A Novel Dataset for Identifying Sex-Biased Genes in Drosophila

Nicholas W. VanKuren\textsuperscript{1,2} and Maria D. Vibranovski\textsuperscript{2,3,\*}

\begin{itemize}
\item \textsuperscript{1}Committee on Genetics, Genomics, and Systems Biology, The University of Chicago, Chicago IL 60637, USA;
\item \textsuperscript{2}Department of Ecology and Evolution, The University of Chicago, Chicago, IL 60637, USA;
\item \textsuperscript{3}Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil 05508
\end{itemize}

\*Corresponding author: Maria D. Vibranovski, E-mail: mdv@ib.usp.br

© Ivyspring International Publisher. This is an open-access article distributed under the terms of the Creative Commons License (http://creativecommons.org/licenses/by-nc-nd/3.0/). Reproduction is permitted for personal, noncommercial use, provided that the article is in whole, unmodified, and properly cited.

Published: 2014.03.10

Abstract

Phenotypic differences between males and females of sexually dimorphic species are caused in large part by differences in gene expression between the sexes, most of which occurs in the gonads. To accurately identify genes differentially expressed between males and females in Drosophila, we sequenced the testis and ovary transcriptomes of D. yakuba, D. pseudoobscura, and D. ananassae and used them to identify sex-biased genes in the latter two species. We highlight the increased sensitivity and improved power of sex-biased gene detection methods when using our testis/ovary data versus male and female whole body transcriptome data. We thus provide a resource specifically designed to accurately identify and characterize sex-biased genes across Drosophila. This dataset is available through NCBI GEO accession GSE52058.

Key words: RNA-seq, testis, ovary, Drosophila pseudoobscura, Drosophila ananassae, Drosophila yakuba

INTRODUCTION

Differences in gene expression account for the majority of phenotypic differences between males and females of sexually dimorphic species [1]. Thus, accurate identification of genes differentially expressed between males and females, \textit{i.e.} sex-biased genes, is crucial for understanding the current state and evolution of the genomic architecture and mechanisms producing sexual dimorphism. The majority of sex-biased expression detected with microarrays in Drosophila occurs in gonads (\textit{e.g.} [2]), suggesting that accurate identification of sex-biased genes should be based on gene expression measurements in these sex-specific organs. To take advantage of the increased sensitivity of whole-transcriptome sequencing (RNA-seq), to avoid the limitations of whole body (WB) samples for detecting sex-biased gene expression, and to better understand sex-biased gene evolution in Drosophila, we sequenced testis and ovary mRNAs from D. pseudoobscura, D. ananassae, and D. yakuba.

METHODS

Flies were grown on cornmeal-molasses agar at 20°C (D. pseudoobscura 14011-0121.94) or 25°C (D. ananassae 14024-0371.13 and D. yakuba CSN). Virgin flies were collected and aged 6-10 days before dissecting 2-3 replicates of testes or ovaries. Total RNA was extracted from testis and ovary samples using the Arcturus® PicoPure® kit. Illumina® TruSeq® RNA kits were then used to poly-A+ select mRNA, reverse-transcribe mRNA using random priming, shear cDNA into 120-200 bp fragments, and produce libraries for 1x50 bp sequencing on an Illumina GAIIx or HiSeq2000 (Table S1). Illumina’s Real Time Analysis v1.13 module processed raw images, called bases, and provided base qualities. We downloaded D.
We identified sex-biased genes in WB, reproductive tract, or testis/ovary samples using Cuffdiff v2.1.0 with default options, which include pooled sample dispersion estimates and geometric normalization of gene-level counts [6], and edgeR v3.4.0 [7]. We generated gene-level count data for edgeR with HTSeq v0.5.4p3 using uniquely-mapped reads and the intersection-nonempty method to assign reads to HTSeq v0.5.4p3 using uniquely-mapped reads and our D. yakuba samples currently consist of one replicate and are likely unsuitable for sex-biased gene identification.

We Cuffdiff and edgeR analyses genes were called sex-biased if the Benjamini-Hochberg [10] false discovery rate was < 0.01.

RESULTS

Cuffdiff and edgeR results are shown in Table 1. In general, Cuffdiff resulted in greater overlap than edgeR of the sex-biased genes found in both WB and testis/ovary analyses (Pearson’s χ², p < 1e-04), while edgeR was more sensitive. There are two key points to Table 1. First, testis/ovary analyses detect more (D. ananassae: 3.3 - 5.0-fold; D. pseudoobscura: 1 - 1.4-fold) sex-biased genes than WB analyses (Pearson’s χ², all p<1e-04). Second, testis/ovary analyses significantly improve our power to detect the smallest class of sex-biased genes found in WB analyses. For example, 5.5-25.3-fold more female-biased genes are found in D. ananassae testis/ovary analyses than WB analyses (Pearson’s χ², p<1e-04; Table 1).

We examined the magnitude of the log fold change of expression levels between testis and ovary or male and female whole body to better understand the difference between the two analyses’ results. Male-biased genes (MBGs) and female-biased genes (FBGs) show larger magnitudes of log; fold changes (i.e. log[expression level in male tissue/ expression level in female tissue]) in testis/ovary analyses than in WB analyses (Figure S1). Three different scenarios could account for this pattern. For MBGs, for example, higher log₂ fold change in expression in testis/ovary relative to WB analyses could be caused by i) lower expression in ovary than in female WB, ii) higher expression in testis than in male WB, or iii) both higher expression in testis and lower expression in ovaries relative to WBs. We examined genes called sex-biased in testis/ovary but not in WB Cuffdiff analyses. Consistent with scenario iii), MBGs have significantly lower expression in ovary and higher expression in testis relative to female and male WB, respectively, in both species (t-tests, all p < 1e-05). FBGs also follow scenario iii) (t-tests, p<1e-05), except D. pseudoobscura female expression levels are not different between WB and ovary. Similar D. pseudoobscura WB and ovary FBG expression levels may be expected if FBGs are enriched with broadly-expressed genes as they are in D. melanogaster [5,11]. Except for the latter observation, these general results are consistent with the idea that gonad samples "concentrate" sex-biased expression relative to WB.

Table 1. Differential expression analyses of whole body and sex-specific organs in Drosophila pseudoobscura and D. ananassae.

| Comparison | Cuffdiff | edgeR |
|------------|----------|-------|
| D. pseudoobscura |          |       |
| whole body | 5269     | 2785  | 2484 |
| testis | 7105     | 3184  | 3921 |
| reproductive tract | 9067     | 4669  | 4398 |
| Overlap (%) | 4477     | 2334  | 2143 |
| (85.0) | (83.8) | (86.3) | (92.4) |
| Total DEa | 13252    | 12575 |
| MBa | 5264     | 5242  |
| FBa | 12738    |       |
| Total Tested | 8284     | 8045  |
| MB | 3043     | 3292  |
| FB | 12738    |       |
| D. ananassae |          |       |
| whole body | 1791     | 1613  | 178 |
| testis-ovary | 8997     | 4494  | 4503 |
| Overlap (%) | 1657     | 1503  | 154 |
| (92.5) | (93.2) | (86.5) | (94.5) |
| Total DEa | 13786    | 13252 |
| MBa | 5224     | 3292  |
| FBa | 13081    |       |
| Total Tested | 3224     | 9429  |
| MB | 2138     | 3456  |
| FB | 13081    |       |

a DE: differentially expressed at false discovery rate <0.01; MB: male-biased; FB: female-biased
b Annotated genes: 16,755
c Numbers and percentages (of smallest value) of genes overlapping between whole body and testis-ovary analyses
d Annotated genes: 16,225

http://www.jgenomics.com
In contrast to sex-biased genes, genes that were tested and unbiased in both testis/ovary and WB analyses do not have significantly different expression levels in whole male/testis or whole female/ovary in either species (t-tests, all \( p > 0.05 \)), except \( D. \) ananassae whole female expression levels are significantly higher than ovary levels (t-test, \( p < 1 \times 10^{-5} \)). This could indicate that \( D. \) ananassae ovary RNA contributes less to the WB RNA pool relative to other species \([2]\), resulting in less detectable female bias in WB samples. These results also highlight the utility of this dataset for determining differences in sex-bias between \( Drosophila \) species, and to assess fine-scale differences in expression across the genus.

Finally, more MB and FB genes were detected in \( D. \) pseudoobscura reproductive tract samples than testis/ovary analyses (Table 1), which agrees with the hypothesis that the majority of sex-biased gene expression occurs in sex-specific organs. For instance, Drosophila male reproductive tracts include seminal vesicles and accessory glands, which have additional sex-biased genes not expressed in testis. Expression profiles of those particular sex-specific organs would also improve the assessment of sex-biased genes.

**ACKNOWLEDGMENTS**

We thank Sidi Chen for help with RNA-seq. MDV was supported by National Institutes of Health grant R01GM078070-01A1, NIH ARRA supplement grant R01 GM078070-03S1 and the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for travel (2013/09714-6). NWV was supported by a National Science Foundation Graduate Research Fellowship and NIH Genetics and Regulation Training Grant T32 GM007197.

Author contributions: MDV designed the study, NWV and MDV collected the samples, NWV analyzed the data, NWV and MDV wrote the paper.

**COMPETING INTERESTS**

The authors have declared that no competing interest exists.

**REFERENCES**

1. Ellegren H, Parsch J. The evolution of sex-biased genes and sex-biased gene expression. Nat Rev Gen. 2007; 8:689-98.
2. Zhang Y, Sturgill D, et al. Constraint and turnover in sex-biased gene expression in the genus Drosophila. Nature. 2007; 450:233-7.
3. Celniker SE, Dillon L, Gerstein M, et al. Unlocking the secrets of the genome. Nature. 2009; 459:927-30.
4. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nat Meth. 2012; 9:357-9.
5. Assis R, Zhou Q, Bachtrog D. Sex-biased transcriptome evolution in Drosophila. Genome Biol Evol. 2011; 4:1189-200.
6. Trapnell C, Hendrickson D, Sauveveu M, et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotech. 2013; 31:36-53.
7. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139-40.
8. [Internet] Anders S. http://www-huber.embl.de/users/anders/HTSeq/.
9. Risso D, Schwartz K, Sherlock G, et al. GC-content normalization for RNA-Seq data. BMC Bioinformatics. 2010; 12:480-97.
10. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J Royal Stat Soc B. 1995; 57:289-300.
11. Meisel RP. Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. Mol Biol Ecol. 2011; 28:1893-900.
Appendix A

Table S1. GEO, SRA, and modENCODE accessions and mapping statistics of datasets used in this study.

| Species  | Tissue               | Rep   | Platform (Illumina) | Source         | GEO Accession | modENCODE Accession | Total reads (x10^6) | Mapped reads (x10^6) | % Mapped |
|----------|----------------------|-------|---------------------|----------------|---------------|---------------------|---------------------|---------------------|----------|
| D. ananassae | Male whole body     | 1     | 2x100 GAIIx        | modENCODE     | GSM1091847  | NA                  | 31.8                | 28.1                | 88.4     |
|          | Male whole body     | 2     | 2x100 GAIIx        | modENCODE     | GSM1091848  | NA                  | 123.6               | 54.0                | 43.7     |
|          | Female whole body   | 1     | 2x100 GAIIx        | modENCODE     | GSM604275   | 3617                | 43.3                | 31.4                | 72.5     |
|          | Female whole body   | 2     | 2x100 GAIIx        | modENCODE     | GSM684276   | 3617                | 134.9               | 66.4                | 49.2     |
|          | Testis              | 1     | 1x50 HiSeq2000     | this paper    | GSM1258041  | -                   | 81.1                | 74.5                | 91.9     |
|          | Testis              | 2     | 1x50 HiSeq2000     | this paper    | GSM1258042  | -                   | 72.6                | 66.7                | 91.9     |
|          | Ovary               | 1     | 1x50 HiSeq2000     | this paper    | GSM1258043  | -                   | 72.2                | 65.9                | 91.2     |
|          | Ovary               | 2     | 1x50 HiSeq2000     | this paper    | GSM1258044  | -                   | 88.9                | 73.5                | 90.8     |
| D. pseudoobscura | Male whole body   | 1     | 2x100 GAIIx        | modENCODE     | GSM694281   | 3620                | 53.2                | 21.2                | 39.9     |
|          | Male whole body     | 2     | 2x100 GAIIx        | modENCODE     | GSM694282   | 3620                | 145.7               | 26.1                | 17.9     |
|          | Female whole body   | 1     | 2x100 GAIIx        | modENCODE     | GSM694279   | 3621                | 48.7                | 20.2                | 41.5     |
|          | Female whole body   | 2     | 2x100 GAIIx        | modENCODE     | GSM694280   | 3621                | 149.4               | 29.4                | 19.7     |
|          | Male rep. tract     | 1     | 2x75 HiSeq2000     | modENCODE     | GSM775500   | 4050                | 212.9               | 167.0               | 78.5     |
|          | Male rep. tract     | 2     | 2x75 HiSeq2000     | modENCODE     | GSM775501   | 4050                | 252.0               | 214.4               | 85.1     |
|          | Female rep. tract   | 1     | 2x75 HiSeq2000     | modENCODE     | GSM775498   | 4049                | 231.5               | 97.1                | 41.9     |
|          | Female rep. tract   | 2     | 2x75 HiSeq2000     | modENCODE     | GSM775499   | 4049                | 227.6               | 105.8               | 46.5     |
|          | Testis              | 1     | 1x50 HiSeq2000     | this paper    | GSM1258036  | -                   | 99.1                | 92.0                | 92.8     |
|          | Testis              | 2     | 1x50 HiSeq2000     | this paper    | GSM1258037  | -                   | 73.5                | 68.5                | 93.2     |
|          | Ovary               | 1     | 1x50 GAIIx         | this paper    | GSM1258038  | -                   | 5.4                 | 4.7                 | 86.6     |
|          | Ovary               | 2     | 1x50 GAIIx         | this paper    | GSM1258039  | -                   | 44.2                | 39.8                | 90.0     |
|          | Ovary               | 3     | 1x50 GAIIx         | this paper    | GSM1258040  | -                   | 5.5                 | 4.9                 | 89.0     |
| D. yakuba | Testis              | 1     | 1x50 GAIIx         | this paper    | GSM1258045  | -                   | 104.1               | 81.8                | 78.6     |
|          | Ovary               | 1     | 1x50 GAIIx         | this paper    | GSM1258046  | -                   | 2.0                 | 1.7                 | 84.4     |
|          | Ovary               | 2     | 1x50 GAIIx         | this paper    | GSM1258047  | -                   | 19.0                | 16.9                | 89.0     |
|          | Ovary               | 3     | 1x50 GAIIx         | this paper    | GSM1258048  | -                   | 76.9                | 65.2                | 84.8     |

* modENCODE reads were not trimmed either before or during mapping, which may account for some low mapping percentages.
Figure S1. Testis versus ovary comparisons result in significantly greater magnitudes of fold change relative to whole body comparisons. WB = whole body comparison, TO = testis-ovary comparison, MBGs = male-biased genes, FBGs = female-biased genes, logFC = log fold change (female / male). The y-axis is the absolute value of the ratio of normalized expression values of female whole body to male whole body or ovary to testis. TO logFCs are significantly higher than their WB counterparts in every case (t-test, p-value < 2.2e-16), except D. ananassae WB FBGs show greater logFCs (Cuffdiff: P<2.2e-16; edgeR: P=3.7e-09).