

Bioinformatics 사용법 (1) – Site List

2015. 8.

연세프로테오믹스연구원 정슬기

| Sequence homology search | |
|---|---|
| http://blast.ncbi.nlm.nih.gov/Blast.cgi | 서열 유사도 분석 |
| http://www.ebi.ac.uk/Tools/msa/clustalo/ | Multiple sequence alignment tool |
| http://www.ebi.ac.uk/interpro/ | Amino acid sequence를 이용 domain/motif 검색 |
| Sequence 분석 | |
| http://web.expasy.org/compute_pi/ | sequence의 pI와 mw를 계산해 주는 툴 |
| http://web.expasy.org/peptide_mass/ | tryptic peptide의 mw를 계산 |
| http://web.expasy.org/peptide_cutter/ | Peptide cleavage site 예측 |
| PTM 예측 | |
| http://www.cbs.dtu.dk/services/NetAcet/ | N-acetyltransferase A 의 substrate site 예측 |
| http://www.cbs.dtu.dk/services/NetCGlyc/ | C-mannosylation site 예측 |
| http://www.cbs.dtu.dk/services/NetNGlyc/ | N-Glycosylation site 예측 |
| http://www.cbs.dtu.dk/services/NetPhos/ | Phosphorylation site 예측 |
| Protein localization/topology 예측 | |
| http://www.psорт.org/ | Subcellular localization 예측 |
| http://www.cbs.dtu.dk/services/SignalP/ | Signal peptide cleavage site 예측 |
| http://www.cbs.dtu.dk/services/TMHMM-2.0/ | Trans-membrane domain 예측 |
| Functional annotation tool | |
| https://david.ncifcrf.gov/ | Functional annotation tool |
| Useful sites | |
| http://www.expasy.org/proteomics | Proteomics 관련 tool list |
| http://www.oxfordjournals.org/our_journals/nar/webserver/c/ | Nucleic Acid Research의 web-server issue에 나온 tool list |

Bioinformatics 사용법 (2) – Fasta format

2015. 8.

연세프로테오믹스연구원 정슬기

FASTA format is a text-based format for representing either nucleotide sequences or peptide sequences, in which base pairs or amino acids are represented using single-letter codes. A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line is distinguished from the sequence data by a greater-than (>) symbol in the first column. It is recommended that all lines of text be shorter than 80 characters in length.

1. 많은 bioinformatics tool들이 fasta format의 서열을 input값으로 받음.
2. 많은 경우 "It is recommended that all lines of text be shorter than 80 characters in length." 이 제한은 지킬 필요 없음

Example

```
> title1
MANEVIKCKAAVAWEAGKPLSIEEIEVAPPKAHEVRIKIATAVCHTDAYTLSGADPEGCFPVILGHEGAGIVESVGEVTKLKAGDVIPLYIPQCG
ECKFCLNPKTNLCQKIRVTQGGKGLMPDGTSRFTCKGKILHYMGTSFSEYTVVADISVAKIDPLAPLDKVCLLGCGISTGYGAAVNTAKLEPGSV
CAVFLGGVGLAVIMGCKVAGASRIIGVDINKDKFARAKEFGATECINPQDFSKPIQEVLIEMTDGGVDYSEFCIGNVKVMRAALEACHKGWG
VSVVVGVAASGEEIATRPFLVLTGRTWKGTAFFGGWKSVESVPLVSEYMSKKIKVDEFVTHNLSFDEINKAFELMHSGKSIRTVVKI

>title2
MFAEIQIQDKDRMGTAGKVIKCKAAVLWEQKQPFSEIEIEVAPPKTKEVRIKILATGICRTDDHVIKGTMVSKFPVIVGHEATGIVESIGEGVTTVKP
```

Additional information

https://en.wikipedia.org/wiki/FASTA_format

<http://zhanglab.ccmb.med.umich.edu/FASTA/>

Bioinformatics 사용법 (3) – BLAST

2015. 8.

연세프로테오믹스연구원 정슬기

BLAST for Basic Local Alignment Search Tool is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences

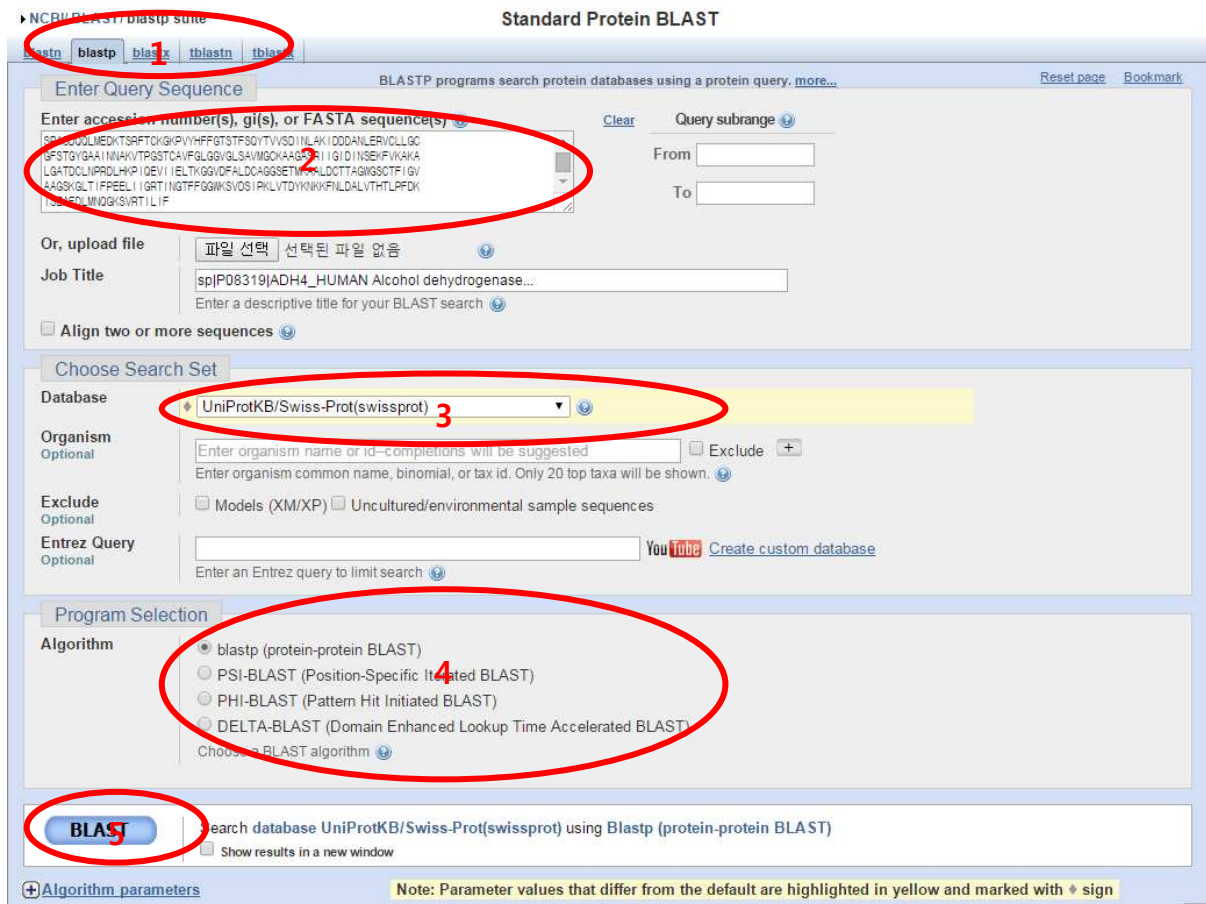
URL: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

Basic BLAST types:

1. blastn (nucleotide blast): Search a nucleotide database using a nucleotide query
2. blastp (protein blast): Search protein database using a protein query
3. blastx: Search protein database using a translated nucleotide query
4. tblastn: Search translated nucleotide database using a protein query
5. tblastx: Search translated nucleotide database using a translated nucleotide query

Example of Standard Protein BLAST

1. blast type 선택
2. query sequence를 입력: fasta format의 서열을 입력
3. database 선택
4. algorithm 선택
5. 실행



References

1. http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs
2. Altschul, SF., Madden, TL., Schäffer, AA., Zhang, J., Zhang, Z. et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs., *Nucleic Acids Res.* **1997**, 25, 3389-402

Bioinformatics 사용법 (4) – ClustalO

2015. 8.

연세프로테오믹스연구원 정슬기

Clustal Omega is a multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences.

URL: <http://www.ebi.ac.uk/Tools/msa/clustalo/>

Example of ADH family alignment

1. 3개 이상의 sequence를 fasta format으로 입력 (e.g)

```
> ADHX_HUMAN
MANEVIKCKAAVAWEAGKPLSIEEIEVAPPKAHEVRIKIATAVCHTDAYTLSGADPEGCFPVILGHEGAGIVESVGEGVTCLKAGDTVIPLYPQCG
ECKFCLNPKTNLCQKIRVTQGGKLMGPDGTSRFTCKGKILHYMGSTFSEYTVVADISVAKIDPLAPLDKVCLLGCGISTGYGAAVNTAKLEPGSV
CAVFLGGVGLAVIMGCKVAGASRIIGVDINKDKFARAKEFGATECINPQDFSKPIQEVLIEMTDGGVDYSFECIGNVKVMRAALEACHKGWG
VSVVVGVAASGEEIATRPFLVLTGRTWKGTAFGGWKSVESVPLVSEYMSKKIKVDEFVTHNLSFDEINKAFELMHSGKSIRTVVKI

> ADH7_HUMAN
MFAEIQIQDKDRMGTAGKVIKCKAAVLWEQKQPFSEIEVAPPKTKVRIKILATGICRTDDHVIKGMTMVSFKFPVIVGHEATGIVESIGEGVTTVKP
GDKVIPLFLPQCRECNACRNPDPGNLCIRSDITGRGVLADGTRFTCKGKPVHFMNTSTFTEYTVVDESSVAKIDDAAPPEKVCCLIGCGFSTGY
GAAVKTGKVKPGSTCVVFLGGVGLSVIMGCKSAGASRIIGIDLNKDKFEKAMAVGATECISPKDSTKPISEVLSEMTGNNVGYTFEIGHLETM
IDALASCHMNYGTSVVVGVPPSAKMLTYDPMMLFTGRTWKGCVFGGLKSRDDVPKLVTEFLAKKFDLDQLITHVLPFKKISEGFELLNSGQSIRT
VLTF

> ADH4_HUMAN
MGTKGKVIKCKAAIAWEAGKPLCIEEIEVAPPKAHEVRIQIATSLCHTDATVIDSKFEGLAFPVIVGHEAAGIVESIGPGVTNVKPGDKVIPLYAPL
CRKCKFCLSPLTNLCGKISNLKSPASDQQLMEDKTSRFTCKGKPVYHFFGTSTFSQYTVVSDINLAKIDDDANLERVCLLGGCFSTGYGAAINNA
KVTPGSTCAVFLGGVGLSAVMGCKAAGASRIIGIDINSEKFKAKALGATDCLNPRDLHKPIQEVIELTKGGVDFALDCAGGSETMKAALDCT
TAGWGSCTFIGVAAGSKGLTIFPEELIIGRTINGTFFGGWKSVDISIPKLVTDYKNKKFNLDALVHTLPLPDKISEAFDLMNQGKSVRILIF
```

2. output format 선택 (optional)

3. 실행

Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

STEP 1 - Enter your input sequences

Enter or paste a set of PROTEIN sequences in any supported format:

```

MGTKGKVKCKAAWWEAGKPLCIEEVEVAPPKAHEVRIQIIATSLCHTDATVIDSKFEG
LAFFPVVHHEAAGIVESIGPGVTNVKPGDKVIPLYAPLCRCKKFLSPLTNLCGKISNLK
SPVSDQQLMEDKTSRFTCKGKPVYHFFGTSTFSQYTVVSDINLAKIDDDANLERVCLLGC
LGFSTGYGAAIINNAKVTGSTCAVFLGGVGLSAVMGCKAAGASRIIGDINSEKFKAKA
LGATDCLNPRDLHKPIQEVIELTKGGVDFALDCAGQSETMKAALDCTTAGWGSCTFIVG
AAGSKGLTFPEELIIGRTINGTFFGGWKSVDISPKLVTDVKNKKNFLDALVTHLTPFDK
ISEAFDLMNQGSVRTILIF
            
```

Or, upload a file: 파일 선택 선택된 파일 없음

STEP 2 - Set your parameters

OUTPUT FORMAT: Clustal w/o numbers

The default settings will fulfil the needs of most users and, for that reason, are not visible.

More options... (Click here, if you want to view or change the default settings.)

STEP 3 - Submit your job

Be notified by email (Tick this box if you want to be notified by email when the results are available)

Submit

3

결과

Alignments
Result Summary
Phylogenetic Tree
Submission Details

Download Alignment File
Hide Colors
Send to ClustalW2_Phylogeny

CLUSTAL O(1.2.1) multiple sequence alignment

```

sp|P40394|ADH7_HUMAN   MFAEIQQIKDKFMGTAGKVIKCKAAWLWEKQPFSEIEEVEVAPPKIIEVRIKILATGICR
sp|P11766|ADHX_HUMAN  -----MANEVIKCKAAWWEAGPLSLSEIEEVEVAPPKAHEVRIKIIATAVCH
sp|P08319|ADH4_HUMAN  -----MGTKGKVIKCKAAWWEAGPLCLIEEVEVAPPKAHEVRIQIIATSLCH
              ***** *  * :*****:*****:** :
sp|P40394|ADH7_HUMAN  TDDHVIKGTMVSKFFPVIHGHEATGIVESIIEGCVTTWPGDKVIILFLPOCFEONACINP
sp|P11766|ADHX_HUMAN  TDAYTLSGADPEEGCFPPVILGHEGAGIVESVGGVTIKLAGDVIILVYIPDCEGCFCLNP
sp|P08319|ADH4_HUMAN  TDATVIDSKFEGLAFFPVIHGHEAAGIVESIIEGCVTTWPGDKVIILVYAPLCRCKKFLSP
** ..... *****:*****:***:* *****: * : * : * : * : *
sp|P40394|ADH7_HUMAN  DGNLCLIRSD----ITGFGVLDGTTIRFTCKGKPVHFMNTSTFTTEYTVDESSVAKIDD
sp|P11766|ADHX_HUMAN  KTNLQKIKRVT----QGIGLMPDGTSRFTCKGKITLHMGTSTFSEYTVVADISVAKIDP
sp|P08319|ADH4_HUMAN  LTNLGKIKISNLKSPASDQDLMEDIKTSRFTCKGKPVYHFFGTSTFSQYTVVSDINLAKIDD
** * : : * : * : * : * : * : * : * : * : * : * : * : * : * : *
sp|P40394|ADH7_HUMAN  AAPPEVCLIGDFSTGVGAAYKTIKVKPGSTCVFGLGGVGLSVIMGCKSAGASRIIGI
sp|P11766|ADHX_HUMAN  LAPLDKVDLLGGISTGVGAAYNTAKLEPGSCAVFGLGGVGLAVIMGCKVAGASRIIGV
sp|P08319|ADH4_HUMAN  DANLERVCLLGGDFSTGVGAAIINNAKVTPGSTCAVFLGGVGLSAVMGCKAAGASRIIGI
* :*****:*****:***:*****:*****:*****:*****:*****:
sp|P40394|ADH7_HUMAN  QLNIDKFEKAMVGAATECISPKDSTPISEVLSEMTGNVGYTFEVI GHLETMIDALASC
sp|P11766|ADHX_HUMAN  DINIKDFARAKEFBATECINPODFKPIDEVLIENTDGGVYVSEFCIIGNVMRAALEAC
sp|P08319|ADH4_HUMAN  DINSEFWAKALGATDQLNPRDLHKPIQEVIELTKGGVDFALDCAGSETMKAALDCT
* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *
sp|P40394|ADH7_HUMAN  HNNVGTSVVVGPPSAKMLTYDPMILLFTGRTIMKGVFGLKSRDDVPKLVTEFLAKFDL
sp|P11766|ADHX_HUMAN  HIGWGVSVVVGVAASGEEIATRPFQLVGTGRTIMKGFAPGGKSVESPKLVSEVMKIKLV
sp|P08319|ADH4_HUMAN  TAGWGSCTFIVGWAAGKGLTIFPEELIIGRTINGTFFGGWKSVDISPKLVTDVKNKKNFL
* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *
sp|P40394|ADH7_HUMAN  DQLI THVLPFKKISEGFELNSGQSI RTVLTFF
sp|P11766|ADHX_HUMAN  DEFVTHLSDFEINKAFELMHSKISIRTVVKI
sp|P08319|ADH4_HUMAN  DALVTHLTPFDKISEAFDLMNQGSVRTILIF
* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *
            
```

References

1. <http://www.ebi.ac.uk/Tools/msa/clustalo/help/>
2. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K. et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* **2011**, 7, 539

Bioinformatics 사용법 (5) – InterProScan

2015. 8.

연세프로테오믹스연구원 정슬기

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites.

URL: <http://www.ebi.ac.uk/interpro/>

Example of ADH6 domain prediction

1. sequence를 fasta format으로 입력
2. 실행

EMBL-EBI Services Research Training About us

InterPro
Protein sequence analysis & classification

Search InterPro... Search

Examples: IPR020405, kinase, P01587, PF02932, GO:0007165

Home Search Release notes Download About InterPro Help Contact

InterPro: protein sequence analysis & classification

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. We combine protein signatures from a number of member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool. [Read more about InterPro](#)

Analyse your protein sequence

```
ELGATE
LNLPQDLKKPIQEVLFDMTDAGIDFCFEA1GNLDVLAALASCNESYGVCVVVG
LPASV
QENSGQLFFSGRSLKGSVFGGWKSRQHIPKLVADYMAEKLNLDP LITHLNLDKI
NEAV
ELMPTGAW
```

Search | Clear | Example protein sequence

Documentation Protein focus Publications

v.53 InterPro 53.0
23rd July 2015

Features include:

- An update to Pfam (28.0).
- Integration of 166 new methods from the CATH-Gene3D (3), Pfam (3), SUPERFAMILY (7) and PANTHER (153) databases.

Download | [Read more](#)

IDA ■

Domain architecture search Search >>

Interproscan 5

[Learn more >>](#)

결과

The screenshot displays the InterProScan results for protein P28332. The page is divided into several sections:

- Overview:** Similar proteins, Structures.
- Filter view on:** Entry type (Family, Domains, Repeats, Site), Status (Unintegrated), Colour by (domain relationship, source database).
- Protein:** P28332, Length: 368 amino acids.
- Protein family membership:** A table showing domain matches. A red circle highlights this section, which includes a table with columns for domain ID, description, and position. The table shows matches for GroES-like (IPR011032) and Alcohol dehydrogenase, N-terminal (IPR013154).
- Detailed signature matches:** A section showing matches for IPR002085 (Alcohol dehydrogenase superfamily, zinc-type) and IPR028633 (Alcohol dehydrogenase family, zinc-type, class-V subfamily).

The 'Export' button and 'Select format' dropdown menu are circled in red. The domain membership section is also circled in red.

1. domain search 결과 위에 마우스 포인터를 올려두면 간략한 설명이 나옴 (클릭하면 InterPro database로 이동)
2. export 명령을 이용 원하는 format의 파일로 저장 가능 (tsv, xml, etc...)

References

1. <http://www.ebi.ac.uk/interpro/training.html>
2. Jones, P., Binns, D., Chang, HY., Fraser, M., Li, W. et al., InterProScan 5: genome-scale protein function classification. *Bioinformatics*. **2014**, 30, 1236-40

Bioinformatics 사용법 (6) – DAVID functional annotation tool

2015. 8.

연세프로테오믹스연구원 정슬기

The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (**DAVID**) 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: (1) Identify enriched biological themes, particularly GO terms; (2) Discover enriched functional-related gene groups; (3) Cluster redundant annotation terms; (4) Visualize genes on BioCarta & KEGG pathway maps; (5) Display related many-genes-to-many-terms on 2-D view; (6) Search for other functionally related genes not in the list; (7) List interacting proteins; (8) Explore gene names in batch; (9) Link gene-disease associations; (10) Highlight protein functional domains and motifs; (11) Redirect to related literatures; (12) Convert gene identifiers from one type to another; (13) And more

URL: <https://david.ncifcrf.gov/>

사용법

1. Click "start analysis" or "Functional Annotation"
2. Submit gene/protein list
3. Select identifier
4. Select list type (Gene list)
5. 실행

DAVID Bioinformatics Resources 6.7
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, gene-ontology, gene functional annotation clustering, BioCarta & KEGG pathway mapping, gene-disease association, homologue 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 captured 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](#)

Recommending: A [paper](#) published in *Nature Protocols* describes step-by-step procedure to use DAVID!

Welcome to DAVID 6.7

2003 - 2015

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 is an update to the sixth version of our original web-accessible programs. 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,

What's Important in DAVID?

- [Current \(v 6.7\) release note](#)
- [New requirement to cite DAVID](#)
- [IDs of Affy Exon and Gene arrays supported](#)
- [Novel Classification Algorithms](#)
- [Pre-built Affymetrix and Illumina backgrounds](#)
- [User's customized gene background](#)
- [Enhanced calculating speed](#)

Statistics of DAVID

Upload List Background

Upload Gene List

[Demolist 1](#) [Demolist 2](#)

[Upload Help](#)

Step 1: Enter Gene List

A: Paste a list

1438_at
1487_at
1494_f_at
1498_g_at

Or

B: Choose From a File

파일 선택 선택된 파일 없음

Multi-List File ?

Step 2: Select Background

AFFYMETRIX_EXON_GENE_ID

Step 3: List Type

Gene List **4**

Background

Step 4: Submit List

Submit List **5**

Functional Annotation Tool

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 6.7 is designed around the "DAVID Gene Concept", a graph theory evidence-based method to agglomerate species-specific gene/protein identifiers from a variety of public genomic 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 6.7 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](#)

결과

Gene List Manager

Select to limit annotations by one or more species [Help](#)

- Use All Species -
Homo sapiens(351)

Select Species

List Manager [Help](#)

List_1

Select List to:

Use Rename
Remove Combine
Show Gene List

Annotation Summary Results [Help and Tool Manual](#)

Current Gene List: List_1
Current Background: Homo sapiens
349 DAVID IDs
Check Defaults Clear All

- Disease (1 selected)
- Functional_Categories (3 selected)
- Gene_Ontology (3 selected)
- General_Annotations (0 selected)
- Literature (0 selected)
- Main_Accessions (0 selected)
- Pathways (3 selected)
- Protein_Domains (3 selected)
- Protein_Interactions (0 selected)
- Tissue_Expression (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

Functional Annotation Clustering
Functional Annotation Chart
Functional Annotation Table

1. 확인하고 싶은 annotation 만 선택하여 분석 가능
2. e.g., gene ontology 의 Biological process 와 Molecular function 만 선택한 경우

Select to limit annotations by one or more species [Help](#)

- Use All Species -
Homo sapiens(351)

Select Species

List Manager [Help](#)

List_1

Select List to:

Use Rename
Remove Combine
Show Gene List

Current Gene List: List_1
Current Background: Homo sapiens
349 DAVID IDs
Check Defaults Clear All

- Disease (0 selected)
- Functional_Categories (0 selected)
- Gene_Ontology (2 selected)

| Gene Ontology Term | Percentage | Count | Chart |
|---|------------|-------|-------|
| GOTERM_BP_1 | 88.5% | 309 | Chart |
| GOTERM_BP_2 | 88.5% | 309 | Chart |
| GOTERM_BP_3 | 87.7% | 306 | Chart |
| GOTERM_BP_4 | 86.5% | 302 | Chart |
| GOTERM_BP_5 | 83.1% | 290 | Chart |
| GOTERM_BP_ALL | 88.5% | 309 | Chart |
| <input checked="" type="checkbox"/> GOTERM_BP_FAT | 86.8% | 303 | Chart |
| GOTERM_CC_1 | 94.3% | 329 | Chart |
| GOTERM_CC_2 | 94.0% | 328 | Chart |
| GOTERM_CC_3 | 94.0% | 328 | Chart |
| GOTERM_CC_4 | 91.4% | 319 | Chart |
| GOTERM_CC_5 | 90.8% | 317 | Chart |
| GOTERM_CC_ALL | 94.3% | 329 | Chart |
| <input checked="" type="checkbox"/> GOTERM_CC_FAT | 80.2% | 280 | Chart |
| GOTERM_MF_1 | 89.1% | 311 | Chart |
| GOTERM_MF_2 | 88.3% | 308 | Chart |
| GOTERM_MF_3 | 77.1% | 269 | Chart |
| GOTERM_MF_4 | 69.3% | 242 | Chart |
| GOTERM_MF_5 | 57.0% | 199 | Chart |
| GOTERM_MF_ALL | 89.1% | 311 | Chart |
| <input checked="" type="checkbox"/> GOTERM_MF_FAT | 78.2% | 273 | Chart |
| PANTHER_BP_ALL | 77.4% | 270 | Chart |
| PANTHER_MF_ALL | 76.5% | 267 | Chart |

3. annotation cluster 결과

88 Cluster(s) [Download File](#)

| Annotation Cluster 1 | | Enrichment Score: 6.07 | | | Count | P_Value | Benjamini |
|--------------------------|---------------|--|----|--|-------|---------|-----------|
| <input type="checkbox"/> | GOTERM_BP_FAT | RNA processing | RT | | 34 | 2.0E-7 | 3.9E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | mRNA processing | RT | | 25 | 2.3E-7 | 2.2E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | mRNA metabolic process | RT | | 27 | 2.4E-7 | 1.6E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | RNA splicing, via transesterification reactions with bulged adenosine as nucleophile | RT | | 16 | 1.7E-6 | 8.0E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | RNA splicing, via transesterification reactions | RT | | 16 | 1.7E-6 | 8.0E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | nuclear mRNA splicing, via spliceosome | RT | | 16 | 1.7E-6 | 8.0E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | RNA splicing | RT | | 21 | 6.0E-6 | 1.4E-3 |
| Annotation Cluster 2 | | Enrichment Score: 4.46 | | | Count | P_Value | Benjamini |
| <input type="checkbox"/> | GOTERM_BP_FAT | DNA metabolic process | RT | | 29 | 9.3E-6 | 2.0E-3 |
| <input type="checkbox"/> | GOTERM_BP_FAT | cellular response to stress | RT | | 31 | 1.0E-5 | 2.0E-3 |
| <input type="checkbox"/> | GOTERM_BP_FAT | DNA replication | RT | | 16 | 2.4E-5 | 4.1E-3 |
| <input type="checkbox"/> | GOTERM_BP_FAT | response to DNA damage stimulus | RT | | 22 | 9.9E-5 | 1.6E-2 |
| <input type="checkbox"/> | GOTERM_BP_FAT | DNA repair | RT | | 18 | 2.2E-4 | 2.8E-2 |
| Annotation Cluster 3 | | Enrichment Score: 3.18 | | | Count | P_Value | Benjamini |
| <input type="checkbox"/> | GOTERM_BP_FAT | in utero embryonic development | RT | | 14 | 1.6E-4 | 2.4E-2 |
| <input type="checkbox"/> | GOTERM_BP_FAT | chordate embryonic development | RT | | 18 | 1.3E-3 | 8.9E-2 |
| <input type="checkbox"/> | GOTERM_BP_FAT | embryonic development ending in birth or egg hatching | RT | | 18 | 1.4E-3 | 9.4E-2 |

4. Annotation chart 결과

216 chart records [Download File](#)

| Sublist | Category | Term | RT | Genes | Count | % | P-Value | Benjamini |
|--------------------------|---------------|--|----|-------|-------|------|---------|-----------|
| <input type="checkbox"/> | GOTERM_BP_FAT | RNA processing | RT | | 34 | 9.7 | 2.0E-7 | 3.9E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | mRNA processing | RT | | 25 | 7.2 | 2.3E-7 | 2.2E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | mRNA metabolic process | RT | | 27 | 7.7 | 2.4E-7 | 1.6E-4 |
| <input type="checkbox"/> | GOTERM_MF_FAT | RNA binding | RT | | 37 | 10.6 | 9.7E-7 | 5.3E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | RNA splicing, via transesterification reactions with bulged adenosine as nucleophile | RT | | 16 | 4.6 | 1.7E-6 | 8.0E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | nuclear mRNA splicing, via spliceosome | RT | | 16 | 4.6 | 1.7E-6 | 8.0E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | RNA splicing, via transesterification reactions | RT | | 16 | 4.6 | 1.7E-6 | 8.0E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | protein localization | RT | | 43 | 12.3 | 2.5E-6 | 9.5E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | protein transport | RT | | 39 | 11.2 | 2.7E-6 | 8.6E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | establishment of protein localization | RT | | 39 | 11.2 | 3.4E-6 | 9.2E-4 |
| <input type="checkbox"/> | GOTERM_BP_FAT | RNA splicing | RT | | 21 | 6.0 | 6.0E-6 | 1.4E-3 |
| <input type="checkbox"/> | GOTERM_BP_FAT | DNA metabolic process | RT | | 29 | 8.3 | 9.3E-6 | 2.0E-3 |
| <input type="checkbox"/> | GOTERM_BP_FAT | cellular response to stress | RT | | 31 | 8.9 | 1.0E-5 | 2.0E-3 |
| <input type="checkbox"/> | GOTERM_BP_FAT | DNA replication | RT | | 16 | 4.6 | 2.4E-5 | 4.1E-3 |
| <input type="checkbox"/> | GOTERM_MF_FAT | transcription factor binding | RT | | 26 | 7.4 | 7.7E-5 | 2.1E-2 |
| <input type="checkbox"/> | GOTERM_BP_FAT | response to DNA damage stimulus | RT | | 22 | 6.3 | 9.9E-5 | 1.6E-2 |
| <input type="checkbox"/> | GOTERM_BP_FAT | in utero embryonic development | RT | | 14 | 4.0 | 1.6E-4 | 2.4E-2 |
| <input type="checkbox"/> | GOTERM_MF_FAT | helicase activity | RT | | 12 | 3.4 | 1.7E-4 | 3.1E-2 |
| <input type="checkbox"/> | GOTERM_BP_FAT | negative regulation of macromolecule metabolic process | RT | | 33 | 9.5 | 2.2E-4 | 3.0E-2 |
| <input type="checkbox"/> | GOTERM_BP_FAT | DNA repair | RT | | 18 | 5.2 | 2.2E-4 | 2.8E-2 |
| <input type="checkbox"/> | GOTERM_MF_FAT | nucleotide binding | RT | | 71 | 20.3 | 2.6E-4 | 3.5E-2 |

5. Download file 기능을 이용하여 tap separated text file 로 down load 가능 (download file 을 click 하면 txt file 이 web-browser 에 바로 열림으로 마우스 오른쪽 버튼 click 하여 “다른이름으로 링크저장”을 사용하는 것이 편함)

References

1. https://david.ncifcrf.gov/helps/functional_annotation.html#summary
2. Huang, da W., Sherman, BT., Lempicki, RA., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* **2009**, 4, 44-57
3. Dennis, G., Sherman, BT., Hosack, DA., Yang, J., Gao, W. et al., DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **2003**, 4, P3