

A genome-wide gene function prediction resource for *Drosophila melanogaster*.

Yan H, Venkatesan K, Beaver JE, Klitgord N, Yildirim MA, Hao T, et al.

Supporting Information

Table S1

Classification of KEGG pathways.
doi:10.1371/journal.pone.0012139.s001
(0.04 MB XLS)

Table S2

GO term and KEGG pathway membership predictions with confidence scores above 0.2 (for GO term predictions) or above 0.1 (for KEGG pathway predictions).
doi:10.1371/journal.pone.0012139.s002
(3.91 MB XLS)

Table S3

KEGG pathway predictions matched to DRSC screening.
doi:10.1371/journal.pone.0012139.s003
(0.03 MB XLS)

Table S4

GO-term predictions matched to DRSC screening.
doi:10.1371/journal.pone.0012139.s004
(0.04 MB XLS)

Table S5

JNK pathway prediction compared to RNAi data at the Japan National Institute of Genetics (NIG).
doi:10.1371/journal.pone.0012139.s005
(0.03 MB DOC)

Table S6

Keyword matching between RNAi screens and gene functions.
doi:10.1371/journal.pone.0012139.s006
(0.02 MB XLS)

Table S7

Additional Expression Profiles from NCBI Gene Expression Omnibus (GEO).
doi:10.1371/journal.pone.0012139.s007
(0.10 MB XLS)

Figure S1

Genetic interaction profiles. A, Both pathway P1 and P2 drive downstream processes to achieve a function; loss of genes within either pathway will not abolish the function. However, when a pair of genes from the two pathways respectively are lost (e.g. a-x, a-y, c-x), both pathways will be broken and a loss-of-function phenotype will emerge. B, Genes within the same pathway have similar genetic interaction profiles, which could be useful in categorizing a,b,c and x,y,z into pathway P1 and P2 respectively.
doi:10.1371/journal.pone.0012139.s008
(1.07 MB TIF)

Figure S2

Code performance of Functional-Specific Classifier model and canonical supervised machine-learning model.
doi:10.1371/journal.pone.0012139.s009
(0.51 MB TIF)

Figure S3

Performance of GO prediction when removing one feature at a time from the model. Receiver Operating Characteristic curves (A) and Precision-Recall curves (B) for the overall performance and the performance when removing one feature at a time in GO term (biological process, BP) prediction. Precision-Recall curves for the GO term prediction model for GO terms with various degrees of specificity, i.e., those that have been annotated with 0–25 genes (C), 25–50 genes (D), 50–100 genes (E), and 100–500 genes (F).

doi:10.1371/journal.pone.0012139.s010

(1.73 MB TIF)

Figure S4

Performance of KEGG prediction when removing one feature at a time from the model. Receiver Operating Characteristic curves (A) and Precision-Recall curves (B) for the overall performance and the performance when removing one feature at a time in the KEGG pathway prediction. Precision-Recall curves for the performance of the model in predicting metabolism only (C), signaling pathway only (D), basic functions (E), and all non-metabolism functions (F).

doi:10.1371/journal.pone.0012139.s011

(2.20 MB TIF)

Figure S5

Performance of KEGG prediction when limited in the gene space of genetic interaction network. Receiver Operating Characteristic curves (A) and Precision-Recall curves (B) for the overall performance and contribution of each feature in the KEGG pathway prediction. Precision-Recall curves for the performance of the model in predicting metabolism only (C), signaling pathway only (D), basic functions (E), and all non-metabolism functions (F).

doi:10.1371/journal.pone.0012139.s012

(1.85 MB TIF)