

Short Research Paper

Exploring Effects of Sex and Diet on *Drosophila melanogaster* Head Gene Expression

Maria D. L. A. Jaime¹✉, Juan Hurtado^{2,3}, Mariana Ramirez Loustalot-Laclette², Brian Oliver¹ and Therese Markow^{3,4}✉

1. Section of Developmental Genomics, Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA;
2. Department of Ecology, Genetics and Evolution, IEGEBA (CONICET-UBA), Faculty of Exact and Natural Sciences, University of Buenos Aires, Ciudad Universitaria, C1428EHA, Buenos Aires, Argentina;
3. National Laboratory for the Genomics of Biodiversity, CINVESTAV, Irapuato, Guanajuato, Mexico;
4. Division of Biological Sciences, University of California San Diego, La Jolla CA 92093, USA.

✉ Corresponding authors: Maria Jaime: maria.jaime@nih.gov Therese Markow: tmarkow@ucsd.edu

© Ivyspring International Publisher. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY-NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2017.08.15; Accepted: 2017.10.04; Published: 2017.10.29

Abstract

Gene expression depends on sex and environment. We stringently explored the contributions of these effects in *Drosophila melanogaster* by rearing three distinct wildtype genotypes on isocaloric diets either high in protein or sugar followed by expression profiling of heads from the sexes. By using different genotypes as replicates we developed robust sex- and diet-biased expression responses.

Key words: High protein diet, High sugar diet, Adult head, *Drosophila melanogaster*.

Introduction

The *Drosophila melanogaster* head is an important integrator of environmental and genetic information. The central nervous system senses the environment through sensory organs concentrated in the head, and produces important hormones such as insulin. Additionally, the head fat body is an energy storage organ that also provides important physiological signals. At least some gene expression in the head shows sexual dimorphism. *Drosophila melanogaster* has become a powerful tool in exploring the influence of diet on human health because they share most basic metabolic steps with humans [1].

The effects of sex, diet, and genotype on *D. melanogaster* gene expression have been described previously [2-7] and it is clear that substantial interactions exist among these three factors. Our interest is sex-biased and diet-biased expression. To inform future investigations aimed at looking in detail at these effects and their interactions, we performed a preliminary experiment rearing flies on two different diets. To determine which differences were robust to

differences in genetic background, we used three strains of *D. melanogaster*. We focused upon head tissue to capture information on changes in gene expression that might ultimately result in differences in behavior and/or physiology. We were specifically interested in sex-biased and diet-biased expression that was robust to differences in genotype or the other tested factors, as we have six replicates each for sex and diet by binning data. The three strains were from iso-female lines collected in San Diego CA [8]. The isocaloric diets were either high protein (HPS, protein/carbohydrate calories = 0.43) or low protein (LPS, protein/carbohydrate calories = 0.1), as previously reported [8]. After rearing flies on these two diets, we collected heads from newly eclosed (2-6 hours) adults of each sex for expression profiling. We processed a total of 12 samples for stranded PolyA+ RNA-sequencing on the Illumina system. We aligned reads to the *D. melanogaster* genome release 6 [9] using TopHat v2.1.0 [10] and generated gene-level expression values with HTseq v0.6.1p1 [11]. While not

sufficiently replicated for drawing strong conclusions on the individual samples and the interactions among the three factors, the range of samples provides a preliminary view of their relative importance. All experimental details and data are available at the Gene Expression Omnibus under accession GSE82245.

We first compared the gene expression profiles by Euclidean distance clustering (Figure 1a). The expression profile similarities separated first, and most dramatically by strain, followed by sex and then diet. This is in agreement with previous work showing that diet contributes less to expression variance than genotype [12]. However, given the high impact of sex on gene expression in many experiments [7], it was somewhat surprising that genotype contributed more than sex. Principal

component analysis resulted in a similar pattern of separation by genotype, followed by sex (not shown). We also generated volcano plots to show the significant differences in gene expression due to sex and diet (Figure 1b, c). For this analysis, we used the sexed flies and the two different diets to compare six replicates of each. Given the genotype differences, this is an especially stringent selection of genes that show a similar expression behavior across strains. We observed 409 genes with sex-biased expression and 27 genes with diet-biased expression (FDR $p < 0.05$) in heads (Table S1). Gene Ontology (GO) term analysis was not informative, other than enrichment for reproductive function for the genes with sex-biased expression (not shown).

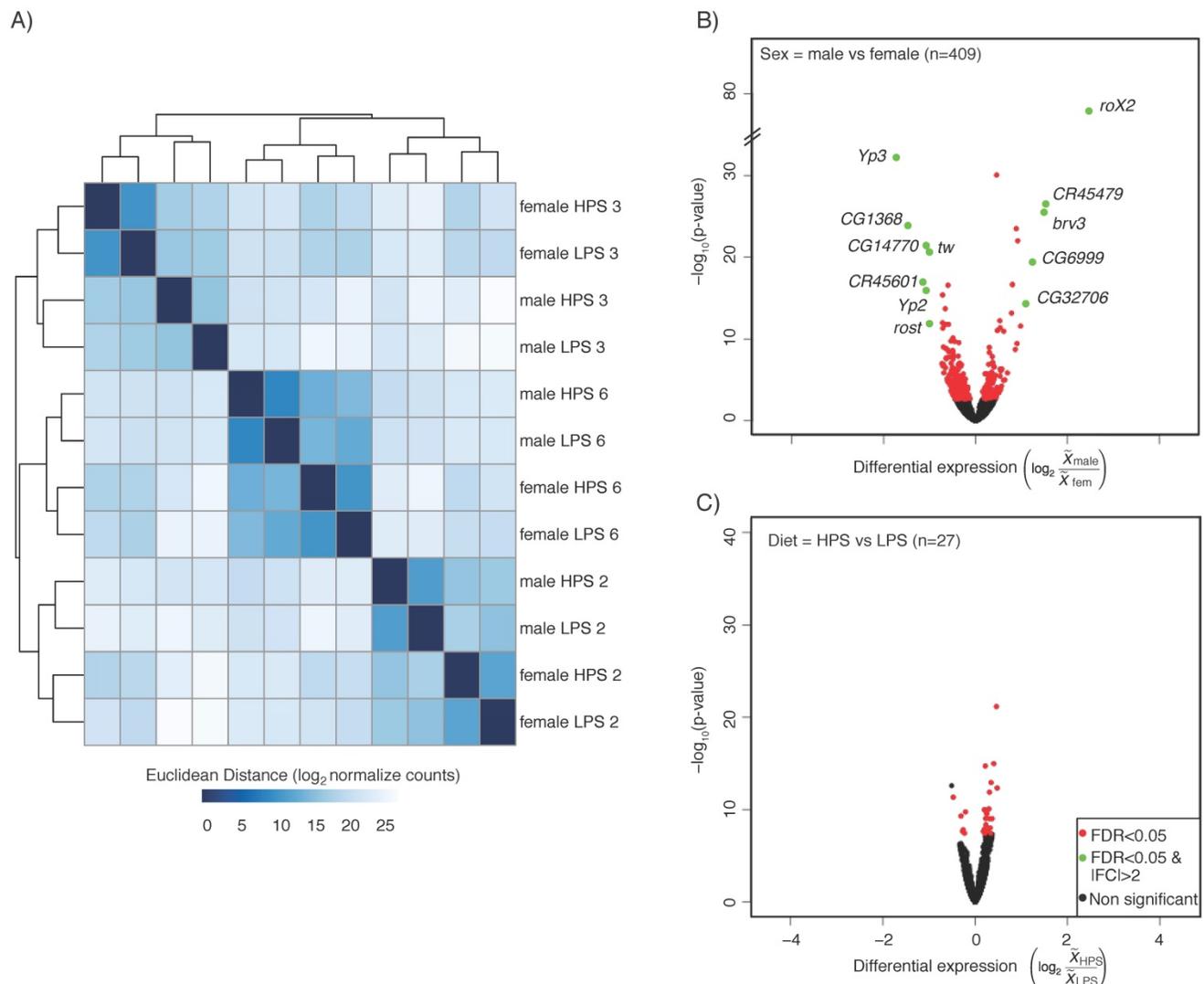


Figure 1. Head expression profiles from different genotypes, sexes, and diets. A) Heat map and dendrogram of sample similarity. Euclidean distance is the square root of the sum of the squared differences in expression (in \log_2 normalized read counts) of each gene between sample pairs. B) Volcano plot of genes with sex biased expression ($\log_2 (\bar{X}_{\text{male}} / \bar{X}_{\text{female}})$). C) Volcano plots of diet-biased expression ($\log_2 (\bar{X}_{\text{HPS}} / \bar{X}_{\text{LPS}})$). B & C) The key (lower right) indicates genes with significantly different expression at a corrected $p < 0.05$ (FDR, red), significantly different with a magnitude (absolute value) > 2 -fold (green), or neither (black). The number of genes differentially expressed in each comparison (n) is shown (upper left) in each panel.

We also browsed the sex-biased and diet-biased gene lists, which confirmed much of what we have come to expect from sex-biased expression patterns in heads, and suggested some interesting new biology. For example, genes with sex-biased expression included many of the usual suspects, such as the *RNA on X* genes (*RoX1* and *RoX2*) that are required for male X chromosome dosage compensation [13] and show male-biased expression, and the *Yolk protein* genes (*Yp1*, *Yp2*, and *Yp3*) that are required for energy storage in developing eggs [14] and show female-biased expression. There were many relatively unstudied genes with sex-biased expression, including non-coding RNAs that would have been missed from non-stranded RNA-seq datasets. For example, *CR45479* showed male-biased expression and is antisense to a cytochrome P450 encoding gene, *Cyp4d14* [15], which also showed male-biased expression. Similarly, *CR45601*, which showed female-biased expression is antisense to *Yp1*. Diet-biased expression was less dramatic, but did include transcription factors such as *kahuli* and predicted sugar transporters (*CG8654*). Some of the diet-biased expression patterns suggested that there may be an effect of diet on neural function. For example, the *CG14446* (*detonator*) gene is required for pain perception [16] and shows LPS-biased expression. Given the importance of the structures and processes in the *Drosophila* head, an influence of pre-adult diet on adult behavior would not be surprising. Collectively, these data provide a resource of sex-biased and diet-biased expression that is robust to dramatic differences in expression profiles due to genotype in the three strains. They further suggest that a more expansive dataset with greater biological replication will uncover diet-biased expression and interactions between diet and sex and genotype.

Methods

A table describing used reagents and resources is provided in Supplementary Table S2. Information included: chemical and biological reagents, commercial kits, fly genotypes, GEO submission of data, and software. The identifier of the resource and provider.

Drosophila Lines and Culture Conditions

We utilize three *D. melanogaster* isofemales lines collected near University of California San Diego in September of 2010, referred as SD2, SD3 and SD6 [17]. Parental lines were reared on standard banana food at 25 °C using a 12 hr light/dark cycle. Mated parents oviposited for 24 hr on 0.5% agar and a sprinkle of yeast.

Larvae were grown on two different diets [8]. Briefly, the diets were composed of sucrose, active dry yeast, yellow cornmeal and agar. Ingredients were mixed and boiled and an antifungal composed of methyl paraben dissolved in ethanol was added once the food was cooled to 55 °C. Immediately after preparation, the food was pipetted in 10-mL aliquots to 8-dram vials. First instar larvae (n=40) were collected and allowed to develop at 25 °C and using a 12 hr light/dark cycle on two different diet.

RNA-seq

We collected 12 samples (2 diets x 2 sexes x 3 lines). Emerging flies, 2-6 hr post-eclosion, were separated by sex. Each sample had 70-80 heads per sex per treatment. Heads were dissected in RNAlater (Qiagen, Valencia, CA), pooled and stored at -20 °C until RNA extraction. Total RNA was extracted using Direct-zol RNA MiniPreps (Zymo, Irvine, CA) following the manufacturer's protocol.

For RNA-seq library preparation we used 100 ng of total RNA following the Truseq PolyA+ stranded mRNA kit (Illumina, San Diego, CA), as described in the manufacturer's protocol, except that we halved reaction volumes. External RNA Control Consortium (ERCC) spike-ins transcribed from a certified plasmid collection (SRM2374) were added to the Fragment-Prime-Finish Mix during the mRNA fragmentation step. We used the previously described RNAs from Pool 78A for all females feed with LPS and HPS and males of strain 6 feed with both diets and 78B for males of strain 2 and 3 feed with HPS (20pg /sample) [18]. Libraries were multiplexed using 12 differently barcoded adapters from TruSeq (Illumina, San Diego, CA). Libraries were quantified with Quant-iT PicoGreen (Life Technologies, Carlsbad, CA), and pooled for multiplexed sequencing. We sequenced in an Illumina HiSeq2500, to produce single-end, 75 bp reads (Illumina, San Diego, CA.).

Base calling was performed using CASAVA v1.8.2 (Illumina, San Diego, CA). BCL files were converted to FASTQ and demultiplexed using bcl2fastq v1.8 (Illumina, San Diego, CA). *D. melanogaster* reads were mapped to the major chromosome arms (chr2L, chr2R, chr3L, chr3R, chr4, chrX, chrY, and chrM) of the *Drosophila* genome assembly release 6 plus ISO1 MT and ERCC sequences using TopHat v2.1.0 [10] with default options except for the following: Libraries were stranded so mapping was done with --library-type fr-firststrand, and we provided a GTF with known set of gene models (FlyBase r6.09 and ERCC) using the -G option. The ERCC reference was based on the DNA

Sequence Library SRM2374 from NIST, the FASTA sequence was corrected according to differences found in [19]. Unique alignments were created by keeping only reads that aligned to a single location (NH:i:1) using grep (v2.20). HTseq v0.6.1p1 [11] was used to create gene level read counts with default options except (-s reverse) was used to get strand specific read counts. *Drosophila melanogaster* Release 6 plus ISO1 MT (GeneBank assembly accession: GCA_000001215.4).

Expression analysis

We utilize DESeq2 v1.12.4 to assess differential expression between diets, using built-in read count normalization [20]. DESeq2 utilize a Generalized Linear Model to count data assuming a negative binomial distribution. This approach uses shrinkage estimation for dispersion and fold changes, combining information across samples and all genes, to improve stability and interpretation of estimates. Unique gene read counts of each sample were fitted to a General Linear Model (Model = sex; Model = diet). Significance testing for group comparisons was done using the Wald test. Wald test estimates of the log2 fold change (LFC) were divided by standard error, resulting in a z-score, which we then converted to a *P*-value [20]. The *p*-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction for FDR at $p < 0.05$ [21]. For main exploratory analysis, we did not require a minimum LFC. Data available of GEO series accession (GSE82245).

Abbreviations

HPS: high protein:sugar ratio diet; LPS: low protein:sugar ratio diet; GO: Gene Ontology.

Supplementary Material

Table S1.

<http://www.jgenomics.com/v05p0128s1.xlsx>

Table S2.

<http://www.jgenomics.com/v05p0128s2.xls>

Acknowledgements

This work supported in part by the Intramural Research program of the National Institutes of Health, NIDDK (to BO), and by Consejo Nacional de Ciencia y Tecnología (CONACyT) grant CB-180385 (to TAM). We are grateful to Justin Fear for data handling advice and comments on the manuscript.

Authorship

TM designed the experiment. MJ, JH, and ML performed the experiment. MJ and BO analyzed the data and wrote the manuscript. All the authors read the manuscript.

Competing Interests

The authors have declared that no competing interest exists.

References

- Rajan A, Perrimon N. Of flies and men: insights on organismal metabolism from fruit flies. *BMC biology*. 2013; 11: 38.
- Goldman TD, Arbeitman MN. Genomic and functional studies of *Drosophila* sex hierarchy regulated gene expression in adult head and nervous system tissues. *PLoS genetics*. 2007; 3: e216.
- Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, et al. Diversity and dynamics of the *Drosophila* transcriptome. *Nature*. 2014; 512: 393-9.
- Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, et al. The developmental transcriptome of *Drosophila melanogaster*. *Nature*. 2011; 471: 473-9.
- Ye J, Cui X, Loraine A, Bynum K, Kim NC, White G, et al. Methods for nutrigenomics and longevity studies in *Drosophila*: effects of diets high in sucrose, palmitic acid, soy, or beef. *Methods in molecular biology* (Clifton, NJ). 2007; 371: 111-41.
- Fear JM, Leon-Novelo LG, Morse AM, Gerken AR, Van Lehmann K, Tower J, et al. Buffering of Genetic Regulatory Networks in *Drosophila melanogaster*. *Genetics*. 2016; 203: 1177-90.
- Lin Y, Chen ZX, Oliver B, Harbison ST. Microenvironmental Gene Expression Plasticity Among Individual *Drosophila melanogaster*. *G3* (Bethesda, Md). 2016; 6: 4197-210.
- Matzkin LM, Johnson S, Paight C, Bozinovic G, Markow TA. Dietary protein and sugar differentially affect development and metabolic pools in ecologically diverse *Drosophila*. *The Journal of nutrition*. 2011; 141: 1127-33.
- Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, et al. The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome research*. 2015; 25: 445-58.
- Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* (Oxford, England). 2009; 25: 1105-11.
- Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* (Oxford, England). 2015; 31: 166-9.
- Reed LK, Williams S, Springston M, Brown J, Freeman K, DesRoches CE, et al. Genotype-by-diet interactions drive metabolic phenotype variation in *Drosophila melanogaster*. *Genetics*. 2010; 185: 1009-19.
- Kuroda MI, Hilfiker A, Lucchesi JC. Dosage Compensation in *Drosophila*—a Model for the Coordinate Regulation of Transcription. *Genetics*. 2016; 204: 435-50.
- Bownes M. The regulation of the yolk protein genes, a family of sex differentiation genes in *Drosophila melanogaster*. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 1994; 16: 745-52.
- Sun W, Margam VM, Sun L, Buczkowski G, Bennett GW, Schemerhorn B, et al. Genome-wide analysis of phenobarbital-inducible genes in *Drosophila melanogaster*. *Insect molecular biology*. 2006; 15: 455-64.
- Honjo K, Mauthner SE, Wang Y, Skene JH, Tracey WD, Jr. Nociceptor-Enriched Genes Required for Normal Thermal Nociception. *Cell reports*. 2016; 16: 295-303.
- Nazario-Yepiz NO, Loustalot-Laclette MR, Carpinteyro-Ponce J, Abreu-Goodger C, Markow TA. Transcriptional responses of ecologically diverse *Drosophila* species to larval diets differing in relative sugar and protein ratios. *PLoS one*. 2017; 12: e0183007.
- Zook JM, Samarov D, McDaniel J, Sen SK, Salit M. Synthetic spike-in standards improve run-specific systematic error analysis for DNA and RNA sequencing. *PLoS one*. 2012; 7: e41356.
- Lee H, Pine PS, McDaniel J, Salit M, Oliver B. External RNA Controls Consortium Beta Version Update. *Journal of genomics*. 2016; 4: 19-22.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. 2014; 15: 550.
- Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 1990; 9: 811-8.