Drosophila melanogaster life-cycle gene expression dataset*

Zareen Gauhar, Murad Ghanim, Tine Herreman, J. David Lambert, Tong-Ruei Li, Chris Mason, Scott Rifkin, Ling Sun, and Kevin P. White#

***Contributors listed alphabetically**
#To whom correspondence may be addressed at kpwhite@uchiago.edu

The *D. melanogaster* life-cycle timecourse data were acquired in the White lab from 2001-2003. The probes printed on the microarray are amplicon products from primer pairs designed on the version 2.0 of the D. melanogaster assembly and annotation, but all amplicons have been mapped forward to assembly and annotation version 5.10. There are 30 time-points over embryo development, 10 time-points over larval development, 18 time-points over metamorphosis, and 4 time-points each for male and female adult flies. These arrays are two channels and all sampled developmental time-points were labeled with Cy3 fluorescent dye and the reference set, consisting of pooled RNA samples across all time-points, was labeled with a Cy5 fluorescent dye. Hybridizations were scanned with an Axon GenePix scanner.

The platform used by Arbeitman, et al.[1] was spotted cDNAs and covered roughly a third of the *D. melanogaster* genes, while the platform used here are spotted amplicons which covers over 85% of the *D.melanogaster* genes. Samples were collected identically to the method described in Arbeitman, et al.[1].

Microarray normalization protocols of the data shown in Flybase were performed by James C Costello[1,2] and Justen R Andrews[1,2,3]

[1] Department of Biology, [2] School of Informatics, [3] Center for Genomics and Bioinformatics
Indiana University, Bloomington, IN

Mapping PCR Products to FlyBase Annotated Genes

Drosophila melanogaster sequences were downloaded from FlyBase[1]. Forward and reverse primer sequences were searched against the version 5 genome assembly of the *D. melanogaster* genome using BLASTn (E-value < 10^{-3}). BLAST results were processed with custom Perl scripts. The physical coordinates of the primer pairs were matched and the predicted amplicon product between the primers (including the primer pairs) were considered "good" if the predicted product was less than 1000 bp and the primer pairs unambiguously produced one predicted amplicon product. The predicted amplicons were then searched against the version 5 genome assembly of the *D. melanogaster* genome using BLASTn (E-value < 10^{-6}). Amplicon BLAST results were processed to ensure the predicted amplicon would hybridize to one place on the genome. The physical coordinates of transcripts associated with a FlyBase annotated gene were from version 5.10 of the *D. melanogaster* genome annotation. A predicted amplicon was

1 ftp://www.flybase.net/genomes/Drosophila_melanogaster/

mapped to a FlyBase gene ID if the BLAST results mapped to transcripts from a unique gene and only one gene.

Microarray Normalization Protocol

Raw data for all timepoints, in the form of GPR (GenePix Results) files, were read into the **R** statistical programming environment[2] through the use of the *marray* package. All control spots and "blank" spots were flagged in addition to any spots flagged by the scanning software. The arrays were then normalized through the *StepNorm* package [2], which evaluates each individual array to determine the best normalization model with respect to spot intensity, print-tip group, print plate, and spatial effects. This package produces log-transformed ratio and intensity values for each spot. Non-flagged replicate spots across biological replicates were averaged.

References

1. Arbeitman, M.N., et al., *Gene expression during the life cycle of Drosophila melanogaster.* Science, 2002. **297**(5590):2270-5.
2. Xiao, Y., Hunt, C.A., Segal, M.R., and Yang, T.H. *Novel stepwise normalization method for two-channel cDNA microarrays*. Conf. Proc. IEEE Eng. Med. Biol. Soc., 2004. **4**:2921-2924.

2 http://www.r-project.org/