

Guide to Functionally Annotating Your ChIP-chip Data

Outline

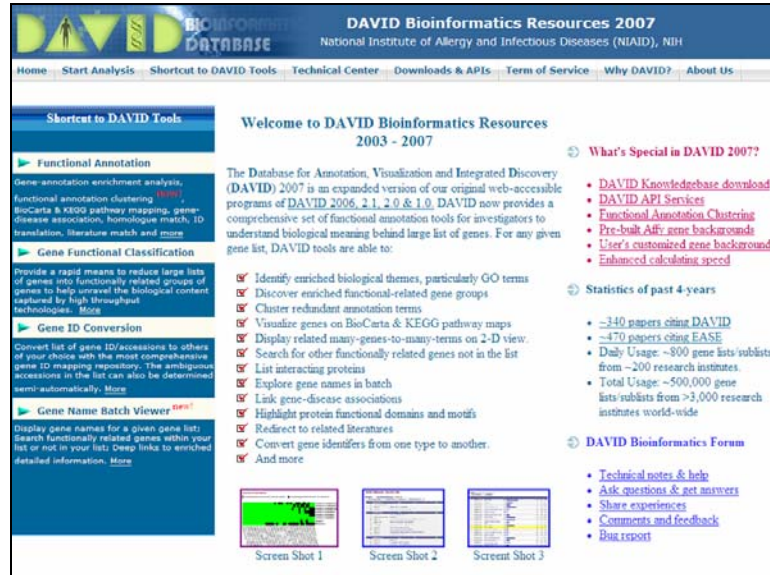
This document describes how to use the Database for Annotation, Visualization and Integrated Discovery (DAVID) to functionally annotate NimbleGen ChIP-chip data.

Step 1. Select the Data to Annotate

1. In Microsoft® Excel®, open one of the following promoter reports from your data DVD:
 - Processed_data_files\Report_All_Peaks\arrayID_ratio_peaks_mapToFeatures_All_Peaks.xls
 - Processed_data_files\Report_Nearest_Peaks\arrayID_ratio_peaks_mapToFeatures_Nearest_Peak.xls
2. (Optional) Sort the promoter report by the most hits:
 - a. Highlight row 1.
 - b. Select **Data** -> **Filter** -> **Auto Filter**.
 - c. Use the **sort** button next to each column header (*FDR*, *peak score*, etc.) to rank the most significant peaks.
Note: To prevent the inclusion of non-informative peaks, apply thresholds of $FDR \leq 0.10$ and $peak\ score \geq 1.0$.
3. Highlight the *ncbi_gene_id* column (column P).
Note: DAVID can analyze up to 3,000 genes. If the *ncbi_gene_id* column contains more than 3,000 gene IDs, highlight $\leq 3,000$ genes to analyze.
4. Select **Edit** -> **Copy** to copy the data in the *ncbi_gene_id* column for subsequent loading into DAVID.

Step 2. Load the Data into DAVID

1. Access the DAVID web site: <http://david.abcc.ncifcrf.gov/home.jsp>



DAVID Bioinformatics Resources 2007
National Institute of Allergy and Infectious Diseases (NIAID), NIH

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

Shortcut to DAVID Tools

- Functional Annotation**
Gene-annotation enrichment analysis, functional annotation clustering, BioCarta & KEGG pathway mapping, gene-disease association, homologous match, ID translation, literature match and more
- Gene Functional Classification**
Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content revealed by high throughput technologies. More
- Gene ID Conversion**
Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. More
- Gene Name Batch Viewer**
Display gene names for a given gene list; Search functionally related genes within your list or not in your list. Deep links to enriched detailed information. More

Welcome to DAVID Bioinformatics Resources 2003 - 2007

The Database for Annotation, Visualization and Integrated Discovery (DAVID) 2007 is an expanded version of our original web-accessible programs of DAVID 2006, 2.1, 2.0 & 1.0. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

- Identify enriched biological themes, particularly GO terms
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathway maps
- Display related many-genes-to-many-terms on 2-D view.
- Search for other functionally related genes not in the list
- List interacting proteins
- Explore gene names in batch
- Link gene-disease associations
- Highlight protein functional domains and motifs
- Redirect to related literatures
- Convert gene identifiers from one type to another
- And more

What's Special in DAVID 2007?

- DAVID Knowledgebase download
- DAVID API Services
- Functional Annotation Clustering
- Pre-built Affy gene backgrounds
- User's customized gene background
- Enhanced calculating speed

Statistics of past 4-years

- ~340 papers citing DAVID
- ~470 papers citing EASE
- Daily Usage ~800 gene lists/sublists from ~200 research institutes
- Total Usage ~500,000 gene lists/sublists from >3,000 research institutes world-wide

DAVID Bioinformatics Forum

- Technical notes & help
- Ask questions & get answers
- Share experiences
- Comments and feedback
- Bug report

Screen Shot 1 | Screen Shot 2 | Screen Shot 3

2. In the left column, click **Functional Annotation**.
3. On the Upload tab:
 - a. In *Step 1: Enter Gene List*, paste (Ctrl + V) the *ncbi_gene_id* column information (Step 1.1 above) into the list box.
 - b. In *Step 2: Select Identifier*, select "ENTREZ_GENE_ID."
 - c. In *Step 3: List Type*, select "Gene List."
 - d. In *Step 4: Submit List*, click **Submit List**.



Functional Annotation Tool
DAVID Bioinformatics Resources 2007, NIAID/NIH

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

Upload | List | Background

Upload Gene List

Demolish Demolish
Upload Help

Step 1: Enter Gene List

A: Paste a list

ncbi_gene_id
3136
3136
3136

Clear

Or

B: Choose From a File

Browse

Step 2: Select Identifier

ENTREZ_GENE_ID

Step 3: List Type

Gene List
Background

Step 4: Submit List

Submit List

Submit your gene list to start the tool!

Tell us how you like the tool
Read technical notes of the tool
Contact us for questions

Key Concepts:

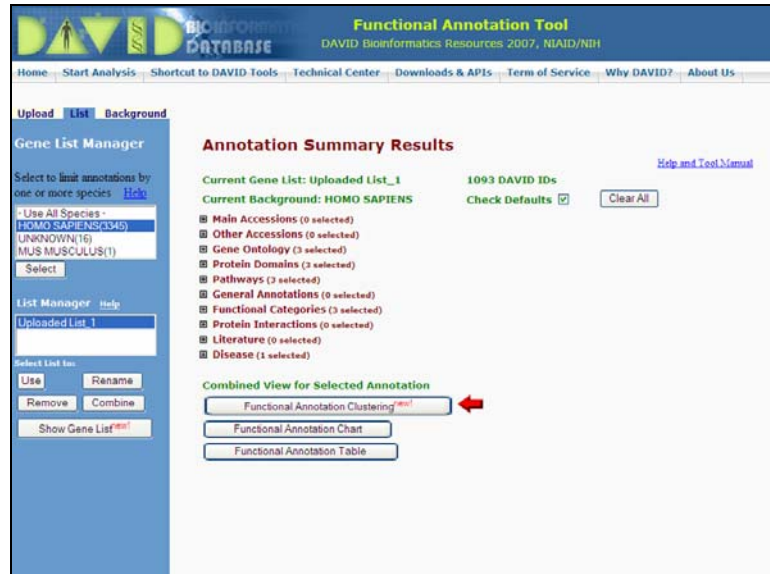
The DAVID Gene Concept
DAVID 2.0 is design around the "DAVID Gene Concept", a graph theory evidence-based method to agglomerate species-specific gene/protein identifiers from a variety of public genome resources including NCBI, PIR and Uniprot/SwissProt. The DAVID Gene Concept method groups tens of million of identifiers from over 65,000 species into 1.5 million unique protein/gene records. More

Term/Gene Co-Occurrence Probability
Ranking functional categories based on co-occurrence with sets of genes in a gene list can rapidly aid in unraveling new biological processes associated with cellular functions and pathways. DAVID 2.0 allows investigators to sort gene categories from dozens of annotation systems. Sorting can be based either the number of genes within each category or by the EASE-score. More

Gene Similarity Search
Any given gene is associating with a set of annotation terms. If genes share similar set of those terms, they are most likely involved in similar biological mechanisms. The algorithm tries to group those related genes based on the agreement of sharing similar annotation terms by Kappa statistics. More

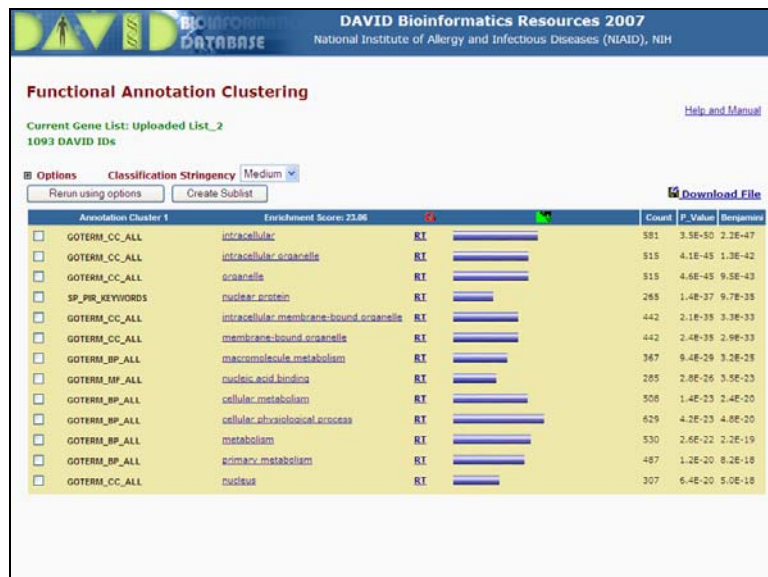
Term Similarity Search
Typically, a biological process/term is done by a corporation of a set of genes. If two or more biological processes are done by similar set of genes, the processes might be related in the biological network somehow. This search function is to identify the related biological processes/terms by quantitatively measuring the degree of the agreement how terms share the similar participating genes. More

- On the List tab, highlight the species and click **Select**.
- In the main body of the window, click **Functional Annotation Clustering** to display the results from your promoter report in a separate browser window.







Step 3. Review Functional Annotation Results

For more information about functional annotation results, click **Help and Manual**.



Annotation Cluster	Enrichment Score: 23.06	Count	P-Value	Benjamin
<input type="checkbox"/> GOTERM_CC_ALL intracellular	RT	581	3.5E-50	2.2E-47
<input type="checkbox"/> GOTERM_CC_ALL intracellular organelle	RT	515	4.1E-45	1.3E-42
<input type="checkbox"/> GOTERM_CC_ALL organelle	RT	515	4.6E-45	9.5E-43
<input type="checkbox"/> SP_PR_KEYWORDS nuclear protein	RT	265	1.4E-37	9.7E-35
<input type="checkbox"/> GOTERM_CC_ALL intracellular membrane-bound organelle	RT	442	2.1E-35	3.3E-33
<input type="checkbox"/> GOTERM_CC_ALL membrane-bound organelle	RT	442	2.4E-35	2.9E-33
<input type="checkbox"/> GOTERM_BP_ALL macromolecule metabolism	RT	347	9.4E-29	3.2E-28
<input type="checkbox"/> GOTERM_MF_ALL nucleic acid binding	RT	285	2.0E-26	3.5E-25
<input type="checkbox"/> GOTERM_BP_ALL cellular metabolism	RT	306	1.4E-23	2.4E-20
<input type="checkbox"/> GOTERM_BP_ALL cellular physiological process	RT	629	4.2E-23	4.0E-20
<input type="checkbox"/> GOTERM_BP_ALL metabolism	RT	530	2.6E-22	2.2E-19
<input type="checkbox"/> GOTERM_BP_ALL primary metabolism	RT	487	1.2E-20	8.2E-18
<input type="checkbox"/> GOTERM_CC_ALL nucleus	RT	307	6.4E-20	5.0E-18

Field	Description
Annotation Cluster	Lists the individual databases DAVID used to functionally annotate genes. DAVID utilizes GO terms to group genes that contribute to that term. Additionally, the method groups related GO terms into a larger cluster. View genes not associated with a cluster by scrolling down the page and clicking N terms were not clustered .

Field	Description
Enrichment Score	Each annotation cluster has an enrichment score. The higher the score, the higher association of your array data with genes in that cluster. To rank the biological significance, the geometric mean (-log) of a GO term's p-value is calculated in a corresponding annotation cluster. Hence, clusters with the highest enrichment score most likely have consistently lower p-values for their GO terms.
	Click to display the complete list of genes that fall within that particular annotation cluster.
 (2D View)	Click to open a 2D view that allows you to visualize the association between a gene and all GO terms within the cluster. All genes present within the cluster are listed on the right and GO terms are listed at the bottom of the view. If a gene is positively associated with a GO term, the box is shaded green. Genes not yet associated with a term are shaded black.
Count	Lists the number of genes associated with the GO term.
P_Value	Lists the score derived from the Modified Fisher Exact P-value. The smaller the value, the more highly associated the genes are with that GO term.
Benjamini	Lists the score corrected for multiple hypothesis testing. The smaller the value, the more highly associated the genes are with that GO term.
<i>annotation cluster term</i> (e.g. intracellular, organelle, nuclear protein, etc.)	Click to display information regarding the particular GO term.
	Click to display the functionally related GO terms (sorted by similarity score) within that particular annotation cluster.
	Click to display a list of all genes present within that GO term. The number of genes in this list is also displayed under the <i>Count</i> column. In addition, links are included for annotation of a particular gene, related genes, and species. To download the entire gene report as text file, click Download File .

References

1. Dennis GJ, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, and Lempicki, RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology* 4(5):P3.
2. Hosack DA, Dennis GJ, Sherman BT, Lane HC, and Lempicki RA. 2003. Identifying Biological Themes within Lists of Genes with EASE. *Genome Biology* 4(6):P4.

Contact Information

Technical Support	(877) NimbleGen / (877) 646-2534 ext. 4 techsupport@nimblegen.com
Customer Services	(877) NimbleGen / (877) 646-2534 ext. 3 customer-services@nimblegen.com
Sales Support (for billing information)	(877) NimbleGen / (877) 646-2534 ext. 1 billing@nimblegen.com