台灣生物資訊核心設施
Taiwan Bioinformatics Institute
Core Facility

國家衛生研究院
National Health Research Institutes

The MamPhEA *and* DroPhEA Tutorials

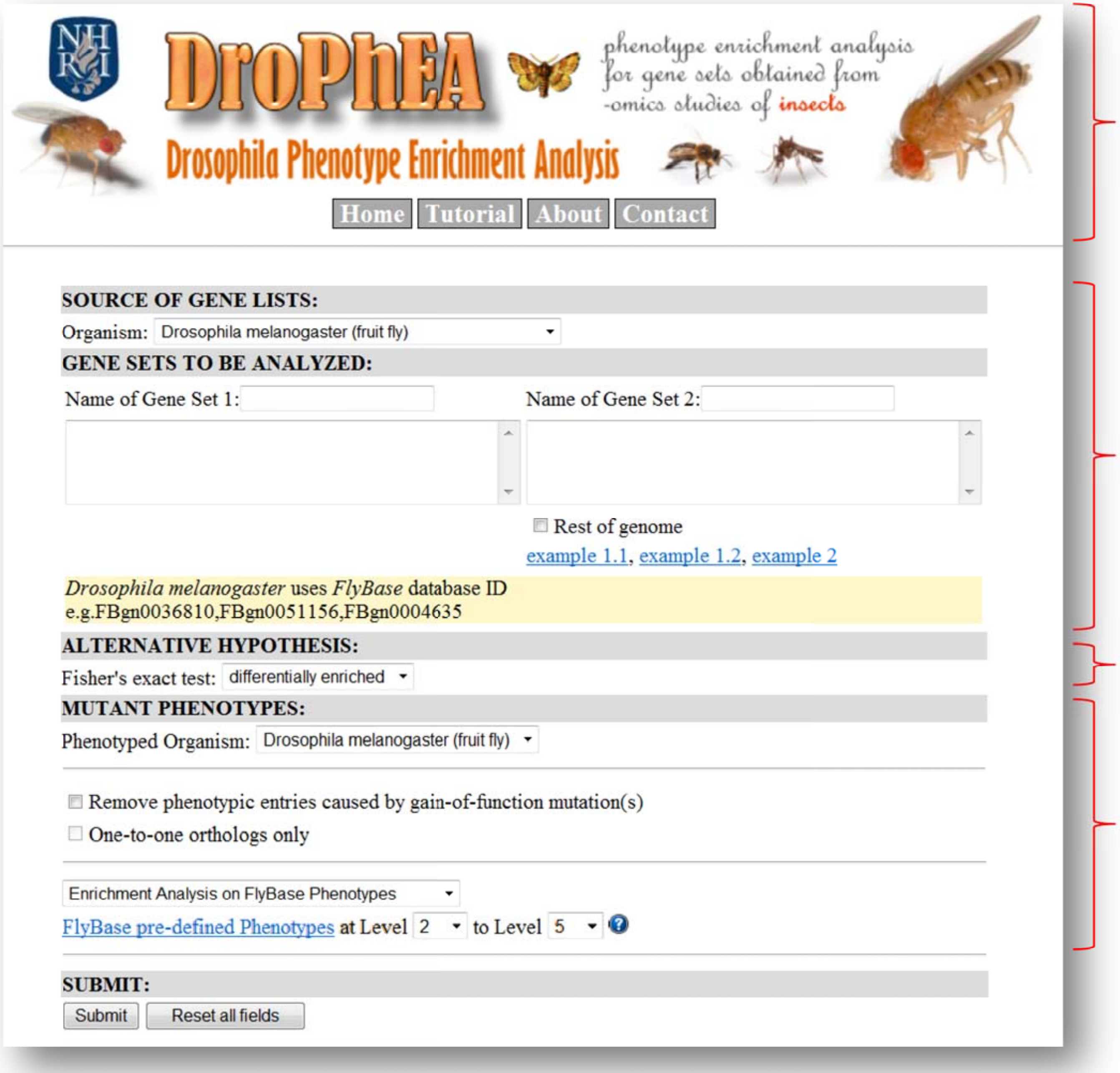
Meng-Pin Weng and Ben-Yang Liao

Division of Biostatistics & Bioinformatics, Institute of
Population Health Sciences, National Health Research
Institutes, Zhunan, Miaoli County 350, Taiwan, R.O.C.

Outline

1. Interface introduction	
2. Data input	Input gene set Select parameter Input phenotype
3. Result	Result display option Hierarchical result Listing result

Interface introduction



SOURCE OF GENE LISTS:

Organism: Drosophila melanogaster (fruit fly)

GENE SETS TO BE ANALYZED:

Name of Gene Set 1:

Name of Gene Set 2:

☐ Rest of genome

example 1.1, example 1.2, example 2

Drosophila melanogaster uses *FlyBase* database ID
e.g.FBgn0036810,FBgn0051156,FBgn0004635

ALTERNATIVE HYPOTHESIS:

Fisher's exact test: differentially enriched

MUTANT PHENOTYPES:

Phenotyped Organism: Drosophila melanogaster (fruit fly)

☐ Remove phenotypic entries caused by gain-of-function mutation(s)

☐ One-to-one orthologs only

Enrichment Analysis on FlyBase Phenotypes

FlyBase pre-defined Phenotypes at Level 2 to Level 5

SUBMIT:

Information

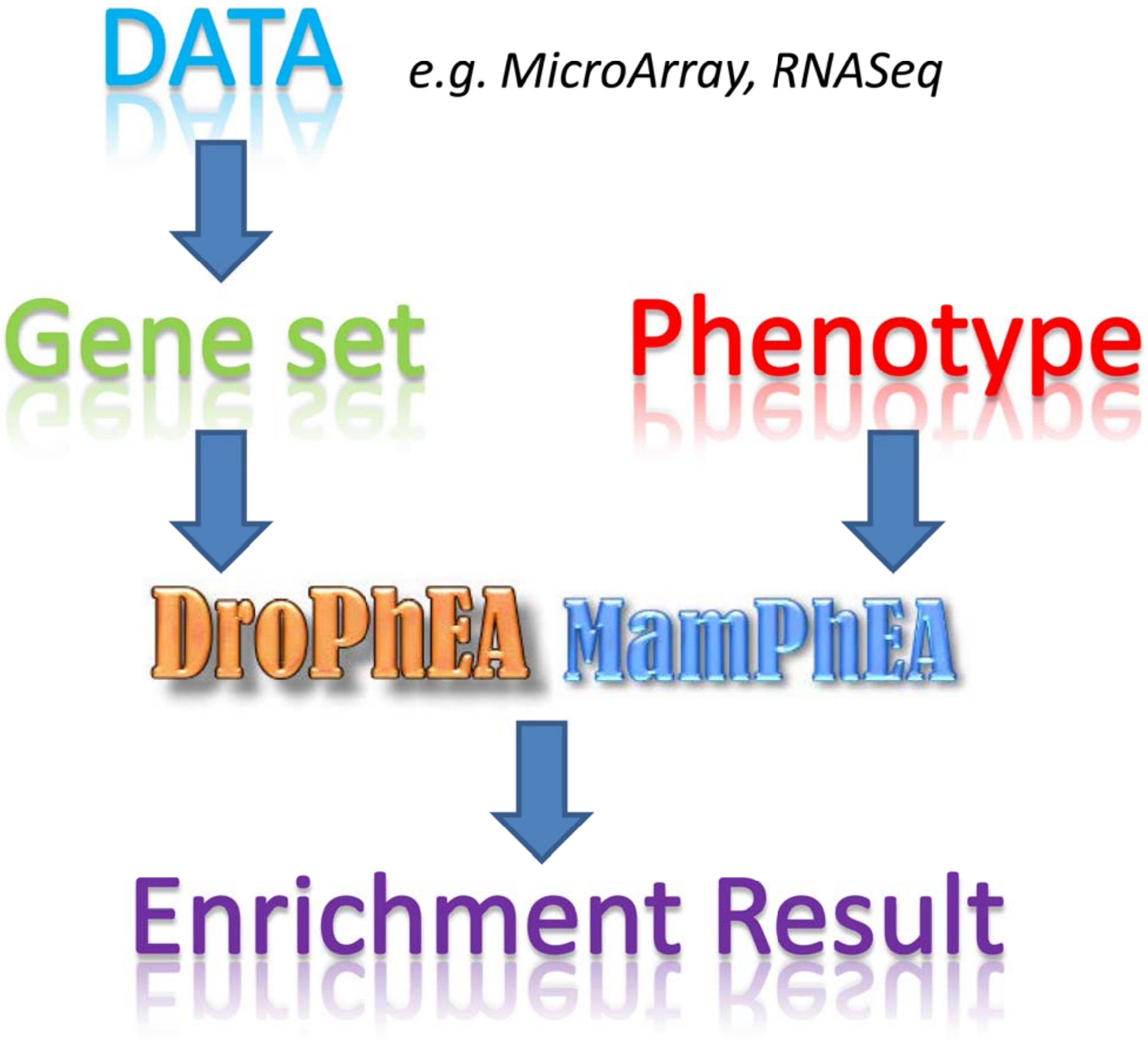
Data input

Test hypothesis

Phenotype

http://evol.nhri.org.tw/phenome/

*Dro*PhEA/*Mam*PhEA working flow chart



DATA

e.g. MicroArray, RNASeq

Gene set

Phenotype

DroPhEA MamPhEA

Enrichment Result

A real case apply to *Dro*PhEA (example 2 of *Dro*PhEA)

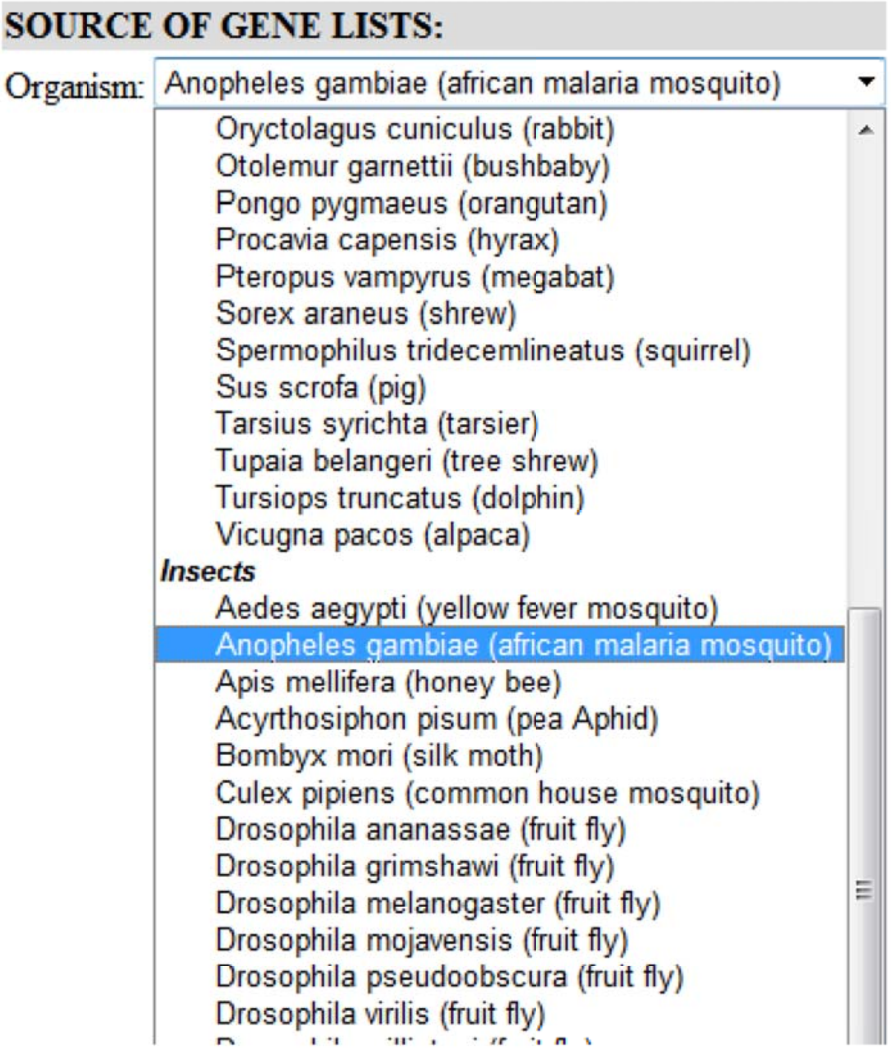
Example: Gene expression difference after a Blood-meal in 24hr

Species: African malaria mosquito
Data set 1: Top 20% highly expressed gene after a Blood-meal
Data set 2: Bottom 80% lowly expressed gene after a Blood-meal

Wu et. al,. (2009) Molecular Biology and Evolution

** You can find this case in DroPhEA example 2*

Data input: select species of input gene



SOURCE OF GENE LISTS:

Organism: Anopheles gambiae (african malaria mosquito)

- Oryctolagus cuniculus (rabbit)
- Otolemur garnettii (bushbaby)
- Pongo pygmaeus (orangutan)
- Procavia capensis (hyrax)
- Pteropus vampyrus (megabat)
- Sorex araneus (shrew)
- Spermophilus tridecemlineatus (squirrel)
- Sus scrofa (pig)
- Tarsius syrichta (tarsier)
- Tupaia belangeri (tree shrew)
- Tursiops truncatus (dolphin)
- Vicugna pacos (alpaca)

Insects

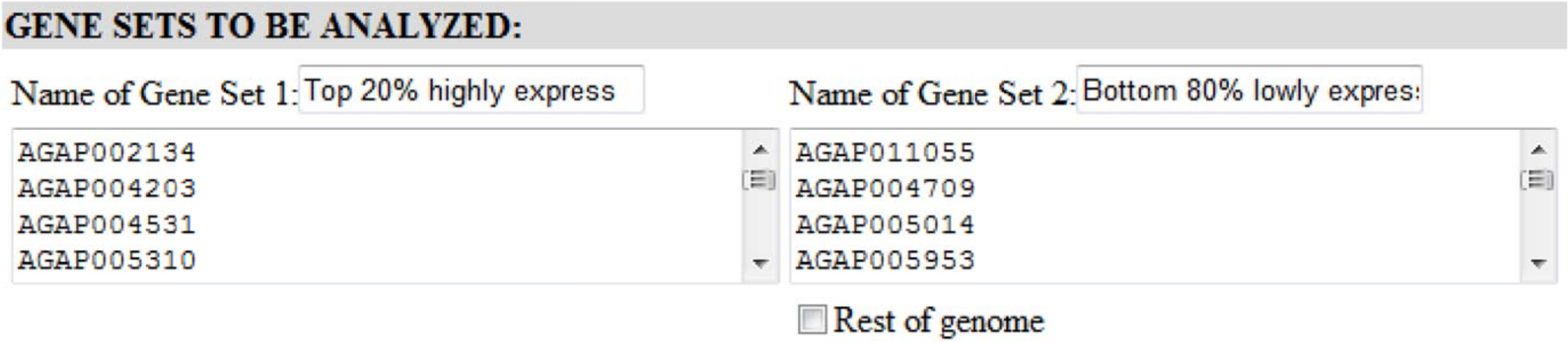
- Aedes aegypti (yellow fever mosquito)
- Anopheles gambiae (african malaria mosquito)
- Apis mellifera (honey bee)
- Acyrthosiphon pisum (pea Aphid)
- Bombyx mori (silk moth)
- Culex pipiens (common house mosquito)
- Drosophila ananassae (fruit fly)
- Drosophila grimshawi (fruit fly)
- Drosophila melanogaster (fruit fly)
- Drosophila mojavensis (fruit fly)
- Drosophila pseudoobscura (fruit fly)
- Drosophila virilis (fruit fly)

Specify the organism of gene set(s) to be analyzed

In total,
35 mammalian species
17 insect species

http://evol.nhri.org.tw/phenome/

Data input: input gene set and assign data name



GENE SETS TO BE ANALYZED:

Name of Gene Set 1: Top 20% highly express

```
AGAP002134
AGAP004203
AGAP004531
AGAP005310
```

Name of Gene Set 2: Bottom 80% lowly expres:

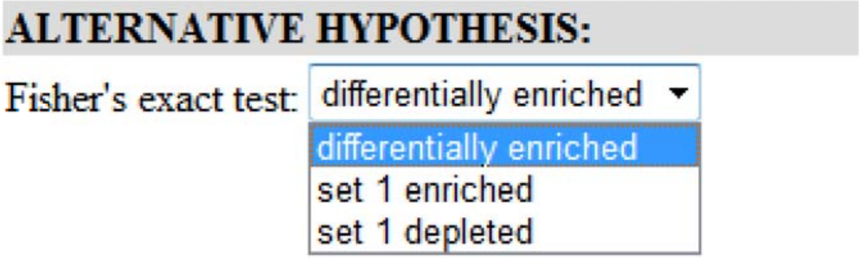
```
AGAP011055
AGAP004709
AGAP005014
AGAP005953
```

☐ Rest of genome

Use (a) a tab, (b) a return, (c) a comma, (d) a semicolon, or (e) a single space to separate genes.

**Gene ID can be any, please see our online document to find our supported gene ID*

Test hypothesis:

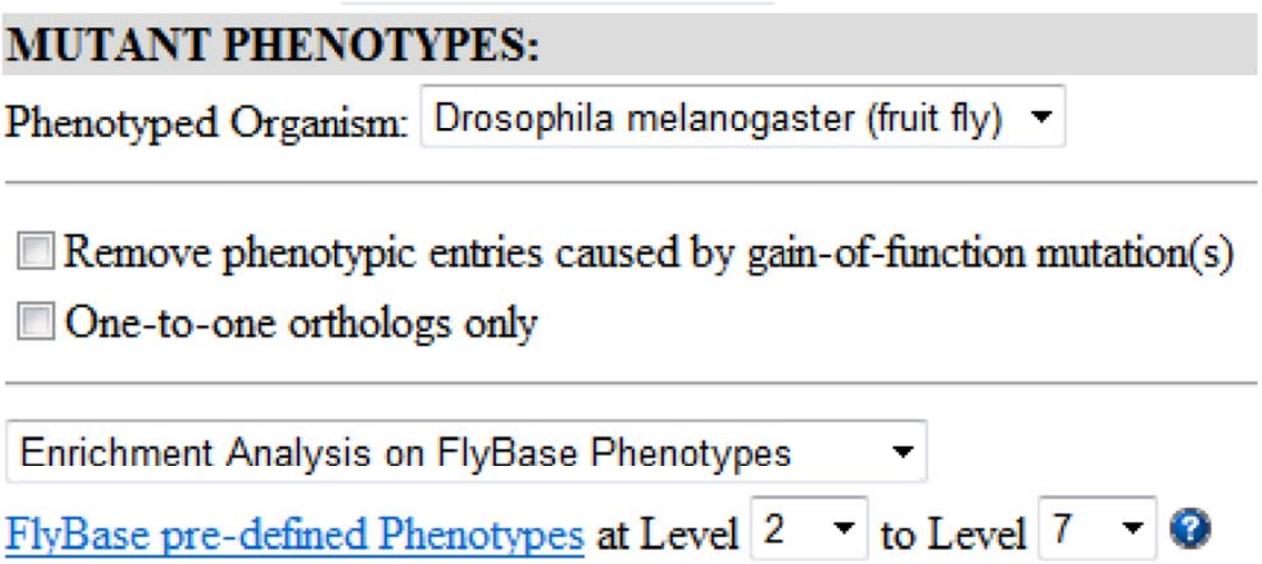


ALTERNATIVE HYPOTHESIS:

Fisher's exact test: differentially enriched

Differentially enriched – genes differently distribute in phenotype
Set1 enriched – gene set1 significantly enrich in phenotype
Set1 depleted – gene set1 significantly depleted in phenotype

Input phenotype: (1) hierarchical structure



MUTANT PHENOTYPES:

Phenotyped Organism: Drosophila melanogaster (fruit fly)

☐ Remove phenotypic entries caused by gain-of-function mutation(s)

☐ One-to-one orthologs only

Enrichment Analysis on FlyBase Phenotypes

FlyBase pre-defined Phenotypes at Level 2 to Level 7

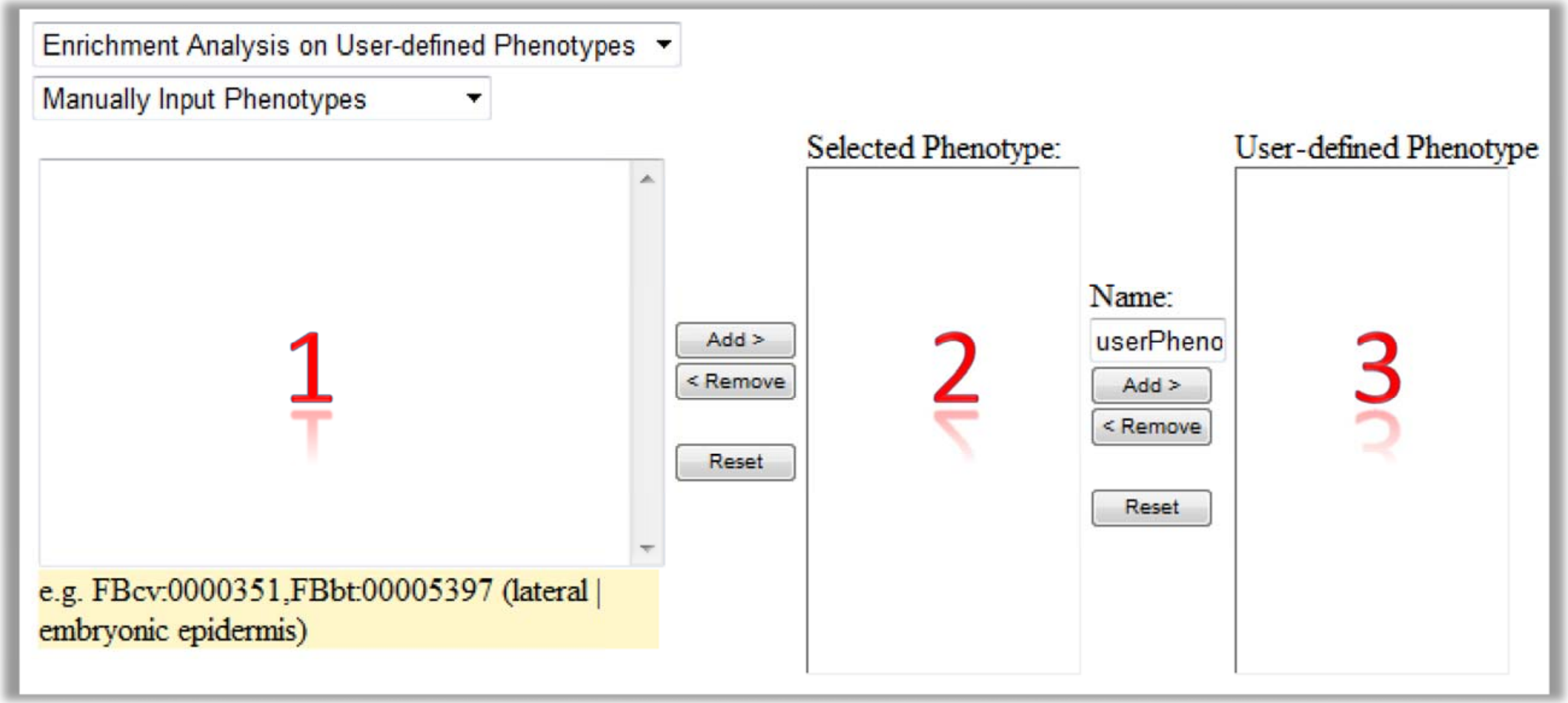
Switch *Dro*PhEA / *Mam*PhEA

Mapping method

Select analysis level

**Dro*PhEA, can specifically exclude gain of function mutation
*Mam*PhEA, can focus on loss-of-function phenotypes only

Input phenotype: (2) User-defined phenotypes



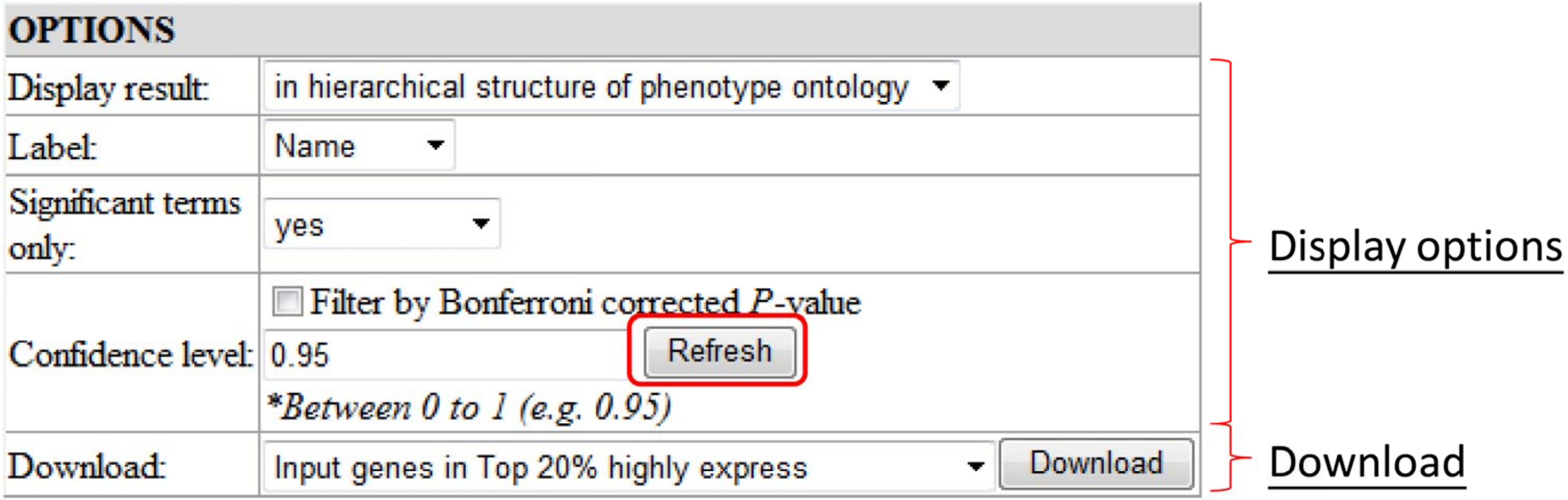
userPheno

e.g. FBcv:0000351,FBbt:00005397 (lateral | embryonic epidermis)

3 steps to customize a phenotype:

(1) search phenotype you want to combine
(2) add to selected phenotype in the middle
(3) give a name and make a customized phenotype

Result: filtering result



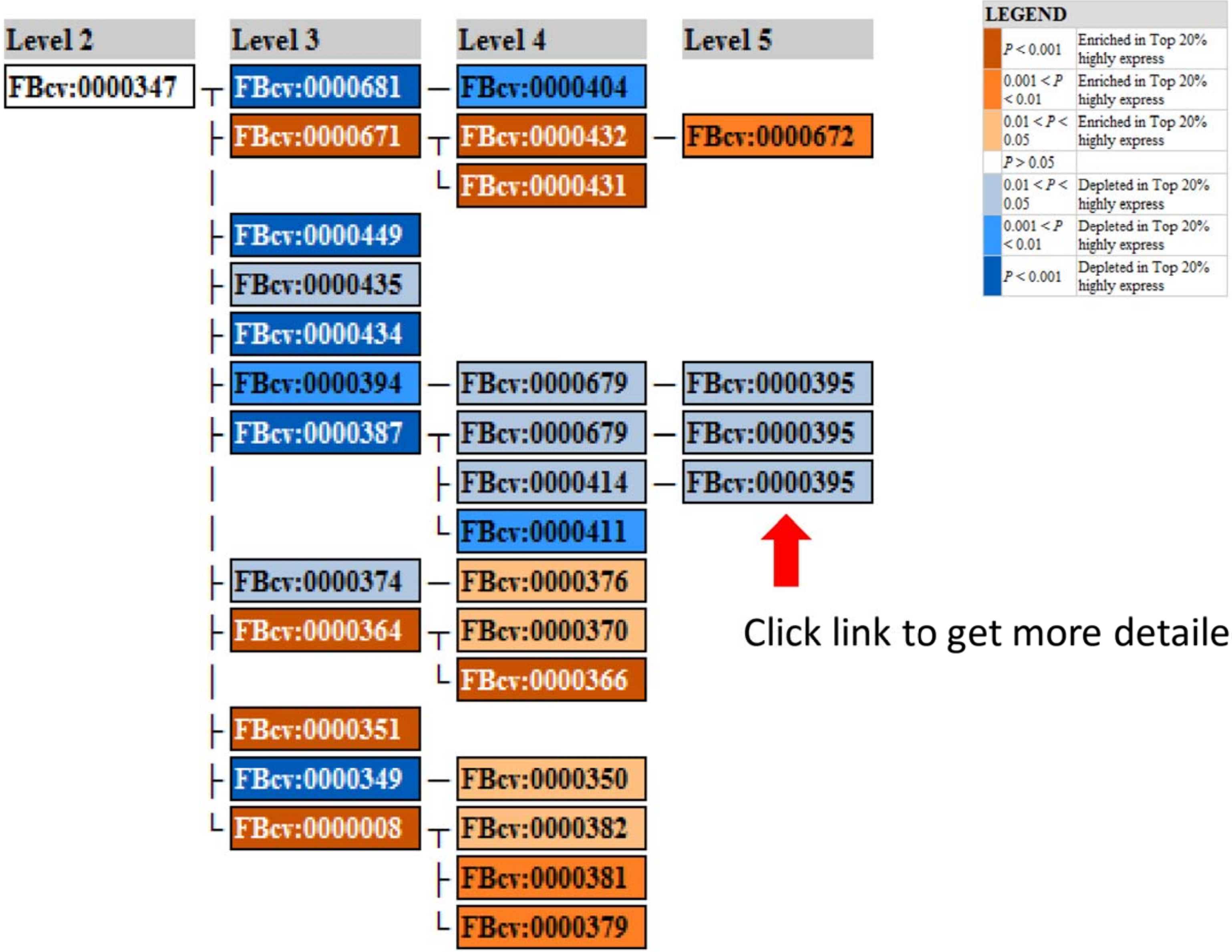
OPTIONS	
Display result:	in hierarchical structure of phenotype ontology
Label:	Name
Significant terms only:	yes
Confidence level:	☐ Filter by Bonferroni corrected *P*-value 0.95 Refresh _*Between 0 to 1 (e.g. 0.95)_
Download:	Input genes in Top 20% highly express Download

2 steps :

(1) Modify your filtering parameter or display style

(2) Hit **REFRESH** !

Result: (1) Hierarchical structure result

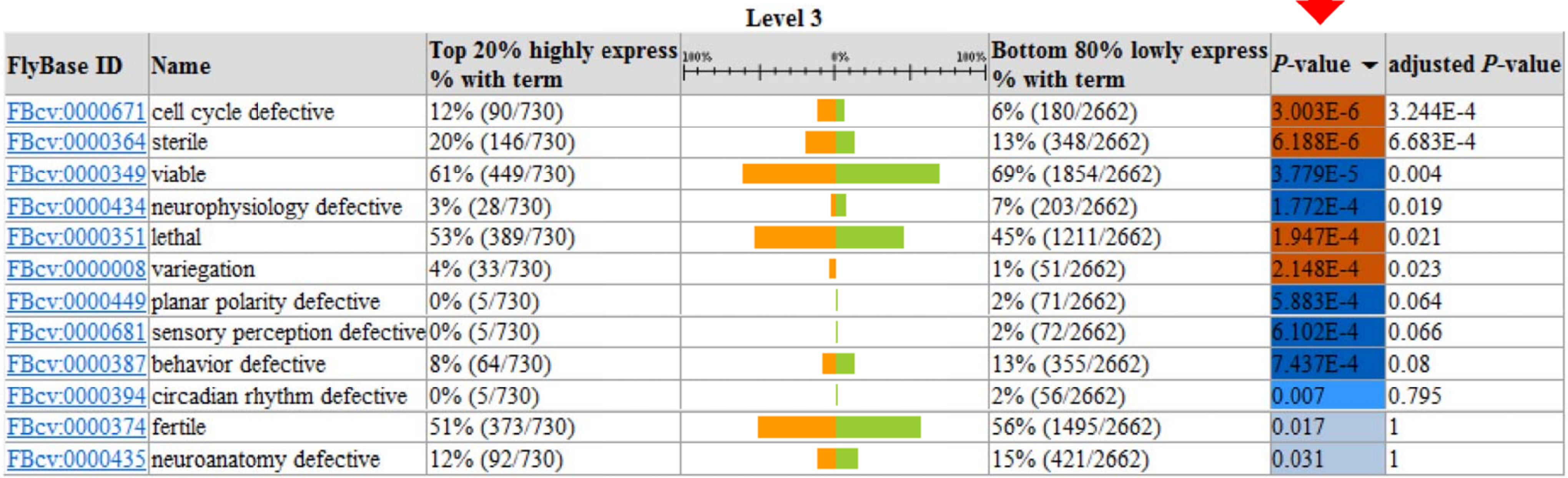


LEGEND		
🟧	$P < 0.001$	Enriched in Top 20% highly express
🟧	$0.001 < P < 0.01$	Enriched in Top 20% highly express
🟧	$0.01 < P < 0.05$	Enriched in Top 20% highly express
⬜	$P > 0.05$	
🟦	$0.01 < P < 0.05$	Depleted in Top 20% highly express
🟦	$0.001 < P < 0.01$	Depleted in Top 20% highly express
🟦	$P < 0.001$	Depleted in Top 20% highly express

Click link to get more detailed result

Result: (2) List of phenotype enrichment

Hit the header to sort

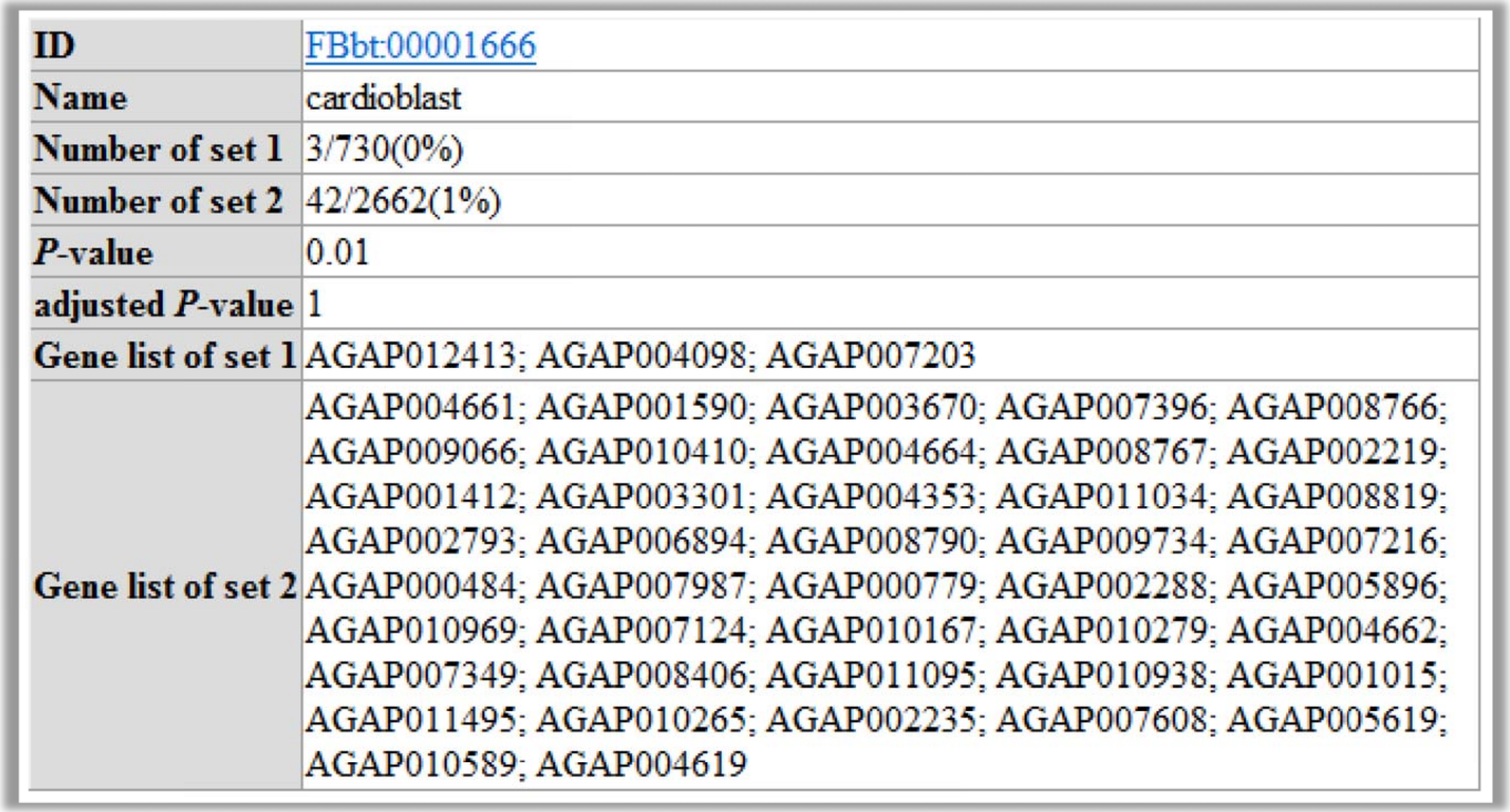


Level 3

FlyBase ID	Name	Top 20% highly express % with term	100% 0% 100%	Bottom 80% lowly express % with term	*P*-value ▾	adjusted *P*-value
FBcv:0000671	cell cycle defective	12% (90/730)		6% (180/2662)	3.003E-6	3.244E-4
FBcv:0000364	sterile	20% (146/730)		13% (348/2662)	6.188E-6	6.683E-4
FBcv:0000349	viable	61% (449/730)		69% (1854/2662)	3.779E-5	0.004
FBcv:0000434	neurophysiology defective	3% (28/730)		7% (203/2662)	1.772E-4	0.019
FBcv:0000351	lethal	53% (389/730)		45% (1211/2662)	1.947E-4	0.021
FBcv:0000008	variegation	4% (33/730)		1% (51/2662)	2.148E-4	0.023
FBcv:0000449	planar polarity defective	0% (5/730)		2% (71/2662)	5.883E-4	0.064
FBcv:0000681	sensory perception defective	0% (5/730)		2% (72/2662)	6.102E-4	0.066
FBcv:0000387	behavior defective	8% (64/730)		13% (355/2662)	7.437E-4	0.08
FBcv:0000394	circadian rhythm defective	0% (5/730)		2% (56/2662)	0.007	0.795
FBcv:0000374	fertile	51% (373/730)		56% (1495/2662)	0.017	1
FBcv:0000435	neuroanatomy defective	12% (92/730)		15% (421/2662)	0.031	1

Click link to get more detailed result

Result: (3) Click the link and get more detailed result



ID	FBbt:00001666
Name	cardioblast
Number of set 1	3/730(0%)
Number of set 2	42/2662(1%)
***P*-value**	0.01
adjusted *P*-value	1
Gene list of set 1	AGAP012413; AGAP004098; AGAP007203
Gene list of set 2	AGAP004661; AGAP001590; AGAP003670; AGAP007396; AGAP008766; AGAP009066; AGAP010410; AGAP004664; AGAP008767; AGAP002219; AGAP001412; AGAP003301; AGAP004353; AGAP011034; AGAP008819; AGAP002793; AGAP006894; AGAP008790; AGAP009734; AGAP007216; AGAP000484; AGAP007987; AGAP000779; AGAP002288; AGAP005896; AGAP010969; AGAP007124; AGAP010167; AGAP010279; AGAP004662; AGAP007349; AGAP008406; AGAP011095; AGAP010938; AGAP001015; AGAP011495; AGAP010265; AGAP002235; AGAP007608; AGAP005619; AGAP010589; AGAP004619

Result: (4) Blood-meal related phenotypes

phenotype_name	phenotype_id	Top 20% highly express	background	p_value	adj_p_value	
sterile	FBcv0000364	145	371	6.88E-06	3.44E-04	+
female sterile	FBcv0000366	110	269	2.75E-05	0.002	+
male germline cell	FBbt00057011	45	83	7.94E-05	0.114	+
male fusome	FBbt00007139	5	1	3.56E-04	0.28	+
spermatocyte fusome	FBbt00004937	5	1	3.56E-04	0.51	+
polar body nucleus	FBbt00000051	7	2	3.59E-04	0.282	+
pronucleus	FBbt00000048	7	2	3.59E-04	0.101	+
germline cell	FBbt00004860	101	266	6.19E-04	0.486	+
spermatocyte ring canal	FBbt00004939	5	1	0.002	1	+
fertile	FBcv0000374	371	1596	0.008	0.425	-
fusome	FBbt00004862	9	10	0.008	1	+
male fertile	FBcv0000376	29	61	0.008	0.682	+
egg	FBbt00000034	47	116	0.009	1	+
cell cycle defective	FBcv0000671	89	196	7.85E-06	3.92E-04	+
mitotic cell cycle defective	FBcv0000432	71	148	1.92E-05	0.002	+
meiotic cell cycle defective	FBcv0000431	32	57	7.27E-04	0.064	+
primordium	FBbt00005495	23	184	3.55E-04	0.1	-
planar polarity defective	FBcv0000449	5	72	9.21E-04	0.046	-
tergum	FBbt00004552	3	58	0.001	1	-
germ layer	FBbt00000110	3	50	0.005	1	-
viable	FBcv0000349	448	1983	1.19E-05	5.95E-04	-
lethal	FBcv0000351	387	1268	6.39E-05	0.003	+
embryonic neuron	FBbt00001446	8	114	2.93E-05	0.023	-
segmental nerve	FBbt00002037	2	55	3.90E-04	0.559	-
motor neuron	FBbt00005123	7	89	4.82E-04	0.378	-
neurophysiology defective	FBcv0000434	28	201	9.29E-04	0.046	-
neuron	FBbt00005106	75	417	0.002	0.486	-
sensory neuron	FBbt00005124	55	319	0.003	1	-
abdominal nerve	FBbt00002179	3	48	0.008	1	-
nerve	FBbt00005105	22	151	0.009	1	-
embryonic myoblast	FBbt00000463	1	33	7.82E-04	1	-
myoblast	FBbt00005083	1	42	9.54E-04	0.749	-
embryonic/larval somatic muscle	FBbt00000464	16	125	0.005	1	-
muscle cell	FBbt00005074	29	189	0.006	1	-
somatic muscle	FBbt00005073	28	179	0.01	1	-
antenna	FBbt00004511	6	74	0.002	1	-
inter-ommatidial cell	FBbt00007250	1	39	0.002	1	-

🟦 reproduction
🟪 cell cycle
🟥 development
🟩 mortality
🟧 neural
🟦 muscle
🟧 sensory
🟪 behavior
🟨 cell-line related
🟩 variegation related
⬜ other category

Thanks for your attention!

Question is welcome!