TUTORIAL

Revised in Apr 2015

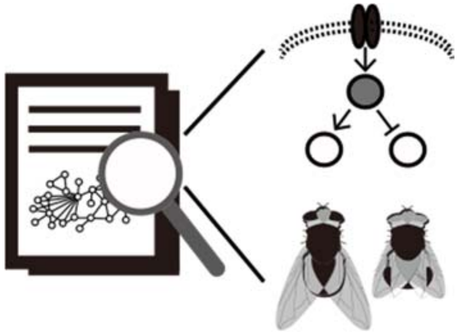
Contents

I. Overview

FlyNet is a network prioritization server that adapts several different prediction pipelines to generate various types of hypothesis during *Drosophila melanogaster*-based researches. FlyNet web server consists of 3 kinds of prioritization options; (1) Function prioritization, (2) Gene prioritization and (3) Human disease prioritization. Also, the integrated gene network and its component sub-networks are provided as text files in the “Download Networks” page.

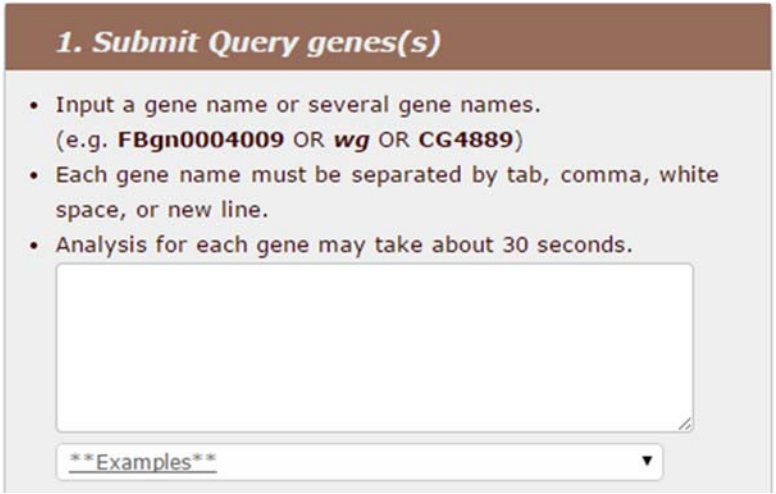
To use the FlyNet prioritization server, click the “Network Prioritization” on the header menu of the web page. Basically, the working procedure of each prioritization option is performed in 3 steps: (1) gene submission, (2) fine-tuning of prioritization parameters, and (3) analysis of reported data.

II. Fly prioritizer - Function prioritization



In "Function prioritization" option, you can expect **the inferences of prioritized biological function and associated RNAi phenotypes** for the uncharacterized genes, which are based on the network analysis.

1. Submit query gene(s)

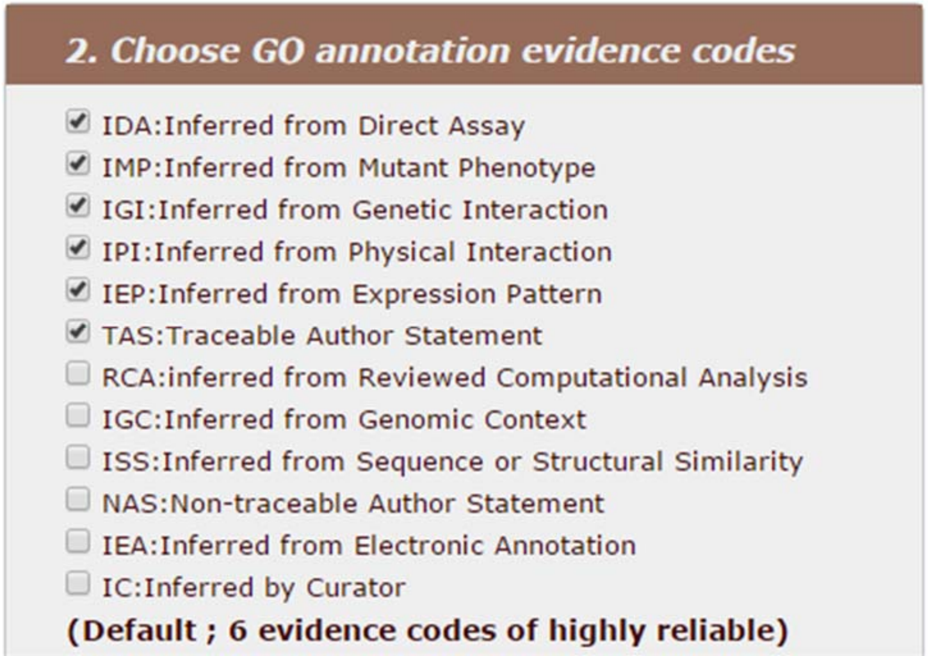


1. Submit Query genes(s)

- Input a gene name or several gene names.
  (e.g. **FBgn0004009** OR ***wg*** OR **CG4889**)
- Each gene name must be separated by tab, comma, white space, or new line.
- Analysis for each gene may take about 30 seconds.

Insert your query gene(s) in the text box. Acceptable gene IDs could be almost all kinds of *D. melanogaster* genes. The gene names must be separated by tab, comma, white space or new line. A maximum of 300 genes are available in a single run. You can also try our example query genes for a test run.

2. Select GO annotation evidence codes by clicking them.



2. Choose GO annotation evidence codes

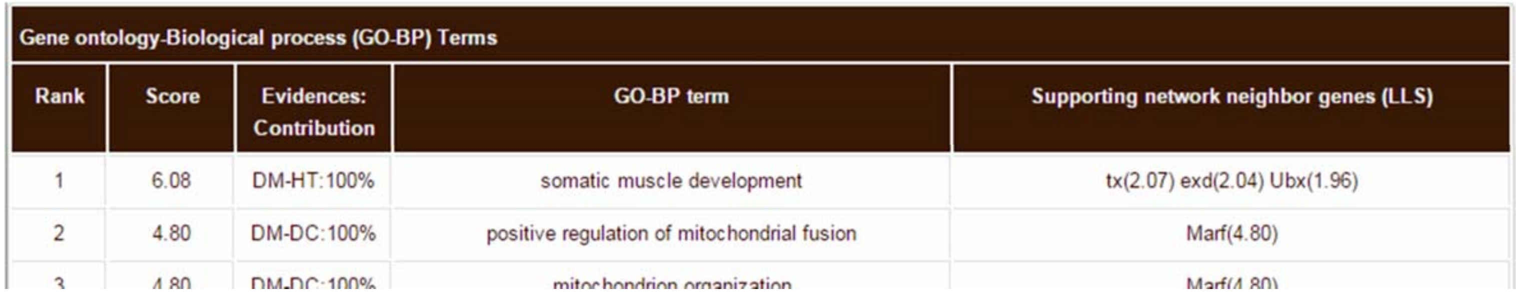
- ☑ IDA:Inferred from Direct Assay
- ☑ IMP:Inferred from Mutant Phenotype
- ☑ IGI:Inferred from Genetic Interaction
- ☑ IPI:Inferred from Physical Interaction
- ☑ IEP:Inferred from Expression Pattern
- ☑ TAS:Traceable Author Statement
- ☐ RCA:inferred from Reviewed Computational Analysis
- ☐ IGC:Inferred from Genomic Context
- ☐ ISS:Inferred from Sequence or Structural Similarity
- ☐ NAS:Non-traceable Author Statement
- ☐ IEA:Inferred from Electronic Annotation
- ☐ IC:Inferred by Curator

(Default ; 6 evidence codes of highly reliable)

Only GO terms from selected evidences will be shown in the following result. You can see the full name of evidences by putting the mouse-pointer on them. Default 6 evidence codes, which are based on experiments and traceable author statements, hence highly reliable.

3. Analysis of reported data

i. Predicted GO-BP terms are listed in rank order.



Gene ontology-Biological process (GO-BP) Terms

Rank	Score	Evidences: Contribution	GO-BP term	Supporting network neighbor genes (LLS)
1	6.08	DM-HT:100%	somatic muscle development	tx(2.07) exd(2.04) Ubx(1.96)
2	4.80	DM-DC:100%	positive regulation of mitochondrial fusion	Marf(4.80)
3	4.80	DM-DC:100%	mitochondrion organization	Marf(4.80)

① Rank: Guilt-by-association principle prioritizes inferred GO-BP terms of the query gene by assessing the connectivity to the neighbor genes and their GO-BP terms.

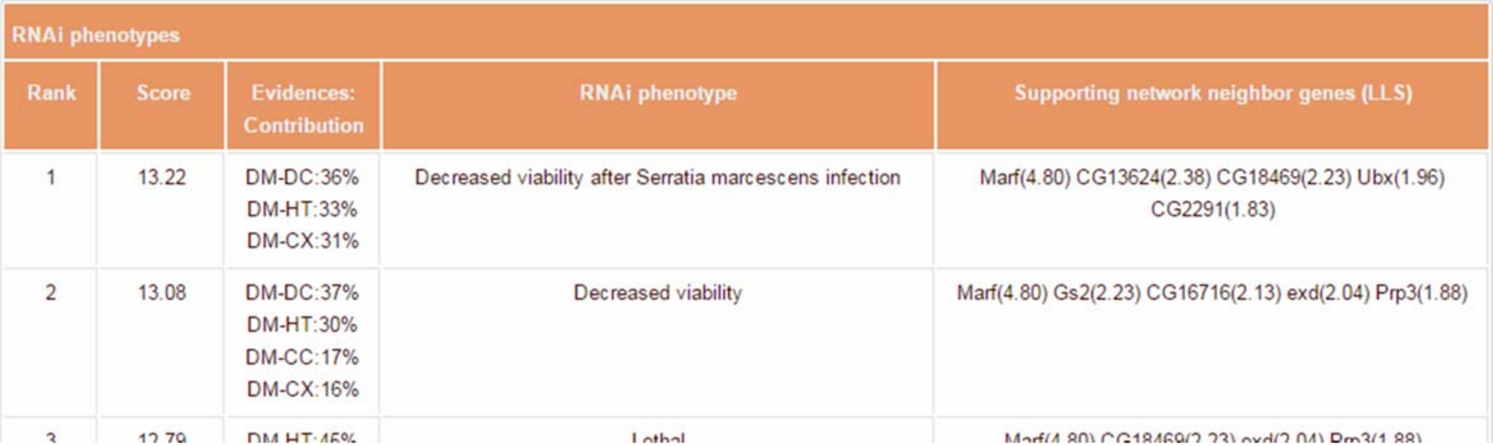
② Score: Sum of LLS (Log Likelihood Scores) of network links to neighboring genes with the specific GO-BP term. For instance, score of 1st-ranked GO-term would be sum of links to all neighboring genes that have the GO-term "*somatic muscle development*".

③ Evidences: list of subnetworks (see network download page) supporting network-based prediction and their fraction of contribution.

④ GO-BP terms: inferred GO biological process annotations of the queried genes. Only GO-BP terms from selected annotation evidence code when submitting the genes will be shown.

⑤ Supporting network neighbor genes: neighbor genes with GO-BP terms from selected annotation evidences and their contribution. (LLS score of the link to the query gene)

ii. Predicted RNAi phenotypes are listed in rank order



RNAi phenotypes

Rank	Score	Evidences: Contribution	RNAi phenotype	Supporting network neighbor genes (LLS)
1	13.22	DM-DC:36% DM-HT:33% DM-CX:31%	Decreased viability after Serratia marcescens infection	Marf(4.80) CG13624(2.38) CG18469(2.23) Ubx(1.96) CG2291(1.83)
2	13.08	DM-DC:37% DM-HT:30% DM-CC:17% DM-CX:16%	Decreased viability	Marf(4.80) Gs2(2.23) CG16716(2.13) exd(2.04) Prp3(1.88)
3	12.79	DM-HT:45%	Lethal	Marf(4.80) CG18469(2.23) exd(2.04) Prp3(1.88)

① Rank: Guilt-by-association principle prioritizes inferred RNAi phenotypes of the query gene by assessing the connectivity to the neighbor genes and their RNAi phenotypes.

② Score: Sum of LLS (Log Likelihood Scores) of network links to neighboring genes with the specific RNAi phenotype. For instance, score of 1st-ranked RNAi phenotype would be sum of links to all neighboring genes that have the RNAi phenotype; “Decreased viability after Serratia marcesces infection”.

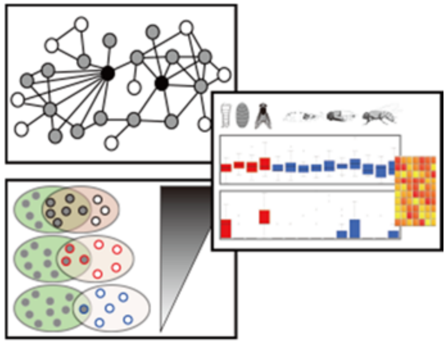
③ Evidences: list of subnetworks (see network download page) supporting network-based prediction and their fraction of contribution.

④ RNAi phenotype: inferred RNAi phenotype that associated with the query genes.

⑤ Supporting network neighbor genes: neighbor genes with RNAi phenotypes and their contribution. (LLS score of the link to the query gene)

TOP ↑

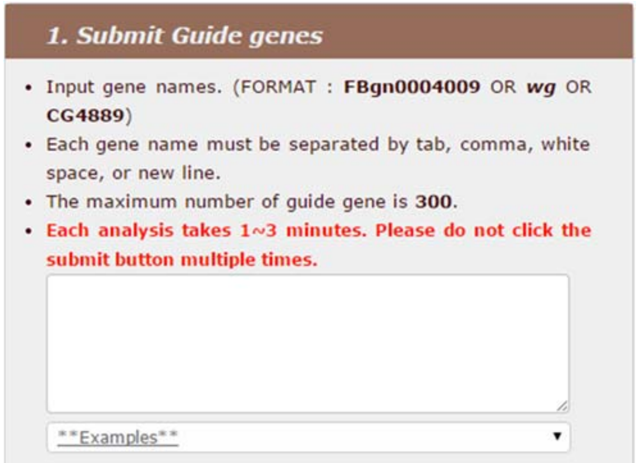
III. Fly prioritizer - Gene prioritization



In “Gene prioritization” option, you can expect various types of network analysis data about ‘guide genes’ and candidate genes. Reporting analysis results are as follows:

- Estimated prediction power of guide genes within the network (AUC score)
- Distribution of spatiotemporal state-specific gene expression and network links
- Network graph of guide and candidate genes (Direct neighborhood case only)
- Prioritized list of guide and candidate genes and those network evidences
- Further ‘Gene set analysis’ of various functional annotations

1. Submit Guide genes

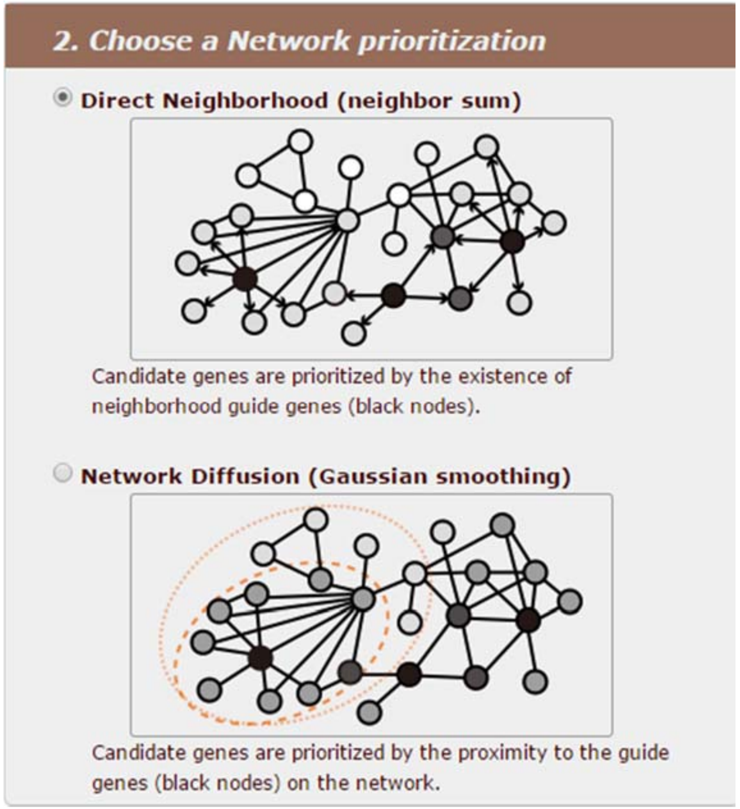


1. Submit Guide genes

- Input gene names. (FORMAT : **FBgn0004009** OR ***wg*** OR **CG4889**)
- Each gene name must be separated by tab, comma, white space, or new line.
- The maximum number of guide gene is **300**.
- **Each analysis takes 1~3 minutes. Please do not click the submit button multiple times.**

Insert your ‘guide genes’ in the text box. **‘Guide genes’ could be a set of genes that are known to participate in certain biological pathways or that lead to certain phenotypic outcomes**. Acceptable gene IDs could be almost all kinds of *D. melanogaster* genes. The gene names must be separated by tab, comma, white space or new line. A maximum of 300 genes are available in a single run. You can also try our example guide genes for a test run.

2. Choose a network prioritization algorithm



2. Choose a Network prioritization

● **Direct Neighborhood (neighbor sum)**

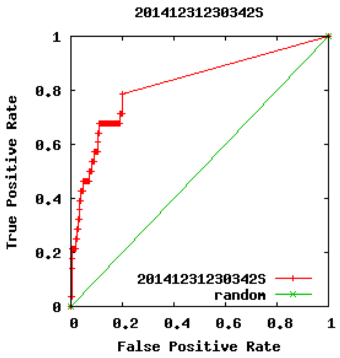
Candidate genes are prioritized by the existence of neighborhood guide genes (black nodes).

○ **Network Diffusion (Gaussian smoothing)**

Candidate genes are prioritized by the proximity to the guide genes (black nodes) on the network.

FlyNet web server allows 2 kinds of network prioritization algorithm; ‘direct neighborhood (neighbor sum)’ and ‘network diffusion (Gaussian smoothing)’. **‘Direct neighborhood’ method prioritizes guide and candidate genes by the existence of neighborhood guide genes**, whereas **‘Network diffusion’ method prioritizes the guide and candidate genes by the proximity to the guide genes on the network**. Each method reports slightly different prioritized lists of both guide genes and candidate genes.

3. Analysis of reported data

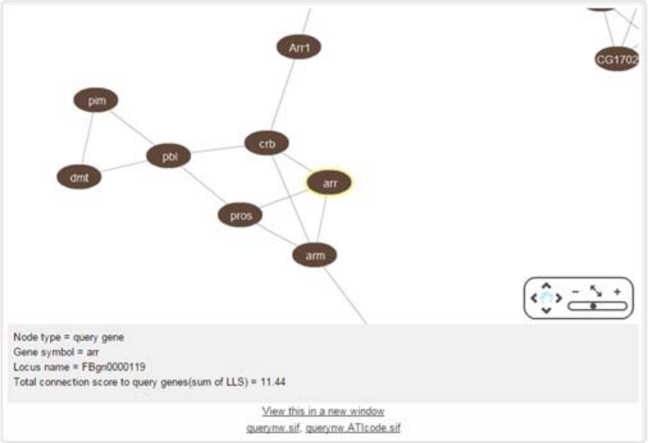


i. Guide gene analysis – AUC (Area Under ROC Curve) score by guide genes

The prediction power of the network model could be estimated by measuring the inter-connectivity among guide genes in the network. This could be assessed by drawing ROC (Receiver Operating Characteristic) curve, therefore AUC score represents a regularized prediction power. In general, the relationship between the prediction power and AUC score is as below:

- AUC = 0.5 : random prediction
- AUC = 1.0 : perfect prediction
- AUC > 0.7 : Indicates significantly high prediction

ii. Guide gene analysis – Genes connection in FlyNet (‘Direct neighborhood’ method only)

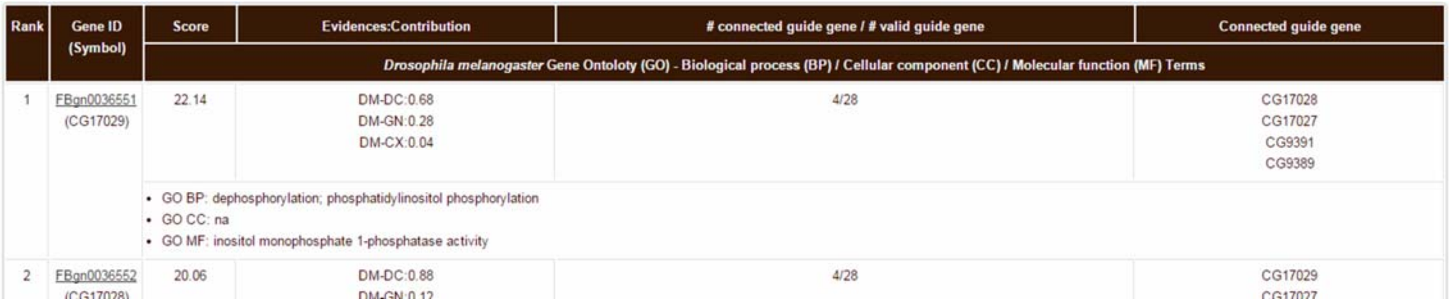


Node type = query gene
Gene symbol = arr
Locus name = FBgn0000119
Total connection score to query genes(sum of LLS) = 11.44

The connection of guide genes is visualized by Cytoscape. This requires the installation of ‘adobe flash player’. Clicking on node or edge will present you the information.

iii. Guide gene analysis – List of guide genes, prioritized based on each algorithms

Each network prioritization algorithm shows slightly different resultant table. In **‘Direct neighborhood’** case,



Rank	Gene ID (Symbol)	Score	Evidences:Contribution	# connected guide gene / # valid guide gene	Connected guide gene
		Drosophila melanogaster Gene Ontoloty (GO) - Biological process (BP) / Cellular component (CC) / Molecular function (MF) Terms			
1	FBgn0036551 (CG17029)	22.14	DM-DC:0.68 DM-GN:0.28 DM-CX:0.04	4/28	CG17028 CG17027 CG9391 CG9389
		• GO BP: dephosphorylation; phosphatidylinositol phosphorylation • GO CC: na • GO MF: inositol monophosphate 1-phosphatase activity			
2	FBgn0036552 (CG17028)	20.06	DM-DC:0.88 DM-GN:0.12	4/28	CG17029 CG17027

① Rank: Guilt-by-association principle prioritizes guide genes by their connectivity to other guide genes.

② Gene ID: Gives the name of query genes. Symbol represents alternative common names. Clicking the names leads to the FlyBase for detailed descriptions

③ Score: Sum of LLS (Log Likelihood Scores) of network links to all other guide genes.

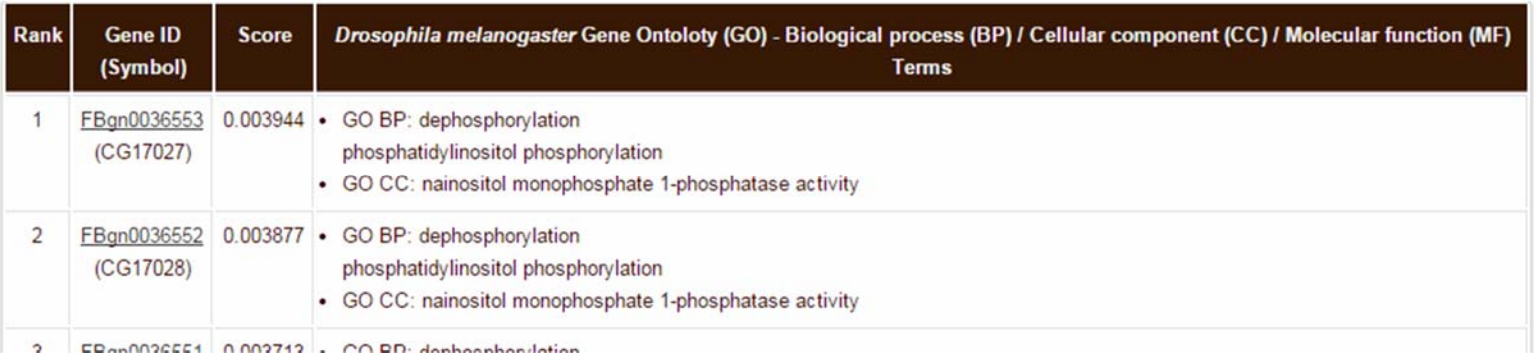
④ Evidences & Contribution: shows the fraction of links from component networks (see network download page) supporting network-based prediction.

⑤ #connected guide gene/ # valid guide gene: The number of linked guide genes over the number of total valid guide genes. ‘Valid’ genes are *D. melanogaster* genes that are known as protein coding genes.

⑥ Connected guide gene: shows a list of linked guide genes.

⑦ GO terms: Corresponding GO-BP/CC/MF annotations that can predict the role of the gene

Whereas in **'Network diffusion' case**, the table is much simpler.



Rank	Gene ID (Symbol)	Score	*Drosophila melanogaster* Gene Ontoloty (GO) - Biological process (BP) / Cellular component (CC) / Molecular function (MF) Terms
1	FBgn0036553 (CG17027)	0.003944	• GO BP: dephosphorylation phosphatidylinositol phosphorylation • GO CC: nainositol monophosphate 1-phosphatase activity
2	FBgn0036552 (CG17028)	0.003877	• GO BP: dephosphorylation phosphatidylinositol phosphorylation • GO CC: nainositol monophosphate 1-phosphatase activity

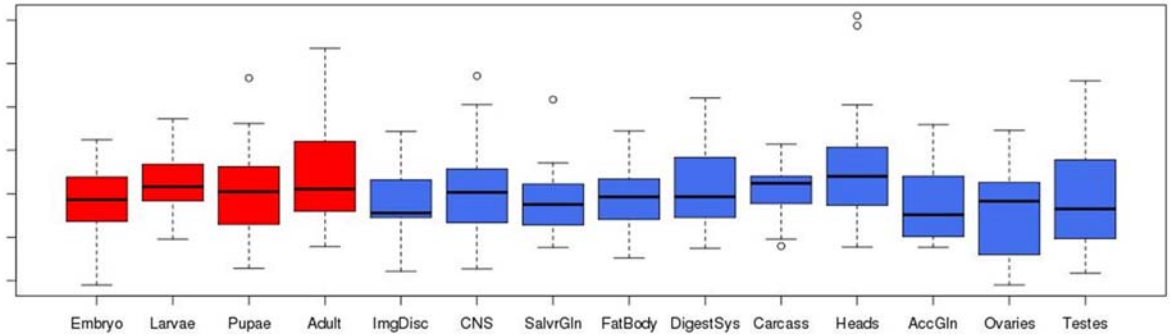
① Rank: Gaussian smoothing prioritizes guide genes by propagating their connectivity to other guide genes.

② Gene ID: Gives the name of query genes. Symbol represents alternative common names. Clicking the names leads to the FlyBase for detailed descriptions

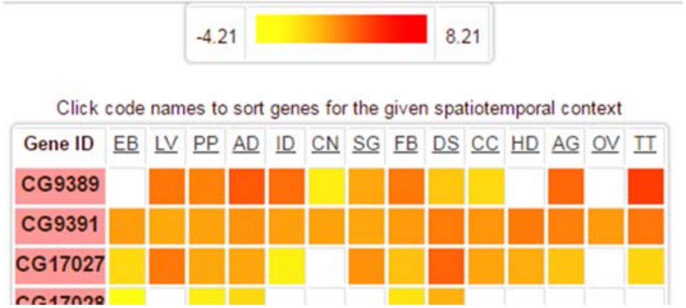
③ Score: Gaussian smoothing score by network diffusion.

④ GO terms: Corresponding GO-BP/CC/MF annotations that can predict the role of the gene

iv. Candidate gene analysis – Spatiotemporal state-specific expression and network links of guide genes



Powered by spatiotemporal state-specific expression data from modENCODE (MDver.3), ‘Gene prioritization’ provides **the distribution of gene expressions of guide genes across 4 developmental stages (red boxes) and 10 tissues (blue boxes)**, which can make an overview of the gene expressions in whole dimensional states of *D. melanogaster*. As the specific patterns of gene expressions are easily interfered by house-keeping gene expressions, ‘Gene prioritization’ also provides **the distribution of links of 14 STN (SpatioTemporal sub-Networks, 4 developmental stages and 10 tissues)**, which is more robust to the existence of house-keeping gene expressions.

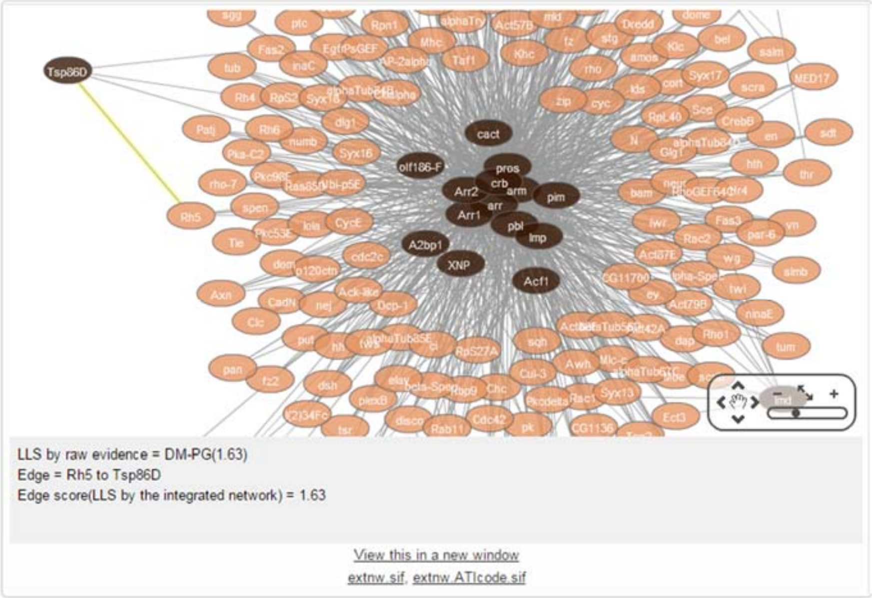


Click code names to sort genes for the given spatiotemporal context

Gene ID	EB	LV	PP	AD	ID	CN	SG	FB	DS	CC	HD	AG	OV	TT
CG9389														
CG9391														
CG17027														
CG17028														

Click on the graph or hyperlink provides **the heat-map of corresponding distribution of gene expression or specific links of STN** in a new window. This heat-map is **re-sortable by click the name of the column of table**, i.e. you can rearrange the order of genes by the spatiotemporal context. Your guide genes are highlighted as bold with red background. The degree of the cell is a relative expression/link counts value and white cell means ‘NA (Not Available)’.

v. Candidate gene analysis – Guide and candidate genes connection in FlyNet (‘Direct neighborhood’ method only)



LLS by raw evidence = DM-PG(1.63)
Edge = Rh5 to Tsp86D
Edge score(LLS by the integrated network) = 1.63

View this in a new window
extnw.sif, extnw.ATIcode.sif

Candidate genes (light nodes) that are connected (grey lines) to guide genes (dark nodes) are visualized in cytoscape. You can get information of edges or nodes by clicking them

vi. Candidate gene analysis – List of top 100 candidate genes, prioritized based on each algorithms

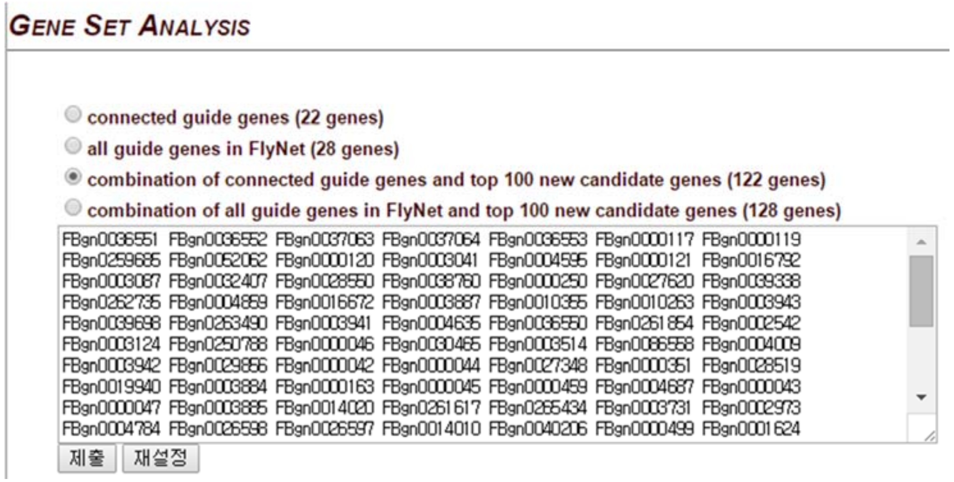
Same as guide gene case except the tables are made from top 100 candidates. See ‘Guide gene analysis – List of guide genes, prioritized based on each algorithms’.

TOP ↑

4. Gene Set Analysis

'Gene Set Analysis' allows an expanded functional analysis of various annotations databases with guide genes and candidate genes combinations. Annotation terms enriched by a queried gene set are reported as prioritized orders. The enrichment scores are calculated by the p-values from hypergeometric test.

i. Submit a gene set



GENE SET ANALYSIS

○ connected guide genes (22 genes)
○ all guide genes in FlyNet (28 genes)
● combination of connected guide genes and top 100 new candidate genes (122 genes)
○ combination of all guide genes in FlyNet and top 100 new candidate genes (128 genes)

```
FBgn0036551 FBgn0036552 FBgn0037063 FBgn0037064 FBgn0036553 FBgn0000117 FBgn0000119
FBgn0259685 FBgn0052062 FBgn0000120 FBgn0003041 FBgn0004595 FBgn0000121 FBgn0016792
FBgn0003087 FBgn0032407 FBgn0028550 FBgn0038760 FBgn0000250 FBgn0027620 FBgn0039338
FBgn0262735 FBgn0004859 FBgn0016672 FBgn0003887 FBgn0010355 FBgn0010263 FBgn0003943
FBgn0039698 FBgn0263490 FBgn0003941 FBgn0004635 FBgn0036550 FBgn0261854 FBgn0002542
FBgn0003124 FBgn0250788 FBgn0000046 FBgn0030465 FBgn0003514 FBgn0086558 FBgn0004009
FBgn0003942 FBgn0029856 FBgn0000042 FBgn0000044 FBgn0027348 FBgn0000351 FBgn0028519
FBgn0019940 FBgn0003884 FBgn0000163 FBgn0000045 FBgn0000459 FBgn0004687 FBgn0000043
FBgn0000047 FBgn0003885 FBgn0014020 FBgn0261617 FBgn0265434 FBgn0003731 FBgn0002973
FBgn0004784 FBgn0026598 FBgn0026597 FBgn0014010 FBgn0040206 FBgn0000499 FBgn0001624
```

The input text box is at the end of the 'Gene prioritization' report page. You can input a gene set to the text box by selecting a prepared combinatorial gene set which consist of guide and candidate genes used in previous network analysis.

ii. Analysis of reported data

Annotation terms from 4 databases will be used; (1) **GO-BP (Gene Ontology – Biological Process) terms**, (2) **KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway terms**, (3) **BioCyc pathway terms**, and (4) **RNAi phenotypes from GenomeRNAi database.** All report table have a same structure.

Rank	ID	Description	p-value	N	m	n	k
1	GO:0007399	nervous system development	1.669e-20	13942	122	101	19
2	GO:0000910	cytokinesis	5.956e-16	13942	122	68	14
3	GO:0001745	compound eye morphogenesis	9.597e-15	13942	122	82	14

① Rank: Rank number of corresponding annotation term.

② ID: Given ID from each database

③ Description: Description provided from each database

④ p-value: from hypergeometric test

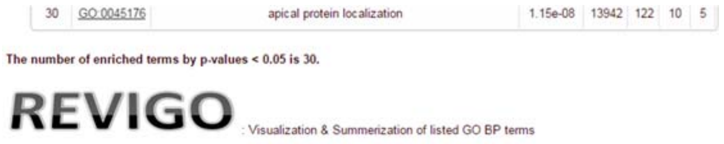
⑤ N: total number of protein coding genes of *D. melanogaster*

⑥ m: number of genes in queried gene set

⑦ n: number of genes consisting each annotations terms

⑧ k: number of genes that are relevant to both the queried gene set and annotation term

Especially in GO-BP term analysis, whole list of enriched terms could be summarized and visualized readily by the application named ‘REVIGO’. Clicking on the icon leads to the REVIGO homepage in a new window

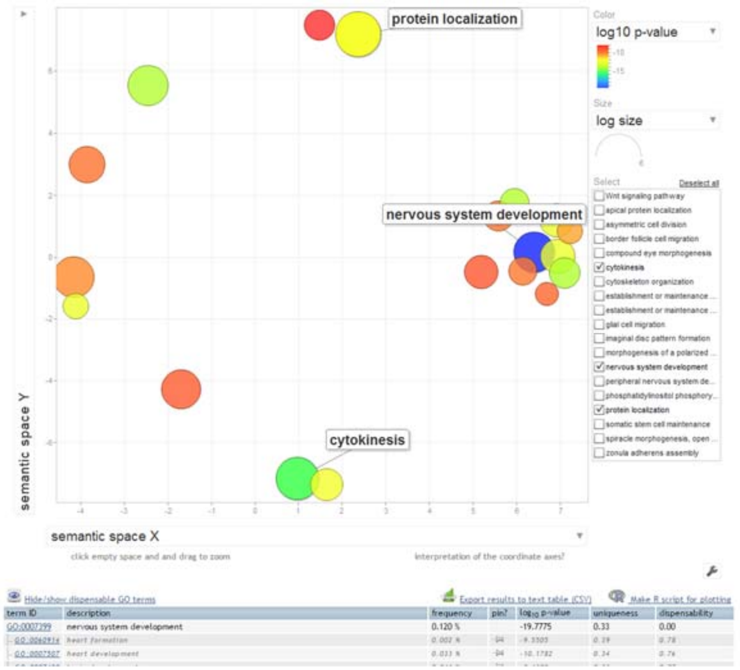


30	GO:0045176	apical protein localization	1.15e-08	13942	122	10	5

The number of enriched terms by p-values < 0.05 is 30.

REVIGO : Visualization & Summerization of listed GO BP terms

< An example result from REVIGO application >



semantic space X

click empty space and and drag to zoom

interpretation of the coordinate axes?

Color: log10 p-value

Size: log size

Select Deselect all
☐ Wnt signaling pathway
☐ apical protein localization
☐ asymmetric cell division
☐ border follicle cell migration
☐ compound eye morphogenesis
☑ cytokinesis
☐ cytoskeleton organization
☐ establishment or maintenance ...
☐ establishment or maintenance ...
☐ glial cell migration
☐ imaginal disc pattern formation
☐ morphogenesis of a polarized ...
☑ nervous system development
☐ peripheral nervous system de...
☐ phosphatidylinositol phosphory...
☑ protein localization
☐ somatic stem cell maintenance
☐ spiracle morphogenesis, open ...
☐ zonula adherens assembly

Hide/show dispensable GO terms   Export results to text table (CSV)   Make R script for plotting

term ID	description	frequency	pin?	$\log_{10}$ p-value	uniqueness	dispensability
GO:0007399	nervous system development	0.120 %		-19.7775	0.33	0.00
- GO:0060914	*heart formation*	0.002 %		-9.3505	0.39	0.78
- GO:0007507	*heart development*	0.033 %		-10.1782	0.34	0.76

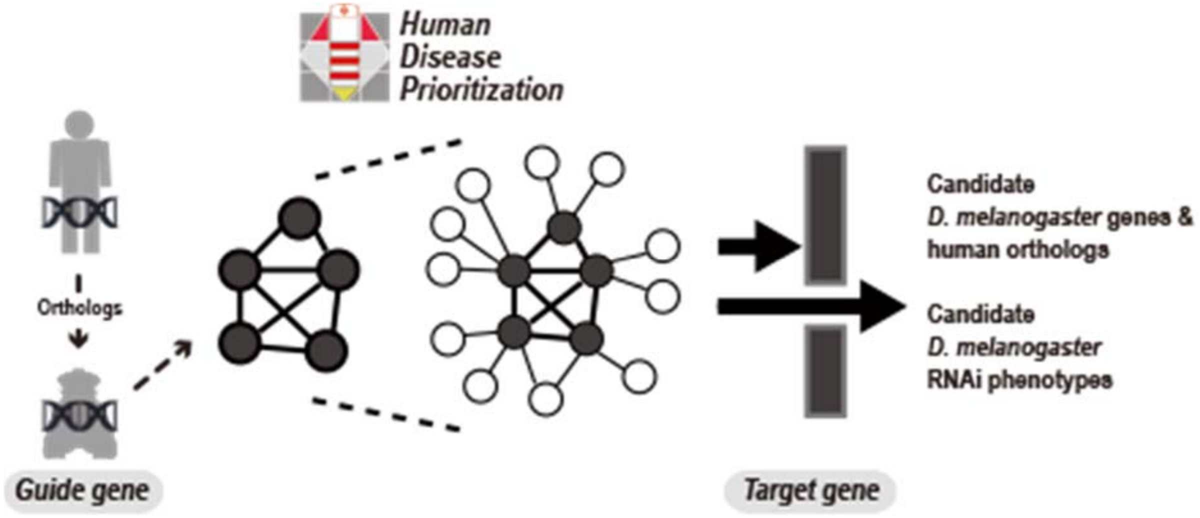
.

TOP ↑

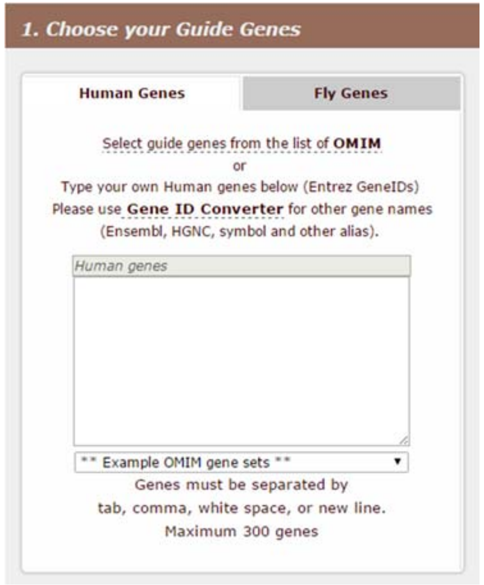
IV. Human prioritizer – Human disease prioritization

‘Human disease prioritization’ option attempts to suggest an alternative to investigate human diseases by means of *D. melanogaster* orthologs and gene perturbation phenotypes. You can expect **prioritized *D. melanogaster* orthologs for genetic analysis** and **associated RNAi phenotype for succeeding experiments**.

‘Human disease prioritization’ starts with the submission of ‘guide genes’ which are related to human diseases. Once the ‘guide genes’ are queried, the following procedure of network analysis runs similar to ‘gene prioritization (with direct neighborhood)’ of Fly prioritizer. Apropos to the candidate *D. melanogaster* genes for guide genes, ‘human disease prioritization’ allows to **confine resultant candidates to a certain set of genes that are gathered for a biological meaning, named ‘target genes’**. Setting ‘target genes’ may make possible to increase the probability of discovering true-positive candidates and to find out the associated RNAi phenotypes more accurately.

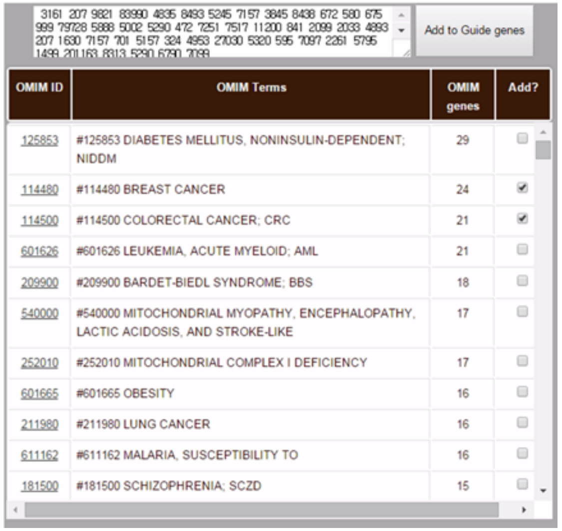


1. Submission of guide genes



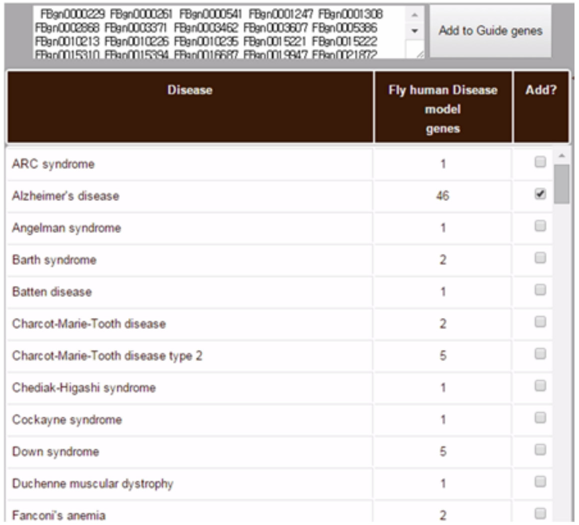
Select guide genes from the list of **OMIM**
or
Type your own Human genes below (Entrez GeneIDs)
Please use **Gene ID Converter** for other gene names
(Ensembl, HGNC, symbol and other alias).

Genes must be separated by
tab, comma, white space, or new line.
Maximum 300 genes



```
3161 207 9821 83990 4835 8493 5245 7157 3845 8438 672 580 675
999 79728 5888 5002 5290 472 7251 7517 11200 841 2099 2033 4893
207 1630 7157 701 5157 324 4953 27030 5320 595 7097 2261 5795
```

OMIM ID	OMIM Terms	OMIM genes	Add?
125853	#125853 DIABETES MELLITUS, NONINSULIN-DEPENDENT; NIDDM	29	☐
114480	#114480 BREAST CANCER	24	☑
114500	#114500 COLORECTAL CANCER; CRC	21	☑
601626	#601626 LEUKEMIA, ACUTE MYELOID; AML	21	☐
209900	#209900 BARDET-BIEDL SYNDROME; BBS	18	☐
540000	#540000 MITOCHONDRIAL MYOPATHY, ENCEPHALOPATHY, LACTIC ACIDOSIS, AND STROKE-LIKE	17	☐
252010	#252010 MITOCHONDRIAL COMPLEX I DEFICIENCY	17	☐
601665	#601665 OBESITY	16	☐
211980	#211980 LUNG CANCER	16	☐
611162	#611162 MALARIA, SUSCEPTIBILITY TO	16	☐
181500	#181500 SCHIZOPHRENIA; SCZD	15	☐



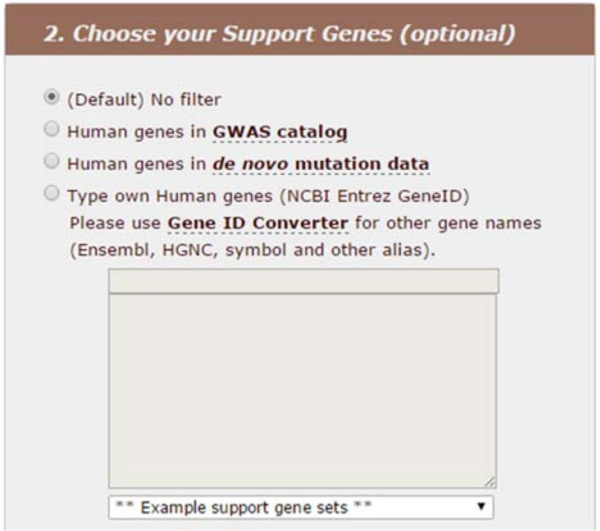
```
FBgn0000229 FBgn0000261 FBgn0000541 FBgn0001247 FBgn0001308
FBgn0002868 FBgn0003371 FBgn0003462 FBgn0003607 FBgn0005386
FBgn0010213 FBgn0010226 FBgn0010235 FBgn0015221 FBgn0015222
```

Disease	Fly human Disease model genes	Add?
ARC syndrome	1	☐
Alzheimer's disease	46	☑
Angelman syndrome	1	☐
Barth syndrome	2	☐
Batten disease	1	☐
Charcot-Marie-Tooth disease	2	☐
Charcot-Marie-Tooth disease type 2	5	☐
Chediak-Higashi syndrome	1	☐
Cockayne syndrome	1	☐
Down syndrome	5	☐
Duchenne muscular dystrophy	1	☐
Fanconi's anemia	2	☐

‘Guide genes’ for ‘Human disease prioritization’ could be **either human disease genes or those *D. melanogaster* orthologs**. Selection of the species of guide genes to be queried are made by selecting the tab above the text box. **Guide genes can be loaded automatically to the text box by selecting the disease name from the lists in the pop-up window**, which were prepared from ‘OMIM morbidmap’ for human gene IDs and ‘FlyBase human disease model alleles’ for *D. melanogaster* gene IDs. **You can also type your own guide genes into the text box**. Please use the ‘gene ID converter’ to submit conventional human gene names properly.

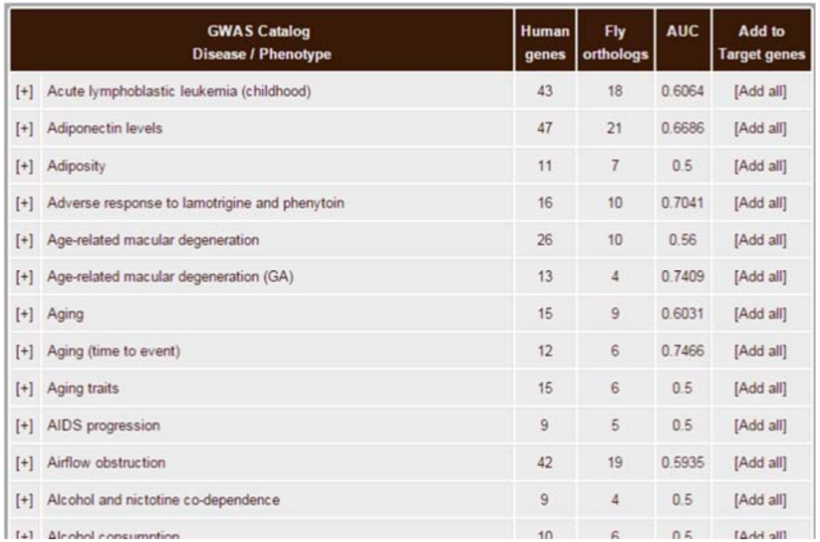
Please confirm that whether the species of the text box is matched with the species of queried gene IDs. The gene names must be separated by tab, comma, white space or new line. A maximum of 300 genes are available in a single run.

2. Submission of target genes



2. Choose your Support Genes (optional)

● (Default) No filter
○ Human genes in **GWAS catalog**
○ Human genes in ***de novo* mutation data**
○ Type own Human genes (NCBI Entrez GeneID)
Please use **Gene ID Converter** for other gene names (Ensembl, HGNC, symbol and other alias).



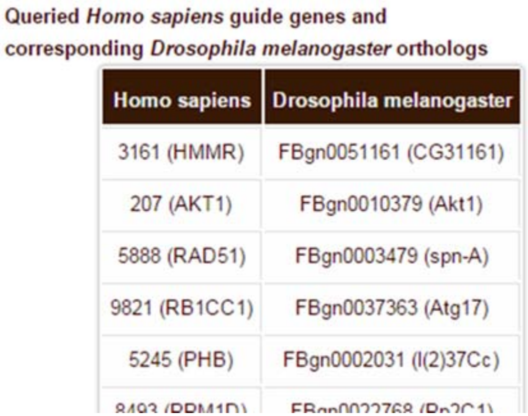
	GWAS Catalog Disease / Phenotype	Human genes	Fly orthologs	AUC	Add to Target genes
[+]	Acute lymphoblastic leukemia (childhood)	43	18	0.6064	[Add all]
[+]	Adiponectin levels	47	21	0.6686	[Add all]
[+]	Adiposity	11	7	0.5	[Add all]
[+]	Adverse response to lamotrigine and phenytoin	16	10	0.7041	[Add all]
[+]	Age-related macular degeneration	26	10	0.56	[Add all]
[+]	Age-related macular degeneration (GA)	13	4	0.7409	[Add all]
[+]	Aging	15	9	0.6031	[Add all]
[+]	Aging (time to event)	12	6	0.7466	[Add all]
[+]	Aging traits	15	6	0.5	[Add all]
[+]	AIDS progression	9	5	0.5	[Add all]
[+]	Airflow obstruction	42	19	0.5935	[Add all]
[+]	Alcohol and nictotine co-dependence	9	4	0.5	[Add all]
[+]	Alcohol consumption	10	6	0.5	[Add all]

In ‘Total FlyNet genes’ option, ‘target genes’ will be the whole *D. melanogaster* genes which have connections to guide genes. Also, **target genes can be confined to the human GWAS (Genome-Wide Association Studies) candidates. Selected target genes that are related to the corresponding diseases/phenotypes will be loaded automatically to the text box by clicking “[Add all]” of the table in the new pop-up window**. To help your choice of GWAS candidates, the list provides counts of human genes, counts of corresponding *D. melanogaster* orthologs and the prediction power made by orthologs in FlyNet (AUC). Some GWAS terms in the list are combination of gene lists from several studies. You can check those references (provided as PubMed ID) and add those genes separately by clicking “[+]” to activate the ‘drop-down menu’. Finally, **you can also type your own target genes of you interest into the text box**.

Please confirm that whether the species of the text box is matched with the species of queried gene IDs. The gene names must be separated by tab, comma, white space or new line. A maximum of 300 genes are available in a single run.

3. Analysis of reported data

i. Information of orthologs



Queried *Homo sapiens* guide genes and corresponding *Drosophila melanogaster* orthologs

Homo sapiens	Drosophila melanogaster
3161 (HMMR)	FBgn0051161 (CG31161)
207 (AKT1)	FBgn0010379 (Akt1)
5888 (RAD51)	FBgn0003479 (spn-A)
9821 (RB1CC1)	FBgn0037363 (Atg17)
5245 (PHB)	FBgn0002031 (l(2)37Cc)
8493 (PPM1D)	FBgn0022768 (Pp2C1)

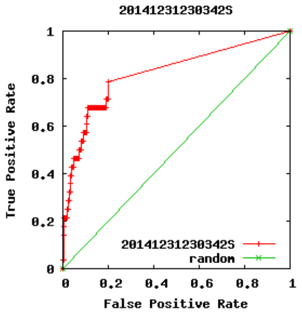
Guide genes invalid in FlyNet

10 genes without *Drosophila melanogaster* ortholog
675, 79728, 999, 83990, 5002, 4835, 5290, 841, 2099, 580

2 kinds of orthologs information are reported. In the first table, queried human guide gene IDs and conventional names are matched to corresponding *D. melanogaster* ortholog gene IDs and symbol names. Only genes that are known as protein coding genes are shown here. The second table is the list of human guide genes that does not have protein coding-*D. melanogaster* orthologs.

If user queried *D. melanogaster* genes as guide genes instead of human genes, this report will not be appeared.

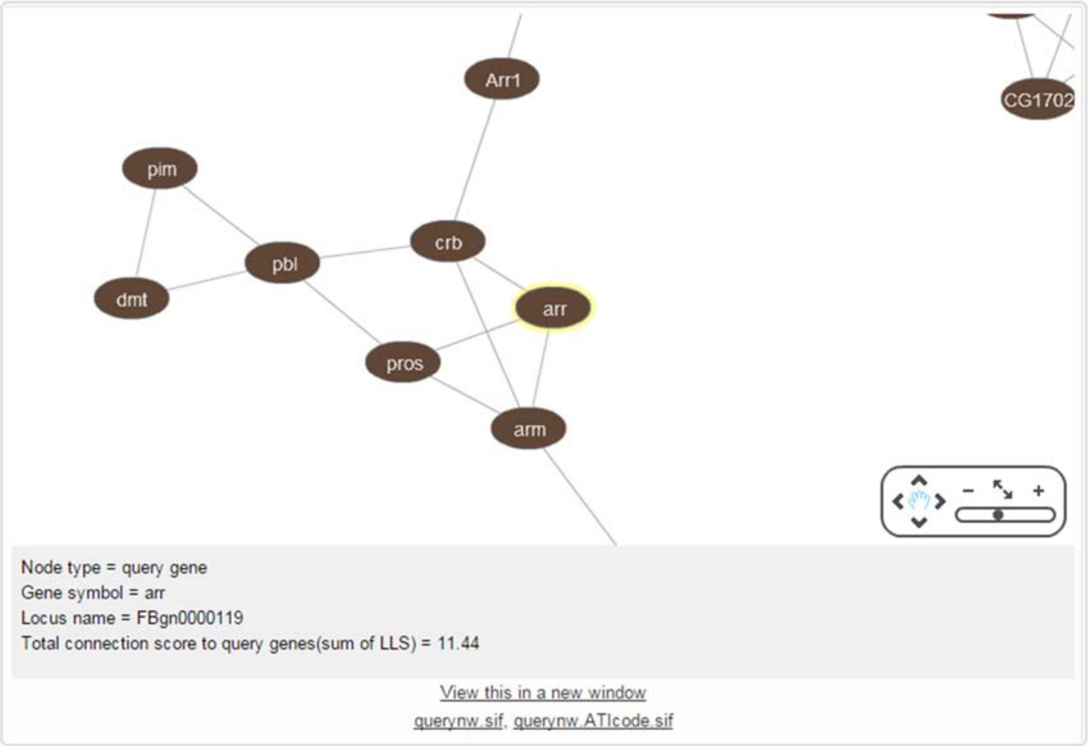
ii. Guide gene analysis - AUC (Area Under ROC Curve) score by guide genes



The prediction power of the network model could be estimated by measuring the inter-connectivity among guide genes in the network. This could be assessed by drawing ROC (Receiver Operating Characteristic) curve, therefore AUC score represents a regularized prediction power. In general, the relationship between the prediction power and AUC score is as below:

- AUC = 0.5 : random prediction
- AUC = 1.0 : perfect prediction
- AUC > 0.7 : Indicates significantly high prediction

iii. Guide gene analysis - Guide genes' (orthologs) connection in FlyNet



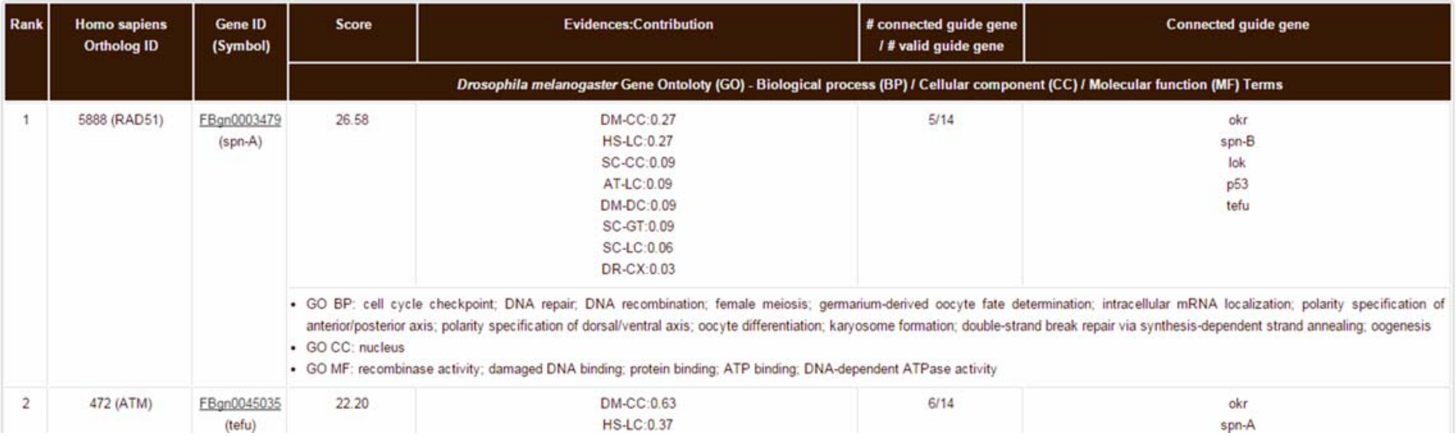
Node type = query gene
Gene symbol = arr
Locus name = FBgn0000119
Total connection score to query genes(sum of LLS) = 11.44

View this in a new window

querynw.sif, querynw.ATIcode.sif

The connections of guide genes (orthologs) are visualized by Cytoscape. This requires the installation of 'adobe flash player'. Clicking on node or edge will present you the information.

iv. Guide genes analysis - List of guide genes, prioritized by neighbor sum



Rank	Homo sapiens Ortholog ID	Gene ID (Symbol)	Score	Evidences:Contribution	# connected guide gene / # valid guide gene	Connected guide gene
			***Drosophila melanogaster* Gene Ontoloty (GO) - Biological process (BP) / Cellular component (CC) / Molecular function (MF) Terms**			
1	5888 (RAD51)	FBgn0003479 (spn-A)	26.58	DM-CC:0.27 HS-LC:0.27 SC-CC:0.09 AT-LC:0.09 DM-DC:0.09 SC-GT:0.09 SC-LC:0.06 DR-CX:0.03	5/14	okr spn-B lok p53 tefu
			• GO BP: cell cycle checkpoint; DNA repair; DNA recombination; female meiosis; germarium-derived oocyte fate determination; intracellular mRNA localization; polarity specification of anterior/posterior axis; polarity specification of dorsal/ventral axis; oocyte differentiation; karyosome formation; double-strand break repair via synthesis-dependent strand annealing; oogenesis • GO CC: nucleus • GO MF: recombinase activity; damaged DNA binding; protein binding; ATP binding; DNA-dependent ATPase activity			
2	472 (ATM)	FBgn0045035 (tefu)	22.20	DM-CC:0.63 HS-LC:0.37	6/14	okr spn-A

① Rank: Guilt-by-association principle prioritizes guide genes by their connectivity to other guide genes.

② *Homo sapiens* Ortholog ID: gives orthologous human NCBI gene ID and conventional gene name of guide *D. melanogaster* genes

③ Gene ID: gives the name of query genes. Symbol represents alternative common names. Clicking the names leads to the FlyBase for detailed descriptions

④ Score: Sum of LLS (Log Likelihood Scores) of network links to all other guide genes.

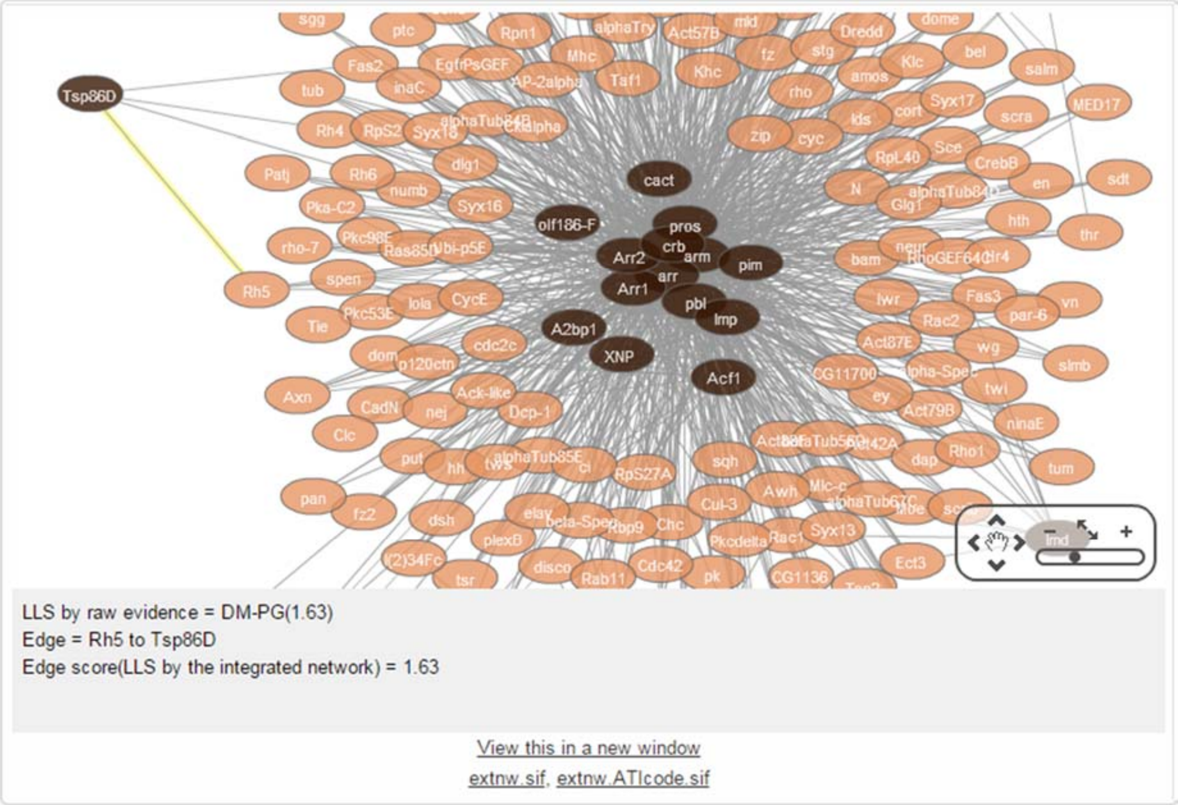
⑤ Evidences & Contribution: shows the fraction of links from component networks (see network download page) supporting network-based prediction.

⑥ #connected guide gene/ # valid guide gene: The number of linked guide genes over the number of total valid guide genes. 'Valid' genes are *D. melanogaster* genes that are known as protein coding genes.

⑦ Connected guide gene: shows a list of linked guide genes.

⑧ GO terms: Corresponding GO-BP/CC/MF annotations that can predict the role of the gene

v. Candidate gene analysis – Guide and candidate genes connection in FlyNet



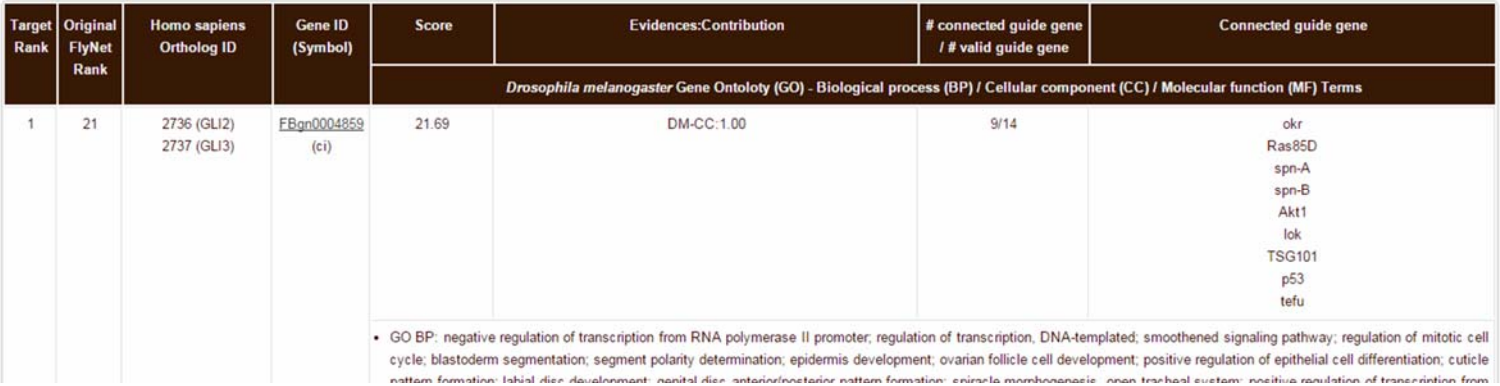
LLS by raw evidence = DM-PG(1.63)
Edge = Rh5 to Tsp86D
Edge score(LLS by the integrated network) = 1.63

View this in a new window

extnw.sif, extnw.ATIcode.sif

Candidate genes (light nodes) that are connected (grey lines) to guide genes (dark nodes) are visualized in cytoscape. You can get information of edges or nodes by clicking them

vi. Candidate gene analysis – List of candidate (target) genes, prioritized by neighbor sum



Target Rank	Original FlyNet Rank	Homo sapiens Ortholog ID	Gene ID (Symbol)	Score	Evidences:Contribution	# connected guide gene / # valid guide gene	Connected guide gene
				Drosophila melanogaster Gene Ontoloty (GO) - Biological process (BP) / Cellular component (CC) / Molecular function (MF) Terms			
1	21	2736 (GLI2) 2737 (GLI3)	FBgn0004859 (ci)	21.69	DM-CC:1.00	9/14	okr Ras85D spn-A spn-B Akt1 lok TSG101 p53 tefu
				• GO BP: negative regulation of transcription from RNA polymerase II promoter; regulation of transcription, DNA-templated; smoothened signaling pathway; regulation of mitotic cell cycle; blastoderm segmentation; segment polarity determination; epidermis development; ovarian follicle cell development; positive regulation of epithelial cell differentiation; cuticle pattern formation; labial disc development; genital disc anterior/posterior pattern formation; spiracle morphogenesis, open tracheal system; positive regulation of transcription from			

① Target Rank: Guilt-by-association principle prioritizes guide and candidate genes by their connectivity to other guide genes. This rank is relevant to confined target genes.

② Original FlyNet Rank: The original rank number without considering target gene.

③ *Homo sapiens* Ortholog ID: gives orthologous human NCBI gene ID and conventional gene name of guide *D. melanogaster* genes

④ Gene ID: gives the name of query genes. Symbol represents alternative common names. Clicking the names leads to the FlyBase for detailed descriptions

⑤ Score: Sum of LLS (Log Likelihood Scores) of network links to all guide genes.

⑥ Evidences & Contribution: shows the fraction of links from component networks (see network download page) supporting network-based prediction.

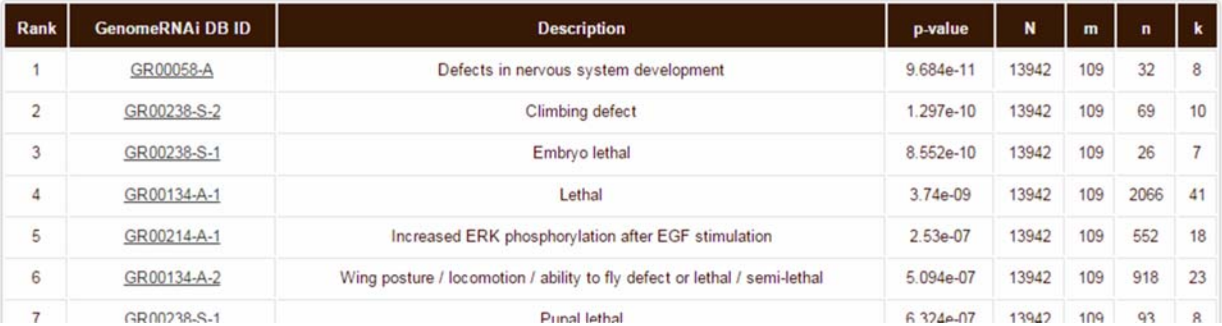
⑦ #connected guide gene/ # valid guide gene: The number of linked guide genes over the number of total valid guide genes. ‘Valid’ genes are *D. melanogaster* genes that are known as protein coding genes.

⑧ Connected guide gene: shows a list of linked guide genes.

⑨ GO terms: Corresponding GO-BP/CC/MF annotations that can predict the role of the gene

vii. Candidate gene analysis – Candidate RNAi phenotype

Finding an RNAi phenotype gene set which has similar gene composition with prioritized genes (guide and target) can suggest a possible experimental start-point to investigate the queried human disease. The significance of enrichment to candidate RNAi phenotypes is calculated by the p-value from hypergeometric test. RNAi phenotypes are reported as prioritized order. Format of the report is same as ‘Gene Set Analysis’.



Rank	GenomeRNAi DB ID	Description	p-value	N	m	n	k
1	GR00058-A	Defects in nervous system development	9.684e-11	13942	109	32	8
2	GR00238-S-2	Climbing defect	1.297e-10	13942	109	69	10
3	GR00238-S-1	Embryo lethal	8.552e-10	13942	109	26	7
4	GR00134-A-1	Lethal	3.74e-09	13942	109	2066	41
5	GR00214-A-1	Increased ERK phosphorylation after EGF stimulation	2.53e-07	13942	109	552	18
6	GR00134-A-2	Wing posture / locomotion / ability to fly defect or lethal / semi-lethal	5.094e-07	13942	109	918	23
7	GR00238-S-1	Pupal lethal	6.324e-07	13942	109	93	8

① Rank: Rank number of corresponding RNAi phenotype term.

② ID: Genome RNAi DB ID

③ Description: Description of the RNAi phenotype

④ p-value: from hypergeometric test

⑤ N: total number of protein coding genes of *D. melanogaster*

⑥ m: number of genes in prioritized gene set (= available guide genes + target genes)

⑦ n: number of genes consisting each RNAi phenotype terms

⑧ k: number of genes that are relevant to both the prioritized gene set and RNAi phenotype term

Climbing defect

Rank	Fly id	Score	Human ID(s)
1	FBgn0000251	20.37	1044,1045
2	FBgn0265434	19.10	22989,4619,4620,4621,4622,4624,4625,4626,4627,4628,4629,57644,79784,8735
3	FBgn0001250	13.56	3674,3678,3685,8516
4	FBgn0264695	13.39	22989,4619,4620,4621,4622,4624,4625,4626,4627,4628,4629,57644,79784,8735
5	FBgn0000042	11.44	100996820,345651,445582,58,59,60,653269,653781,70,71,728378,72
6	FBgn0040505	11.28	238,4058
7	FBgn0003984	9.43	9499,9542
8	FBgn0000045	9.39	100996820,345651,445582,58,59,60,653269,653781,70,71,728378,72

Clicking on the hyperlink on Genome RNAi DB ID allows the list of overlapped *D. melanogaster* genes (‘k’ in the table above) and those human orthologous gene IDs (NCBI gene ID form) in a new window. These genes are also prioritized by the score that is used in the prioritization of guide/target genes (Sum of LLS of network links to all guide genes).

TOP ↑

<END>