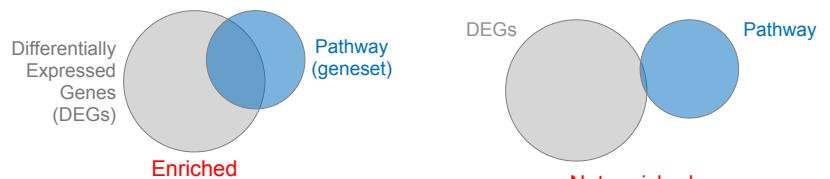


Pathway analysis! (a.k.a. geneset enrichment)

Use bioinformatics methods to help extract biological meaning from such lists...



Pathway analysis (a.k.a. geneset enrichment) Principle



-
- Variations of the math: overlap, ranking, networks... > *Not critical, different algorithms show similar performances*
 - DEGs come from your experiment > *Critical, needs to be as clean as possible*
 - Pathway genes ("geneset") come from annotations > *Important, but typically not a competitive advantage*

Pathway analysis (a.k.a. geneset enrichment)

Limitations

- **Geneset annotation bias:** can only discover what is already known
 - **Post-transcriptional regulation** is neglected
 - **Tissue-specific** variations of pathways are not annotated
 - e.g. NF- κ B regulates metabolism, not inflammation, in adipocytes
 - **Size bias:** stats are influenced by the size of the pathway
 - **Non-model organisms:** no high-quality genesets available
 - Many pathways/receptors **converge** to few regulators
 - e.g. Tens of innate immune receptors activate four TFs: NF- κ B, AP-1, IRF3/7, NFAT

Side-note

Starting point for pathway analysis: **Your gene list**

- You have a list of genes/proteins of interest
 - You have quantitative data for each gene/protein
 - Fold change
 - p-value
 - Spectral counts
 - Presence/absence

228018	ENSG000000090339	NP	C20orf58
226	ENSG00000000030	ND	000192
207	ENSG00000000001	3383	057219
225	ENSG00000000001	51513	055029
221	ENSG00000000001	3613	000585
1553	ENSG00000000001	7124	06125
2184	ENSG00000000001	757	89495
2049	ENSG00000000001	92370	01032249
2026	ENSG00000000001	79646	8870
23095	ENSG00000000001	56892	4515
22801	ENSG00000000001	124540	83893
15540	ENSG00000000001	253982	412
20312	ENSG00000000001	140688	NMB
225182	ENSG00000000001	10457	069
225079	ENSG00000000001	9518	PA2
243010	ENSG00000000001	2013	83
230668	ENSG00000000001	4050	MEM50B
218541	ENSG00000000001	0_33666	01
224225	ENSG00000000001	NP_002332	MP2
207339	ENSG00000000001	524	MS12
202637	s_at	W03F8.6	C20orf58

Translating between identifiers

- Many different identifiers exist for genes and proteins, e.g. UniProt, Entrez, etc.
 - Sometimes you have to translate one set of ids into another
 - A program might only accept certain types of ids
 - You might have a list of genes with one type of id and info for genes with another type of id

Translating between identifiers

- Many different identifiers exist for genes and proteins, e.g. UniProt, Entrez, etc.
 - Sometimes you have to translate one set of ids into another
 - A program might only accept certain types of ids
 - You might have a list of genes with one type of id and info for genes with another type of id
 - **Various web sites translate ids -> *best for small lists***
 - UniProt <www.uniprot.org>; IDConverter <idconverter.bioinfo.cnio.es>

Translating between identifiers: UniProt < www.uniprot.org >

The screenshot shows the UniProt ID Mapping tool. At the top, there's a search bar and a 'Query' field. Below that are buttons for 'Search', 'Blast', 'Align', 'Retrieve', and 'ID Mapping' (which is highlighted with a red box). The main area has sections for 'Identifiers', 'From' (set to EMBL/GenBank/DDBJ), and 'To' (set to UniProtKB AC). There are also 'Map', 'Swap', and 'Clear' buttons. To the right of the 'From' dropdown is a large, empty text input field.

Translating between identifiers

- Many different identifiers exist for genes and proteins, e.g. UniProt, Entrez, etc.
- Sometimes you have to translate one set of ids into another
 - A program might only accept certain types of ids
 - You might have a list of genes with one type of id and info for genes with another type of id
- Various web sites translate ids -> *best for small lists*
 - [UniProt < www.uniprot.org>; IDConverter < \[idconverter.bioinfo.cnio.es\]\(http://idconverter.bioinfo.cnio.es\) >](http://www.uniprot.org)
- VLOOKUP in Excel - good if you are an excel whizz - I am not!**
 - Download flat file from Entrez, Uniprot, etc; Open in Excel; Find columns that correspond to the 2 IDs you want to convert between; Sort by ID; Use vlookup to translate your list

Translating between identifiers: Excel VLOOKUP

`VLOOKUP(lookup_value, table_array, col_index_num)`

The screenshot shows an Excel spreadsheet with two tables. The first table, 'Data Table', has columns A through F. The second table, 'Annotation Table', has columns G through K. The formula `=VLOOKUP(A3,G3:O30490,2,FALSE)` is entered in cell B3. The data includes gene symbols like Kif1c, Gabrp, Camkk1, and their corresponding RefSeq numbers and expression values (Exp1, Exp2, Exp3). The 'Annotation Table' provides additional information such as Entrez ID, Unigene ID, and RefSeq numbers for each gene.

Translating between identifiers

- Many different identifiers exist for genes and proteins, e.g. UniProt, Entrez, etc.
- Sometimes you have to translate one set of ids into another
 - A program might only accept certain types of ids
 - You might have a list of genes with one type of id and info for genes with another type of id
- Various web sites translate ids -> *best for small lists*
 - [UniProt < www.uniprot.org>; IDConverter < \[idconverter.bioinfo.cnio.es\]\(http://idconverter.bioinfo.cnio.es\) >](http://www.uniprot.org)
- VLOOKUP in Excel -> good if you are an excel whizz - I am not!**
 - Download flat file from Entrez, Uniprot, etc; Open in Excel; Find columns that correspond to the two ids you want to convert between; Use vlookup to translate your list
- Use the `merge()` or `mapIDs()` functions in R - fast, versatile & reproducible!**
 - Also `clusterProfiler::bitr()` function and many others... [[Link to clusterProfiler vignette](#)]

bitr: Biological Id Translator

clusterProfiler provides `bitr` and `bitr_kegg` for converting ID types. Both `bitr` and `bitr_kegg` support many species including model and many non-model organisms.

```
x <- c("GPX3", "GLRX", "LBP", "CRYAB", "DEFB1", "HCLS1", "SOD2", "HSPA2", "ORM1", "IGFBP1", "PTHLH", "GPC3", "IGFBP3", "T0B1", "MITE", "NDRG1", "NR1H4", "FGFR3", "PVR", "IL6", "PTPRM", "ERBB2", "NID2", "LAMB1", "COMP", "PLS3", "MCAM", "SPP1", "LAMC1", "COL4A2", "COL4A1", "MYOC", "ANXA4", "TFPI2", "CST6", "SLPI", "TIMP2", "CPM", "GGTM", "NNMT", "HAL", "EEF1A2", "HGD", "TCN2", "CDA", "PCCA", "CRYM", "PDXX", "STC1", "WARS", "HMOX1", "FXYD2", "RBP4", "SLC6A12", "KDELR3", "ITM2B")  
eg = bitr(x, fromType="SYMBOL", toType="ENTREZID", OrgDb="org.Hs.eg.db")  
head(eg)
```

```
## SYMBOL ENTREZID  
## 1 GPX3 2878  
## 2 GLRX 2745  
## 3 LBP 3929  
## 4 CRYAB 1410  
## 5 DEFB1 1672  
## 6 HCLS1 3059
```

See package vignette:
<https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>

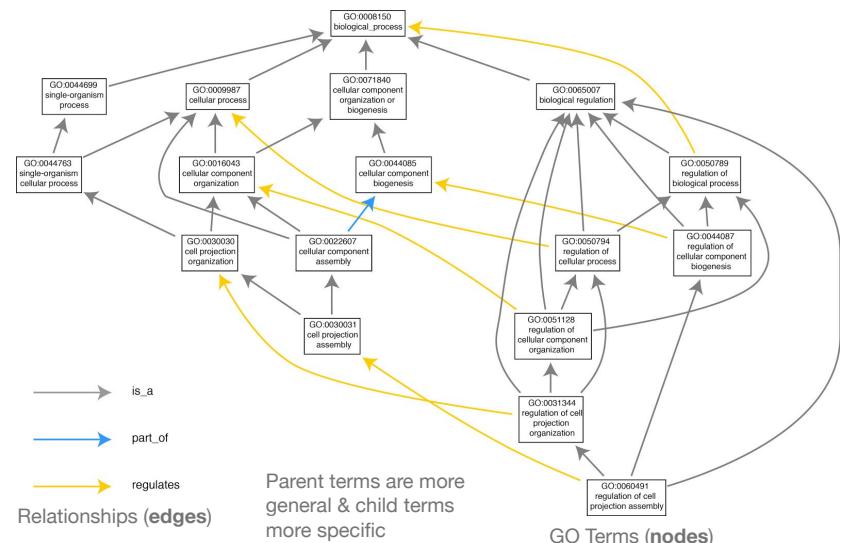
GO database < www.geneontology.org >

- What function does HSF1 perform?
 - response to heat; sequence-specific DNA binding; transcription; etc
- **Ontology** => a structured and controlled vocabulary that allows us to annotate gene products consistently, interpret the relationships among annotations, and can easily be *handled by a computer*
- GO database consists of 3 ontologies that describe gene products in terms of their associated **biological processes**, **cellular components** and **molecular functions**

What functional set databases do you want?

- Commonly used
 - **Gene Ontology (GO)**
 - **KEGG Pathways** (mostly metabolic)
 - **GeneGO MetaBase** 
 - **Ingenuity Pathway Analysis (IPA)** 
 - **MSigDB** (Molecular Signatures Database: gene sets based on chromosomal position, cis-regulatory motifs, GO terms, etc)
- Many others...
 - **Enzyme Classification, PFAM, Reactome, Disease Ontology, Chemical Entities of Biological Interest, Network of Cancer Genes** etc...
 - See: Open Biomedical Ontologies (www.obofoundry.org)

GO is structured as a “directed graph”



GO Annotations

- GO is not a database of genes/proteins or sequences
- Gene products get annotated with GO terms by organism specific databases, such as Flybase, Wormbase, MGI, ZFIN, UniProt, etc
- Annotations are available through AmiGO <amigo.geneontology.org>

The screenshot shows the AmiGO homepage with a search bar at the top. Below it, there's a search result table with columns for 'Evidence code', 'Evidence code description', 'Source of evidence', 'Manually checked', and 'Current number of annotations*'. One row in the table is circled in red, highlighting the 'IEA' entry which has a value of 15,687,382. At the bottom of the page, there's a note: '*October 2007 release'.

Evidence code	Evidence code description	Source of evidence	Manually checked	Current number of annotations*
IDA	Inferred from direct assay	Experimental	Yes	71,050
IEP	Inferred from expression pattern	Experimental	Yes	4,598
IGI	Inferred from genetic interaction	Experimental	Yes	8,311
IMP	Inferred from mutant phenotype	Experimental	Yes	61,549
IPI	Inferred from physical interaction	Experimental	Yes	17,043
ISS	Inferred from sequence or structural similarity	Computational	Yes	196,643
RCA	Inferred from reviewed computational analysis	Computational	Yes	103,792
IGC	Inferred from genomic context	Computational	Yes	4
IEA	Inferred from electronic annotation	Computational	No	15,687,382
IC	Inferred by curator	Indirectly derived from experimental or computational evidence made by a curator	Yes	5,167
TAS	Traceable author statement	Indirectly derived from experimental or computational evidence made by the author of the published article	Yes	44,564
NAS	Non-traceable author statement	No 'source of evidence' statement given	Yes	25,656
ND	No biological data available	No information available	Yes	132,192
NR	Not recorded	Unknown	Yes	1,185

*October 2007 release

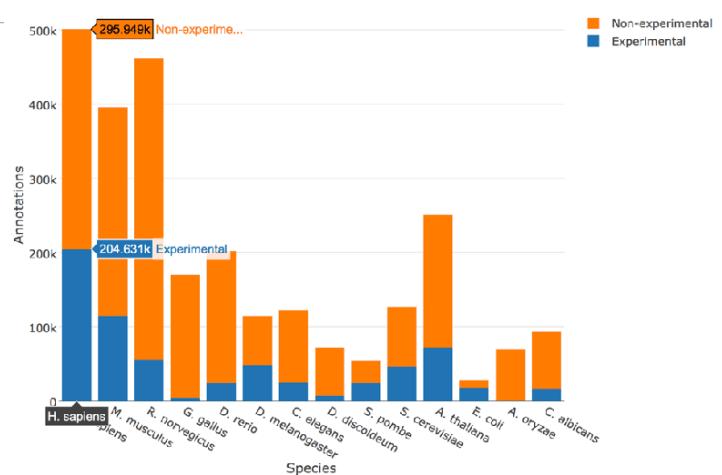
GO evidence codes

Evidence code	Evidence code description	Source of evidence	Manually checked	Current number of annotations*
IDA	Inferred from direct assay	Experimental	Yes	71,050
IEP	Inferred from expression pattern	Experimental	Yes	4,598
IGI	Inferred from genetic interaction	Experimental	Yes	8,311
IMP	Inferred from mutant phenotype	Experimental	Yes	61,549
IPI	Inferred from physical interaction	Experimental	Yes	17,043
ISS	Inferred from sequence or structural similarity	Computational	Yes	196,643
RCA	Inferred from reviewed computational analysis	Computational	Yes	103,792
IGC	Inferred from genomic context	Computational	Yes	4
IEA	Inferred from electronic annotation	Computational	No	15,687,382
IC	Inferred by curator	Indirectly derived from experimental or computational evidence made by a curator	Yes	5,167
TAS	Traceable author statement	Indirectly derived from experimental or computational evidence made by the author of the published article	Yes	44,564
NAS	Non-traceable author statement	No 'source of evidence' statement given	Yes	25,656
ND	No biological data available	No information available	Yes	132,192
NR	Not recorded	Unknown	Yes	1,185

*October 2007 release

Use and misuse of the gene ontology annotations
Seung Yon Rhee, Valerie Wood, Kara Dolinski & Sorin Draghici
Nature Reviews Genetics 9, 509-515 (2008)

Experimental annotations by species



- See AmiGO for details: http://amigo.geneontology.org/amigo/base_statistics

Can now do gene list analysis with GeneGO

The screenshot shows the PANTHER Classification System interface. The main window is titled 'Gene List Analysis' and includes tabs for 'Browse', 'Sequence Search', 'CSNP Scoring', and 'Keyword Search'. On the left, there's a sidebar with links like 'About', 'PANTHER Data', 'PANTHER Tools', 'Workspace', 'Downloads', and 'Help/Tutorial'. A message at the top says 'Please refer to our article in *Nature Protocols* for detailed instructions on how to use this page.' Below it, an error message says 'Error parsing request, no input specified'. The main form has several sections: 'Help Tips Steps:', 'Enter IDs: Supported IDs', 'Upload IDs: File Format', 'Select List Type:', 'Select organism.', and 'Select Analysis: Functional classification viewed in gene list'. There are also checkboxes for 'ID list', 'Previously reported list search results', 'Workspace list', 'PANTHER Generic Mapping File', and 'VCF File - Ranking region: 20 Kb'.

DAVID at NIAID < david.abcc.ncifcrf.gov >

Analysis Wizard
DAVID Bioinformatics Resources 2008, NIAID/NIH

Home Start Analysis Shortcut to DAVID Tools Technical Center Downloads & APIs Term of Service Why DAVID? About Us

Analysis Wizard

Upload Gene List

Demolist 1 Demolist 2 Upload Help

Step 1: Enter Gene List
A: Paste a list

Clear

B: Choose From a File
Choose File or file selected

Step 2: Select Identifier
AFFY_ID

Step 3: List Type
Gene List (radio button selected)
Background

Step 4: Submit List
Submit List

Tell us how you like the tool Contact us for questions

Note: Affy Exon IDs and Affy Gene Array IDs are now supported in DAVID, as "affy_id" type.

An example:
1007_s_at
1008_s_at
117_at
121_at
1255_g_at
1301_at
1316_at
1320_at
1431_s_at
1431_t_at
1438_at
1487_at
1494_f_at
1598_g_at

DAVID

- Notice that you can pick a *Background* (Universe)

Analysis Wizard

Upload List Background

Gene List Manager

Select to limit annotations by one or more species [Help](#)
Use All Species HOMO SAPIENS(4402) SYNTHETIC CONSTRUCT(S)
Select

List Manager Help

Uploaded List_2

Select List to:
Use Rename Remove Combine Show Gene List new!

Step 1. Successfully submitted gene list
Current Gene List: Uploaded List_2
Current Background: HOMO SAPIENS

Step 2. Analyze above gene list with one of DAVID tools [Which DAVID tools to use?](#)

Functional Annotation Tool
 Functional Annotation Clustering
 Functional Annotation Chart
 Functional Annotation Table

Gene Functional Classification Tool
 Gene ID Conversion Tool
 Gene Name Batch Viewer

DAVID

- Functional Annotation Tool

Annotation Summary Results

Current Gene List: Uploaded List_3
Current Background: HOMO SAPIENS

2320 DAVID IDs Check Defaults Clear All

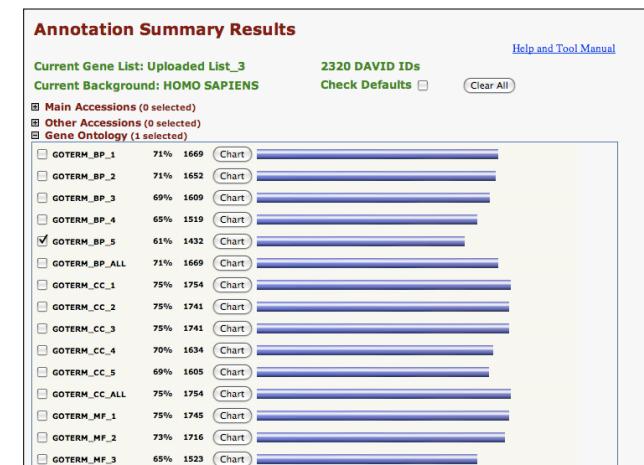
Main Accessions (0 selected)
 Other Accessions (0 selected)
 Gene Ontology (4 selected)
 Protein Domains (3 selected)
 Pathways (3 selected)
 General Annotations (0 selected)
 Functional Categories (3 selected)
 Protein Interactions (0 selected)
 Literature (0 selected)
 Disease (1 selected)
 Tissue Expression

Combined View for Selected Annotation
 Functional Annotation Clustering new! red arrow
 Functional Annotation Chart
 Functional Annotation Table

Help and Tool Manual

DAVID

- Specify functional sets



DAVID

- Let's look at the *Functional Annotation Chart*

Annotation Summary Results

Current Gene List: Uploaded List_3
Current Background: HOMO SAPIENS

2320 DAVID IDs Check Defaults Clear All

Main Accessions (0 selected)
Other Accessions (0 selected)
Gene Ontology (4 selected)
Protein Domains (3 selected)
Pathways (3 selected)
General Annotations (0 selected)
Functional Categories (3 selected)
Protein Interactions (0 selected)
Literature (0 selected)
Disease (1 selected)
Tissue Expression

Combined View for Selected Annotation

Functional Annotation Clustering [new!](#)

Functional Annotation Chart

Functional Annotation Table

DAVID

- Functional Annotation Chart*

Functional Annotation Chart

Current Gene List: Uploaded List_1
Current Background: Homo sapiens
2316 DAVID IDs

Options

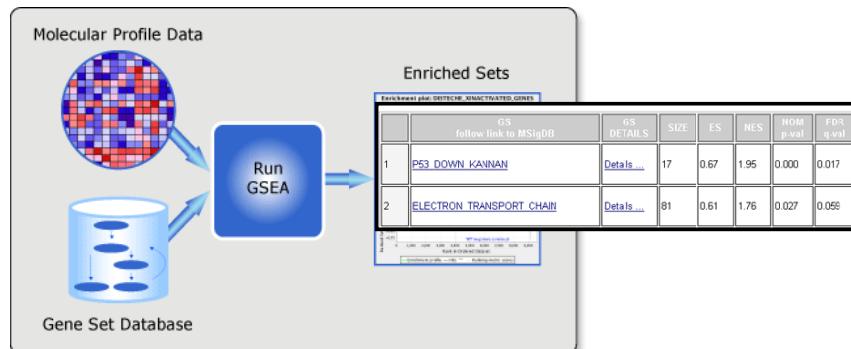
Return Using Options Create Sublist Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
	GOTERM_BP_S	regulation of progression through cell cycle	RT	98	4.2	3.3E-7	8.6E-4	
	GOTERM_BP_S	apoptosis	RT	131	5.7	1.6E-6	2.1E-3	
	GOTERM_BP_S	cell death	RT	136	5.9	3.8E-6	3.3E-3	
	GOTERM_BP_S	regulation of transcription from RNA polymerase II promoter	RT	83	3.6	3.7E-5	2.4E-2	
	GOTERM_BP_S	protein kinase cascade	RT	71	3.1	4.7E-5	2.4E-2	
	GOTERM_BP_S	regulation of kinase activity	RT	48	2.1	5.4E-5	2.3E-2	
	GOTERM_BP_S	negative regulation of cell proliferation	RT	48	2.1	1.0E-4	3.7E-2	
	GOTERM_BP_S	regulation of cell size	RT	41	1.8	1.2E-4	3.9E-2	
	GOTERM_BP_S	monocarboxylic acid metabolic process	RT	48	2.1	1.3E-4	3.6E-2	
	GOTERM_BP_S	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	RT	61	2.6	1.5E-4	3.8E-2	
	GOTERM_BP_S	positive regulation of cellular metabolic process	RT	72	3.1	1.7E-4	3.8E-2	

Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources
Da Wei Huang, Brad T Sherman & Richard A Lempicki
Nature Protocols 4, 44 - 57 (2009)

GSEA < www.broadinstitute.org/gsea >

- Download GSEA desktop application



- Excellent tutorial, user's guide and example datasets to work through

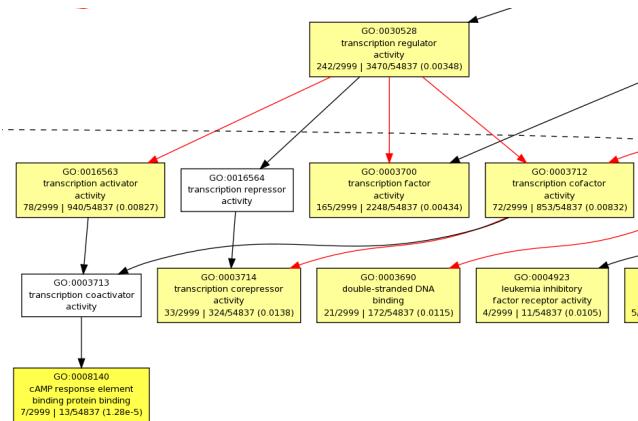
Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles
Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, ...
PNAS 102, 15545-15550 (2005)

Overlapping functional sets

- Many functional sets overlap, in particular those from databases that are hierarchical in nature (e.g. GO)
- Hierarchy enables:
 - Annotation flexibility (e.g. allow different degrees of annotation completeness based on what is known)
 - Computational methods to “understand” function relationships (e.g. ATPase function is a subset of enzyme function)
- Unfortunately, this also makes functional profiling trickier

GOEast < omicslab.genetics.ac.cn/GOEAST >

- Graphical view of enriched GO terms and their relationships



DAVID Functional Annotation Clustering

- Based on shared genes between functional sets



GO SLIMs

- Cut-down versions of the GO ontologies containing a subset of the terms in the whole GO
- GO FAT (DAVID):
 - filters out very broad GO terms based on a measured specificity of each term



Want more?

- GeneGO** < portal.genego.com >
 - MD/PhD curated annotations, great for certain domains (eg, Cystic Fibrosis)
 - Nice network analysis tools
 - Email us for access
- Oncomine** < www.oncomine.org >
 - Extensive cancer related expression datasets
 - Nice concept analysis tools
 - Research edition is free for academics, Premium edition \$\$\$
- Lots and lots other R/Bioconductor packages in this area!!!**

Hands-on time!

https://bioboot.github.io/bggm213_S18/lectures/#15

Also: R Quiz Online

Do it Yourself!

Pathway analysis (a.k.a. geneset enrichment) Principle



- Variations of the math: overlap, ranking, networks... ➤ *Not critical, different algorithms show similar performances*
- DEGs come from your experiment ➤ *Critical, needs to be as clean as possible*
- Pathway genes ("geneset") come from annotations ➤ *Important, but typically not a competitive advantage*

Data structure: counts + metadata

countData

gene	ctrl_1	ctrl_2	exp_1	exp_1
geneA	10	11	56	45
geneB	0	0	128	54
geneC	42	41	59	41
geneD	103	122	1	23
geneE	10	23	14	56
geneF	0	1	2	0
...

countData is the count matrix
(number of reads coming from
each gene for each sample)

First column of **colData** must match column names of **countData** (-1st)

colData

id	treatment	sex	...
ctrl_1	control	male	...
ctrl_2	control	female	...
exp_1	treatment	male	...
exp_2	treatment	female	...

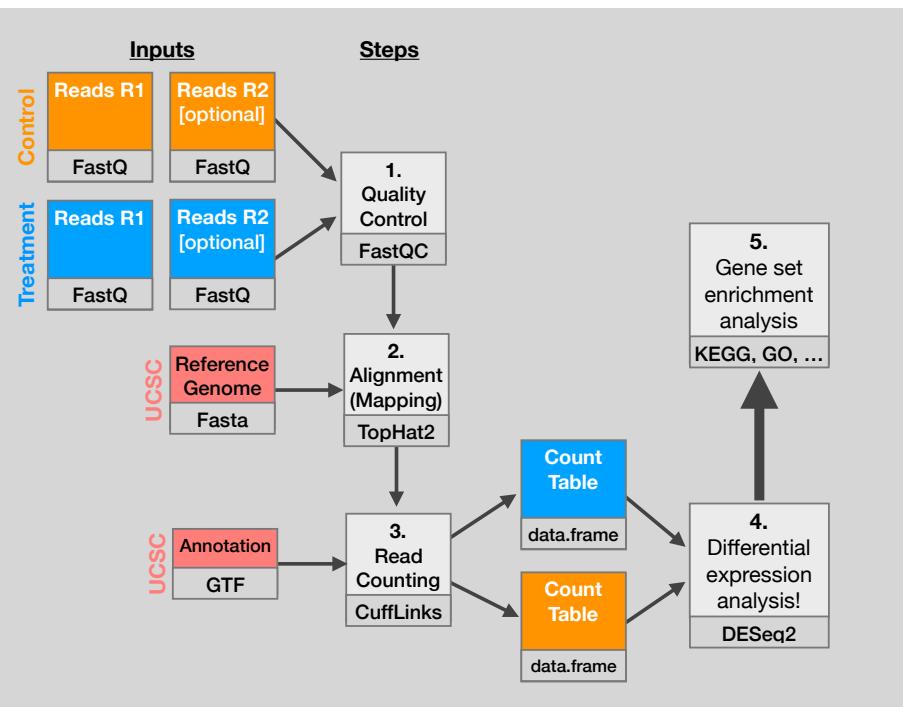
Sample names:
ctrl_1, ctrl_2, exp_1, exp_2

colData describes metadata
about the *columns* of countData

Pathway analysis (a.k.a. geneset enrichment) Limitations

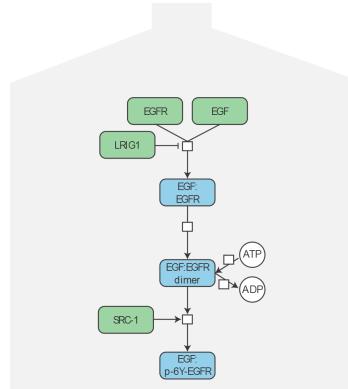
Side-note:

- **Geneset annotation bias:** can only discover what is already known
- **Post-transcriptional regulation** is neglected
- **Tissue-specific** variations of pathways are not annotated
 - e.g. NF- κ B regulates metabolism, not inflammation, in adipocytes
- **Size bias:** stats are influenced by the size of the pathway
- **Non-model organisms:** no high-quality genesets available
- Many pathways/receptors **converge** to few regulators
 - e.g. Tens of innate immune receptors activate four TFs: NF- κ B, AP-1, IRF3/7, NFAT

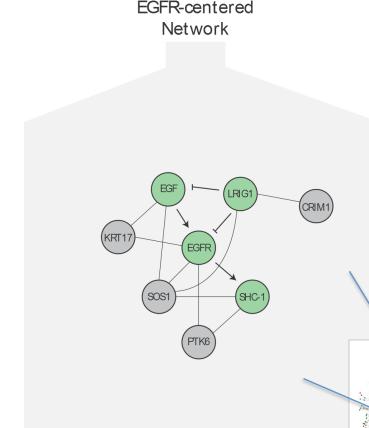


Pathways vs Networks

EGFR-centered Pathway

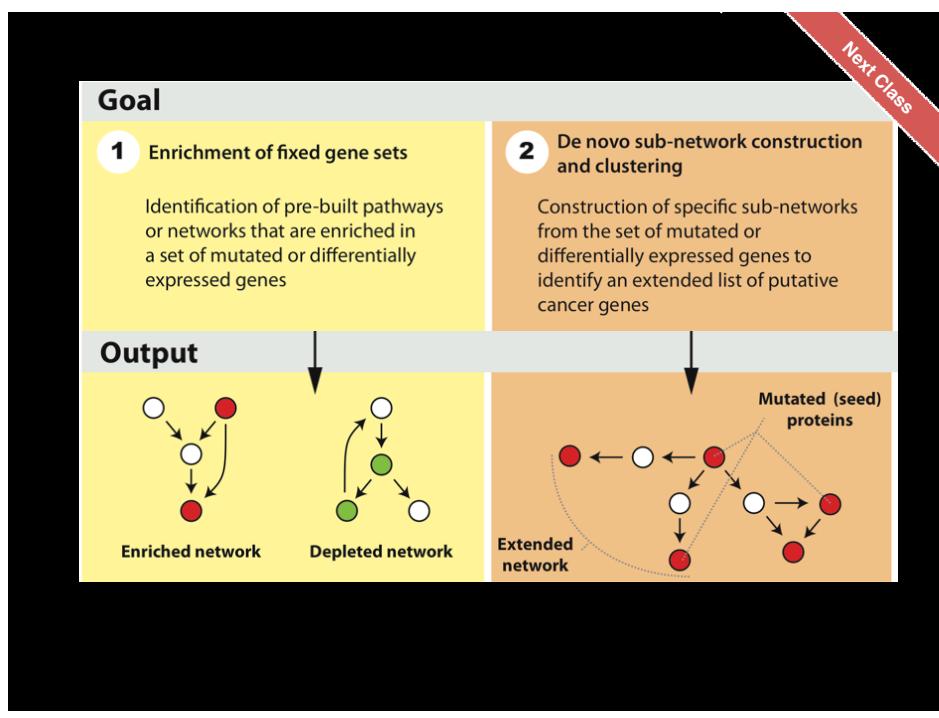


- Detailed, high-confidence consensus
 - Biochemical reactions
 - Small-scale, fewer genes
 - Concentrated from decades of literature



- Simplified cellular logic, noisy
 - Abstractions: directed, undirected
 - Large-scale, genome-wide
 - Constructed from *omics* data integration

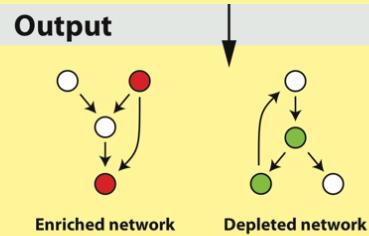
Next Class



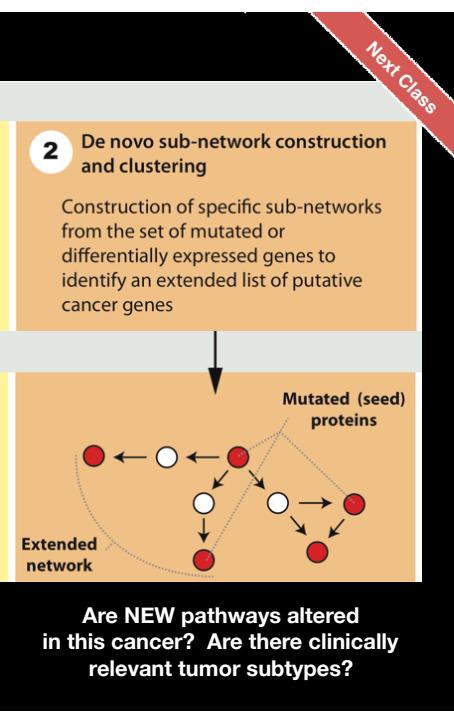
Goal

- ## 1 Enrichment of fixed gene sets

Identification of pre-built pathways or networks that are enriched in a set of mutated or differentially expressed genes



What biological process is altered in this cancer?



Are NEW pathways altered
in this cancer? Are there clinically
relevant tumor subtypes?

Network analysis approaches

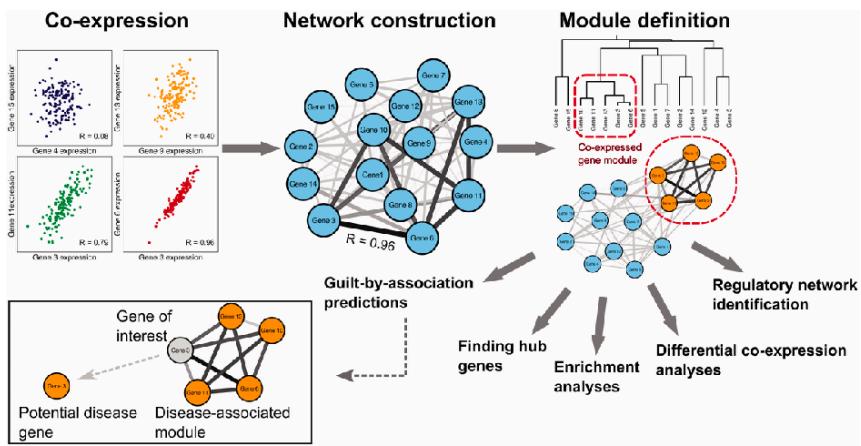


Image from: van Dam et al. (2017) <https://doi.org/10.1093/bib/bbw139>

Next Class

Do it Yourself!

R Quiz time!

https://bioboot.github.io/bggm213_S18/lectures/#15

Also: R Quiz Online