Please cite this article in press as: Coolon et al., Genomic Imprinting Absent in *Drosophila melanogaster* Adult Females, Cell Reports (2012), http://dx.doi.org/10.1016/j.celrep.2012.06.013

Cell Reports
Report

Genomic Imprinting Absent in *Drosophila melanogaster* Adult Females

Joseph D. Coolon,[1,*] Kraig R. Stevenson,[2] C. Joel McManus,[3,4] Brenton R. Graveley,[4] and Patricia J. Wittkopp[1,2]
[1]Department of Ecology and Evolutionary Biology
[2]Department of Computational Medicine and Bioinformatics
University of Michigan, Ann Arbor, MI 48109, USA
[3]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[4]Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, CT 06030, USA
*Correspondence: jcoolon@umich.edu
http://dx.doi.org/10.1016/j.celrep.2012.06.013

SUMMARY

Genomic imprinting occurs when expression of an allele differs based on the sex of the parent that transmitted the allele. In *D. melanogaster,* imprinting can occur, but its impact on allelic expression genome-wide is unclear. Here, we search for imprinted genes in *D. melanogaster* using RNA-seq to compare allele-specific expression between pools of 7- to 10-day-old adult female progeny from reciprocal crosses. We identified 119 genes with allelic expression consistent with imprinting, and these genes showed significant clustering within the genome. Surprisingly, additional analysis of several of these genes showed that either genomic heterogeneity or high levels of intrinsic noise caused imprinting-like allelic expression. Consequently, our data provide no convincing evidence of imprinting for *D. melanogaster* genes in their native genomic context. Elucidating sources of false-positive signals for imprinting in allele-specific RNA-seq data, as done here, is critical given the growing popularity of this method for identifying imprinted genes.

INTRODUCTION

More than 50 years ago, Helen Crouse coined the term ''imprinting'' to describe a case in Sciard flies in which the sex of the parent influenced the inheritance of a chromosome (Crouse, 1960). Since that time, the definition of imprinting has been expanded to include any parent-of-origin-dependent chromosome marking, especially those causing differential gene activity or expression (Ferguson-Smith, 2011). Recently, genomic scans for imprinting at the level of RNA abundance in plants and mammals have shown that (1) only a small percentage of genes (typically on the order of 100 genes) appear to be imprinted (Babak et al., 2008; Gehring et al., 2011; Hsieh et al., 2011; Luo et al., 2011; Wang et al., 2008, 2011; Waters et al., 2011; Wolff et al., 2011); (2) these genes are sometimes found in clusters within the genome (Ferguson-Smith, 2011; Wood and Oakey, 2006); and (3) their imprinting is often required for normal development (McGrath and Solter, 1984; Surani et al., 1984) and physiology (Buiting et al., 1995; Weksberg et al., 1993).

In *Drosophila melanogaster*, studies of imprinting have yielded conflicting results. Euchromatic genes inserted onto the heterochromatic Y chromosome and genes located on chromosomes with deletions, duplications, rearrangements, and/or translocations can show differences in their activity depending on the parent from which they are inherited, demonstrating that *D. melanogaster* is capable of imprinting (Anaka et al., 2009; Golic et al., 1998; Haller and Woodruff, 2000; Joanis and Lloyd, 2002; Lloyd et al., 1999; MacDonald et al., 2010; Maggert and Golic, 2002; Menon and Meller, 2009). However, when Wittkopp et al. (2006) tested for evidence of imprinting by analyzing allele-specific expression of eight genes that showed strong parent-of-origin effects on total gene expression in a genomic survey of *D. melanogaster* (Gibson et al., 2004), no evidence of imprinting was observed. Furthermore, gynogenetic and androgenetic *D. melanogaster*, which inherit all of their genetic information from a single parent, are viable, suggesting that imprinting is not essential in this species (Fuyama, 1984; Komma and Endow, 1995). Consequently, even though it is clear that *D. melanogaster* can form parent-of-origin-specific imprints that affect gene activity, the prevalence of imprinted genes in their native genomic context within the *D. melanogaster* genome remains unclear (Menon and Meller, 2010).

RESULTS AND DISCUSSION

To search for imprinting genome-wide, we used Illumina sequencing in conjunction with a novel bioinformatics pipeline to infer allele-specific RNA transcript abundance in progeny from reciprocal crosses. This method uses transcribed sequence polymorphisms to distinguish sequencing reads derived from each of the two parental alleles in F_1 offspring. To maximize the proportion of sequencing reads informative for allele-specific expression, we used a cosmopolitan (M-type) and an African (Z-type) line of *D. melanogaster* (Hollocher et al., 1997). The M-type line used was the zygotic hybrid rescue line (zhr) first described by Sawamura and colleagues (1993) and the Z-type line was a Zimbabwean isofemale line (z30) isolated in 1990 (Begun and Aquadro, 1993; Wu et al., 1995). To improve the accuracy of allele assignments, we sequenced the M-type (zhr)

Please cite this article in press as: Coolon et al., Genomic Imprinting Absent in *Drosophila melanogaster* Adult Females, Cell Reports (2012), http://dx.doi.org/10.1016/j.celrep.2012.06.013

Cell PRESS
Open ACCESS

and Z-type (z30) genomes to 23.2X and 21.5X coverage (Table S1) and used these data to assemble line-specific genomic sequences (see Extended Experimental Procedures).

M-type females were crossed to Z-type males, producing F_1 hybrids hereafter referred to as MZ. Likewise, Z-type females were crossed to M-type males, producing F_1 hybrids hereafter referred to as ZM (Figure S1A). MZ and ZM hybrid flies were collected 7–10 days after eclosion, and total RNA was extracted from a pool of 20 hybrid females for each genotype. MZ and ZM RNA samples were used to make cDNA sequencing libraries, which were sequenced using an Illumina GAIIx machine. The resultant paired-end (2X76bp) sequencing reads (Table S1) were aligned to the strain-specific M-type and Z-type genomes. Using two strain-specific genome sequences for mapping avoids mapping biases introduced by using only a single reference genome (Degner et al., 2009; Graze et al., 2012). Of the reads from the MZ and ZM samples, 86% and 87%, respectively, were aligned without mismatches to unique genomic loci (Table S1). In each case, 21% of the uniquely mapping reads aligned perfectly to only one genome and were used to infer allele-specific expression (Figure S1B; Table S1).

The power to infer allele-specific expression using RNA-seq data (which is necessary to test for imprinting with this method) depends upon the expression level of a gene, as well as the density of transcribed polymorphisms within it (Fontanillas et al., 2010). Prior work has shown that obtaining at least 20 allele-specific reads for a gene results in reproducible measures of relative allelic expression (McManus et al., 2010). Retaining only genes with 20 or more allele-specific reads (allele 1 + allele 2 $\geq$ 20) in both the MZ and ZM samples, 7,206 genes were tested for allelic expression patterns consistent with imprinting (Table S2). This includes 3% of the 4,875 genes with a number of fragments per kilobase per million mapped reads (FPKM) less than 1, 51% of the 1,706 genes with an FPKM between 1 and 5, and 83% of the 7,430 genes with an FPKM greater than 5 (Figure S2). The modENCODE consortium used a threshold of FPKM = 1 to classify *D. melanogaster* genes as "expressed" or "not expressed" (Graveley et al., 2011) and according to this definition, we tested 77% of the 9,136 genes expressed (in the 7- to 10-day-old adult females we examined) for imprinting.

To assess the accuracy of our allele-specific expression measurements, we compared the allelic expression ratios determined by RNA-seq to estimates from pyrosequencing (Ahmadian et al., 2000) of individual genes. Ten genes selected at random were used for pyrosequencing of the same MZ and ZM samples used for RNA-seq (Table S3). Pyrosequencing measurements were highly correlated (R^2 = 0.88) with estimates from RNA-seq (Table S3; Figure S3A), suggesting that RNA-seq produces reliable measures of relative allelic expression. This is consistent with previous comparisons of RNA-seq and pyrosequencing measures of allelic expression that used distinct bioinformatic pipelines (McManus et al., 2010; Emerson et al., 2010).

To identify genes that might be imprinted, we tested for differences in relative allele-specific expression between MZ and ZM using the Fisher's exact test (FET). This test evaluates whether differential allelic expression (when present) is equal in magnitude and direction in the two genotypes. At a false discovery rate (FDR) of 5%, 119 (1.65%) of the 7,206 genes analyzed had significant differences (FET, q < 0.05) in relative allelic expression between the two types of F_1 hybrid progeny (Figure 1; Table S2). To evaluate the accuracy of RNA-seq measurements of allele-specific expression specifically for putatively imprinted genes (PIGs), we used pyrosequencing to independently measure allele-specific expression for four genes in this class using the same ZM and MZ samples as those used for RNA-seq. We again observed strong concordance (R^2 = 0.85, Figure S3B) between pyrosequencing and RNA-seq measures of allele-specific expression, suggesting that inaccurate quantification of expression levels in cDNA pools by RNA-seq is unlikely to explain the observed differences in relative allelic expression between hybrid genotypes.

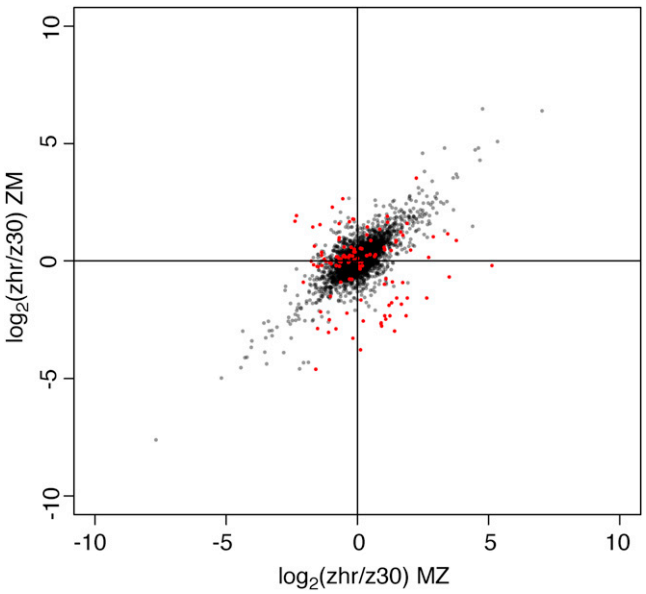

Figure 1. Allelic Expression from Reciprocal Crosses Suggests that <2% of Genes in the Genome Might Be Imprinted

Log_2-transformed allelic expression ratios (zhr/z30) from MZ on the *x* axis and log_2(zhr/z30) allelic expression ratio from ZM on the *y* axis. Each point represents one gene. Points are color-coded by significance in false-discovery-rate-corrected Fisher's exact tests, where red points indicate q < 0.05. Note that the power to detect differences in allelic expression between ZM and MZ differs from gene to gene and is dependent upon the number of Illumina sequencing reads obtained that map to that gene. See also Figures S1–S3 and Tables S1 and S2.

Putatively Imprinted Genes Are Clustered in the Genome

In mammals, imprinted genes are often found in clusters throughout the genome (Ferguson-Smith, 2011; Wood and Oakey, 2006), and this clustering might relate to the mechanism by which they are regulated (Caspary et al., 1998; Mancini-Dinardo et al., 2006; Lewis et al., 2004; Lopes et al., 2003). To determine if this was also true for the PIGs in the *D. melanogaster* genome, we used a sliding-window Monte Carlo sampling approach with FDR-corrected approximate permutation tests to investigate potential clustering. We found that there were four regions in the *D. melanogaster* genome

Cell PRESS
Open ACCESS

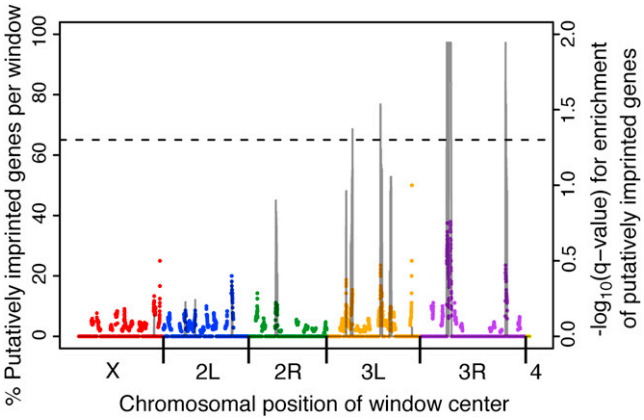

Figure 2. Putatively Imprinted Genes Clustered Significantly on Chromosomes

Using a sliding-window analysis, the proportion of genes within each 500 kb window that were identified as putatively imprinted is indicated for positions across the genome. Each chromosome arm is indicated on the *x* axis, with one point representing each window. Using a Monte Carlo sampling approach and approximate permutation tests that control for differences in the number of genes within each window, and following these steps with a multiple testing correction, we identified regions of the genome that were significantly enriched for PIGs. FDR-corrected p-values are indicated by the solid line, and the dotted line indicates the threshold used to identify significant clusters ($q < 0.05$).

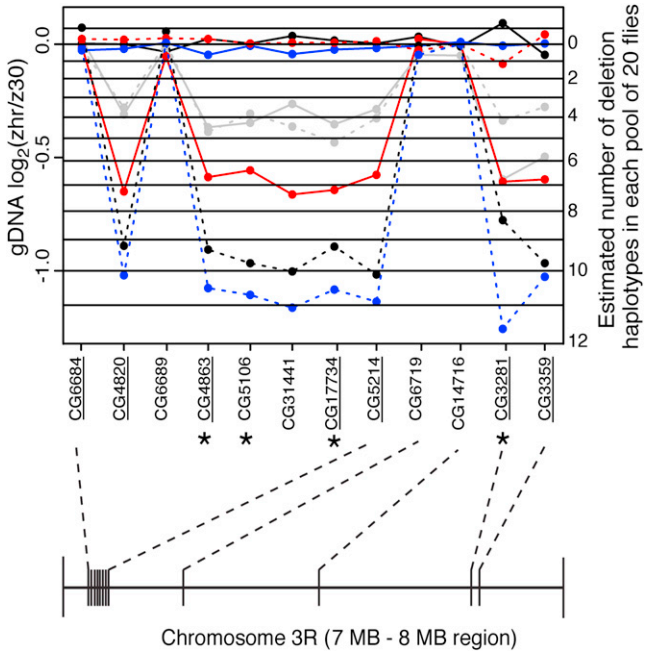

Figure 3. Replicate Pools of Flies Showed Different Allele Frequencies in Genomic DNA for Putatively Imprinted Genes Located in a Cluster

For 12 genes in the region containing the largest cluster of putatively imprinted loci (7,000,000–8,000,000 on chromosome 3R), seven that were identified as putatively imprinted (underlined) and five that were not, we used pyrosequencing to determine the relative abundance of zhr and z30 alleles in genomic DNA in additional biological replicate pools containing 20 F_1 heterozygous flies each. The log2(zhr/z30) ratio is plotted for gDNA from each biological replicate pool, with the four ZM pools indicated by solid lines and the four MZ pools indicated by dotted lines. Replicates are arbitrarily colored blue, gray, red and black. The genomic arrangement of these genes is shown below the plot. Genes labeled with an asterisk were also genotyped in individual flies (Table S4). Note that *CG6684* is underlined because it showed significant evidence of allelic expression differences between MZ and ZM in the RNA-seq data; however, this gene does not appear to be included within the deleted region(s). Pyrosequencing analysis of *CG6684* showed no evidence of differential allelic expression between MZ and ZM and normal variance among replicate biological samples, suggesting that it was a false positive in the RNA-seq data. *CG5106* and *CG31441* appear to be included within the deleted region but showed no significant evidence of an imprinting-like pattern of allelic expression in the RNA-seq data, probably due to lack of power, as these two genes had the lowest read counts of those tested. See also Tables S3 and S4.

that showed significant clustering (permutation test, $q < 0.05$) of PIGs (Figure 2). Interestingly, all four significant clusters were found on chromosome 3, with two on the left arm (3L) and two on the right arm (3R) of the chromosome. Together, these four regions contain 27% (32/119) of the PIGs, with one cluster located on chromosome arm 3R (6,550,000–8,280,000) containing 17% (20/119) of all PIGs (Figure 2). Clustering of PIGs in the genome is consistent with previously described mechanisms of imprinting, but it could also be caused by other factors.

Low-Frequency Deletion(s) Responsible for Some Cases of Apparent Imprinting

To further test for evidence of imprinting, we more closely examined 12 genes within the largest and most significant cluster of PIGs (3R 6.5–8.3 MB region, Figure 2). Seven of these genes were PIGs and five were genes that showed no significant differences in relative allelic expression between ZM and MZ. Pyrosequencing was again used to obtain an independent measure of relative allelic expression, except that instead of testing the same biological sample used for RNA-seq (as described above), we analyzed four independent biological replicate pools of ZM and MZ flies, each containing twenty 7- to 10-day-old adult females (Table S3). From each pool, we sequentially extracted genomic DNA (gDNA) and RNA.

F_1 flies produced by crossing two highly inbred lines are expected to be genetically identical; thus, analysis of gDNA serves as a control for differential amplification of the two alleles during PCR prior to pyrosequencing (Landry et al., 2005; Wittkopp, 2011; Wittkopp et al., 2004, 2008). Surprisingly, and unlike the case for the 34 genes located outside of clustered PIGs that we analyzed (data not shown), relative allelic abundance differed greatly for the gDNA samples among the biological replicates—between the MZ and ZM genotypes as well as among replicate MZ or ZM samples (Figure 3). When present, deviations from equal allelic abundance in gDNA were similar for genes throughout the cluster within a replicate pool but differed among pools. The M-type (zhr) allele was always the allele underrepresented (Figure 3).

A polymorphic deletion(s) in the M-type (zhr) strain or a polymorphic duplication(s) in the Z-type (z30) strain could account for the differences in gDNA content observed among replicate pools of F_1 flies. To directly test for evidence of a deletion or

duplication, we used pyrosequencing to genotype 48 individual F_1 progeny (24 MZ and 24 ZM) at four loci within the 3R 6.5–8 MB region that showed a cluster of PIGs (indicated with asterisks in Figure 3), as well as at two loci on other chromosomes. All but two of the 48 hybrid flies showed evidence of one M-type and one Z-type allele at all six loci tested, as expected. The remaining two hybrids showed evidence of only the Z-type (z30) allele at the four loci within the cluster, but both flies showed both alleles at the two loci tested on other chromosomes (Table S4); the presence of these heterozygous sites demonstrates that these two flies are in fact F_1 hybrids and not contaminating flies with parental genotypes. Based on these data, we conclude that the M-type (zhr) strain contains one or more deletion(s) in this region on 3R that remains heterozygous despite years of inbreeding followed by 10 generations of pair mating immediately prior to the start of this experiment. Residual heterozygosity such as this has also been reported in *D. melanogaster* following extensive inbreeding in lines used for genomic sequencing (Mackay et al., 2012).

The presence of this deletion haplotype at low frequency in the zhr line used to produce MZ and ZM hybrids suggests that differences in its frequency in the pools of 20 MZ and 20 ZM hybrid flies used for RNA-seq are more likely than imprinting to be responsible for the observed difference in relative allelic expression. Indeed, after controlling for differences in the alleles present in gDNA among the replicate pools analyzed by pyrosequencing (see Experimental Procedures), relative allelic expression in cDNA samples was not significantly different ($p > 0.05$ for all tests). It remains to be seen whether genotypic differences between the MZ and ZM pools of flies used for RNA-seq are also responsible for differences in relative allelic expression observed for other clustered PIGs, but we believe it is likely.

Nonclustered PIGs Have Higher-than-Normal Intrinsic Noise

Our initial RNA-seq survey for imprinting identified as PIGs all genes with significant differences in relative allelic expression between F_1 hybrid progeny from reciprocal crosses; however, imprinting is often defined in a more limited way, such that only one allele of a gene (either the maternally or paternally inherited allele) accounts for the majority (or all) of the expression of the imprinted gene. Among the original set of 119 PIGs, only 18 showed patterns of allelic expression consistent with this more strict definition (Table S2; Figure S4), and none of these were located in the clusters described above (Figure 2). To further test these 18 genes for evidence of imprinting, we analyzed allelic expression for each gene in the MZ and ZM biological replicates described in the preceding section (Table S3). Unlike for clustered PIGs examined in these samples, no significant differences in allele frequency were found among replicate gDNA samples for any of these 18 genes.

The relative allelic expression for these genes in the four MZ and four ZM biological replicates was still not typical; however, these 18 genes showed greater variance in relative allelic expression among the biological replicate pools than most genes that we have analyzed with pyrosequencing. Indeed, a Wilcoxon rank-sum test showed that the standard errors of $\log_2$-transformed allelic expression ratios were significantly greater for the 18 PIGs than for 16 genes selected at random ($W = 260$, $p = 2.68 \times 10^{-7}$; Figure 4). Additional statistical tests showed no evidence for imprinting of these genes ($q > 0.05$ for all tests). Given (1) the high degree of variability we observed for these genes among replicate pools with the same genotype (MZ or ZM), (2) the lack of evidence for imprinting found by pyrosequencing, and (3) that we only analyzed one pool of flies for each genotype by RNA-seq, we conclude that significant differences observed between MZ and ZM for relative allelic expression in the RNA-seq data are most likely caused by sampling error.

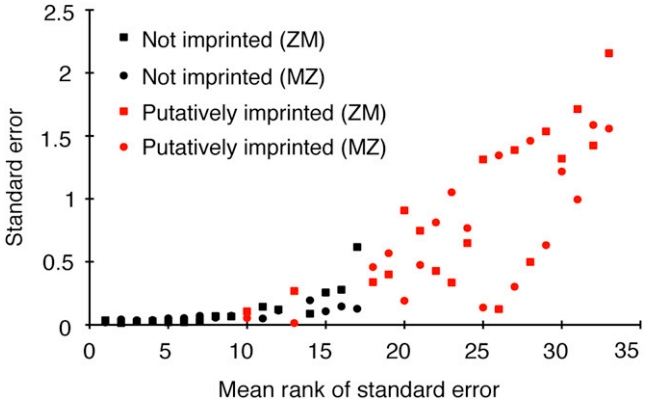

Figure 4. Putatively Imprinted Genes Have High Intrinsic Noise

For each gene for each sample type (ZM or MZ), the standard error for $\log_2$-transformed allelic expression ratios from biological replicate pools of flies is shown, with black points representing genes selected at random from the genome (none of which showed significant evidence of imprinting) and red points representing PIGs. Square marks represent the ZM sample and circles represent the MZ sample, with one rank for each gene tested. The *x* axis is rank of standard error for the two samples for each gene. See also Figure S4 and Table S3.

What Role Does Imprinting Play in Regulating *D. melanogaster* Gene Expression?

As described above, RNA-seq analysis (validated by pyrosequencing) identified 119 of 7,206 genes as having differences in relative allele-specific expression in reciprocal hybrids; however, analysis of gDNA and cDNA from additional replicate biological samples identified other factors (the presence of a polymorphic deletion(s) and using a single measurement to represent a highly variable phenotype) that are more likely than imprinting to be responsible for the differences in allelic expression observed in our RNA-seq data. Consequently, we conclude that these data provide no convincing evidence that imprinting affects expression of endogenous *D. melanogaster* genes in their native genomic contexts—at least in the 7- to 10-day-old adult females we examined.

Given the evidence of imprinting in other studies of *D. melanogaster*, why do we fail to find evidence of it in our genomic analysis? We cannot rule out the possibility that imprinting affects allelic activity in males, at other developmental stages, in limited tissues (with the signal masked by the absence of imprinting in the majority of cells sampled), or for genes with

Cell PRESS
Open ACCESS

expression and/or polymorphism levels that cause them to be below our detection threshold, but there is also no evidence suggesting that imprinting is occurring under any of these conditions. In addition, as described by Menon and Meller (2010), evidence of imprinting in *D. melanogaster* comes from studying particular genotypes, and imprinting might not impact gene expression in all genotypes: "In *Drosophila*, imprints are detected by alteration in expression of genes on rearranged chromosomes, but there is little to suggest that expression of any gene in karyotypically normally (*sic*) flies is governed by imprinting." We tested 77% of the expressed genes in the *D. melanogaster* genome for imprinting in this study, and evidence that imprinting affects the expression of genes in their native genomic context is still lacking.

Genomic Surveys for Imprinting Using RNA-seq: Proceed with Caution

In addition to providing insight into imprinting in *D. melanogaster*, this study identifies important considerations for using RNA-seq to test for imprinting in any species. RNA-seq has been used to search for imprinted loci in both plants and animals, including mouse (Babak et al., 2008; Wang et al., 2008, 2011), *Arabidopsis* (Gehring et al., 2011; Hsieh et al., 2011; Wolff et al., 2011), maize (Waters et al., 2011), and rice (Luo et al., 2011); but this approach is not without its pitfalls. For example, a study using RNA-seq to identify imprinted genes in various mouse tissues reported over 1,000 imprinted loci (Gregg et al., 2010a, 2010b), but most of these loci were subsequently shown to be false positives caused by biased sequencing and the failure to measure and account for technical and biological variability (DeVeale et al., 2012).

Data presented here and in DeVeale et al. (2012) clearly show the importance of validating putatively imprinted genes identified by RNA-seq with independent techniques (and, ideally, independent biological samples) prior to concluding that they are imprinted. To focus validation efforts on the loci most likely to be imprinted, RNA-seq experiments should include both biological and technical replicates, as well as, whenever possible, the analysis of gDNA extracted from the same tissue homogenate as the RNA. This final control is particularly important when working with small organisms (e.g., flies), for which multiple inbred individuals (that could have residual heterozygosity) are typically pooled prior to RNA extraction and cDNA sequencing, but it can also detect and control for differences in genomic content that might exist among cells from the same individual due to somatic mutations. For example, Shibata et al. (2012) have recently shown that microdeletions can cause genomic heterogeneity among mouse and human cells. Sequencing gDNA and cDNA derived from the same tissue sample can also allow corrections for bias introduced during the library preparation and sequencing. With more and more researchers turning to RNA-seq to study genomic imprinting, it is important to keep these caveats in mind.

EXPERIMENTAL PROCEDURES

Fly Strains, Rearing, and Collections

The *D. melanogaster* strain zhr carrying the hybrid rescuing Zhr1 chromosome (full genotype, XYS.YL.Df(1)Zhr; Ferree and Barbash, 2009; Sawamura et al., 1993) and the Zimbabwean isofemale line z30 (Begun and Aquadro, 1993; Wu et al., 1995) were used for this study. All flies were reared on cornmeal medium on 16:8 light:dark cycle at 20°C. Prior to crossing, both strains were subjected to 10 generations of sibling pair matings to reduce genome-wide heterozygosity, and this was followed by three generations of population expansion to generate the quantity of flies needed for crosses. For each reciprocal cross performed, 10 vials were set up with 3 female and 3 male flies. Virgin female progeny were allowed to mate from the time of eclosion to 3 days posteclosion, then males and females were separated and females aged to 7–10 days post eclosion. All flies were collected during the same time of day to minimize the effects of circadian rhythm, and flies were snap-frozen in liquid N_2.

Library Preparation and Illumina Sequencing

Pools of 20 female flies were used for total RNA extraction with TRIzol reagent according to manufacturer instructions (Invitrogen). Illumina sequencing libraries were prepared (see Extended Experimental Procedures) as previously reported (McManus et al., 2010). Two lanes of paired-end (2X76 bp) Illumina GAIIx sequencing were performed.

Quantifying Total and Allele-Specific Expression from Sequencing Reads

We developed a bioinformatics pipeline to quantify gene expression from the Illumina sequencing output (Figure S1B; Extended Experimental Procedures). Briefly, we aligned each mate of the paired-end RNA-seq reads separately to the newly built *D. melanogaster* genomes (zhr and z30; Extended Experimental Procedures), keeping only those reads that aligned to one genomic location. Reads that did not map were trimmed by 13 bases and realigned in three iterations. Reads that did not align were then discarded. We then converted zhr and z30 genomic coordinates of aligned reads to sequenced *D. melanogaster* genomic coordinates using the liftOver utility from the UCSC Genome Browser (Kent et al., 2002). Aligned sequence reads were then filtered based on their alignment to a previously identified set of overlap filtered constitutively expressed exons within the *D. melanogaster* genome (McManus et al., 2010) using the intersectBed module of BedTools (Quinlan and Hall, 2010) (Version 2.12.0).

Remaining sequencing reads that aligned to only one of the two line-specific genomes were used for quantification of allele-specific gene expression. Down-sampling followed by rounding to the nearest integer was used to account for differences in overall sequencing output between MZ and ZM and differences in mappability between zhr and z30 alleles. For each gene, allele-specific expression levels are reported (Table S2). To reduce the effect of sampling error (Fontanillas et al., 2010; McManus et al., 2010), we analyzed only genes that had more than 20 allele-specific reads (allele 1 + allele 2 $\geq$ 20) in both ZM and MZ. To test for unequal allelic expression between ZM and MZ, we performed Fisher's exact tests using zhr and z30 allelic counts. Due to the multitude of tests performed, a false discovery rate (FDR) significance threshold of 5% was used to determine significance (Benjamini and Hochberg, 1995). Statistical analyses were performed in R (version 2.12.2, CRAN).

FPKM values reflecting total expression levels for individual genes were calculated by dividing the total number of paired-end reads mapped to a gene (including reads that were and were not informative for allele-specific expression) by the length of the sequence representing that gene in kilobases and then dividing this value by the number of millions of mapped reads from that sample.

Sliding-Window Analyses with Monte Carlo Sampling and Approximate Permutation Tests

Genomic clustering of putatively imprinted genes was analyzed using a sliding-window approach where we divided the genome into 11,726 overlapping 500 kb windows and moved stepwise, offsetting by 10 kb with each step. For each window, we counted the number of total genes and PIGs within each region. To test whether the observed clustering was significant, we used a Monte Carlo sampling approach to approximate the null distribution of imprinted genes randomly scattered along the genome. A Monte Carlo sampling approach was used to approximate the null distribution, because the number of permutations required for an exact test in this case was exceedingly large

(7.8×10^{261}). From the total set of 7,206 genes, we randomly sampled 119 genes without replacement, assigned them imprinting status, and aggregated new imprinting counts for each window. This was done 10,000 times, resulting in an approximate null distribution of the number of imprinted genes expected by chance in each window. To calculate an approximate p-value for each window, we summed the number of occurrences where the permuted value exceeded the observed value. Due to the multitude of tests performed, an FDR-corrected significance threshold of 5% was used to determine significance ($q < 0.05$). Significant windows were collapsed to four regions based on overlap (Figure 2).

Pyrosequencing

To evaluate the accuracy of allelic expression measurements derived from our RNA-seq data and analysis, new cDNA pools were synthesized from the same RNA samples used for Illumina sequencing and used for pyrosequencing. cDNA was synthesized from total RNA using T(18)VN primers and Superscript II (Invitrogen) according to manufacturer recommendations. Both cDNA and gDNA were analyzed using pyrosequencing. For each gene assayed, PCR was performed in triplicate on both the cDNA and gDNA samples (separately) and followed by pyrosequencing (QIAGEN). The genomic DNA was extracted from an independent pool of F_1 flies and was used to normalize cDNA measurements (Wittkopp, 2011). Log_2-transformed cDNA allelic expression ratios from Illumina and pyrosequencing were compared after normalization using type 2 regressions in R.

To investigate allelic expression within a cluster of genes on chromosome 3R, we constructed four new replicate pools of 20 individuals each for both ZM and MZ samples and coextracted RNA and gDNA from a single tissue homogenate of each pool of flies using the Promega SV total RNA extraction system with modified protocol (Wittkopp, 2011). cDNA was made from total RNA as above, and both gDNA and cDNA were used for PCR followed by pyrosequencing. To account for differences in gDNA allelic abundance among replicate pools of flies, the log_2 allelic expression ratio for gDNA from a particular pool was subtracted from the log_2 allelic expression ratios for cDNA samples derived from the same pool of flies (Wittkopp et al., 2004, 2006, 2008; Wittkopp, 2011).

The four biological replicates were used to investigate variation in allelic expression for a set of randomly chosen genes, and this was compared to a set of putatively imprinted genes. The standard error for the log_2 allelic expression ratio was calculated for each assay-sample combination for the randomly chosen genes and nonclustered PIGs, and these two sets were compared using a Wilcoxon rank-sum test in R.

ACCESSION NUMBERS

The sequencing data from this study have been submitted to the National Center for Biotechnology Information Sequence Read Archive under accession number SRA052065.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2012.06.013.

LICENSING INFORMATION

ACKNOWLEDGMENTS

We would like to thank all members of the Wittkopp lab, especially Brian Metzger and Bing Yang, for helpful comments on experimental design, data analysis, and manuscript presentation, Sebastian Zoellner for computational resources and Chung-I Wu for the z30 fly line. This work was supported by a National Science Foundation grant to P.J.W (MCB-1021398) and a National Institutes of Health grant to B.R.G. (5R01GM095296), as well as a National Research Service Award from the National Institutes of Health to J.C.D. (1F32GM089009-02). P.J.W. is a Research Fellow of the Alfred P. Sloan Foundation.

Received: February 16, 2012
Revised: April 27, 2012
Accepted: June 12, 2012
Published online: July 19, 2012

REFERENCES

Ahmadian, A., Gharizadeh, B., Gustafsson, A.C., Sterky, F., Nyrén, P., Uhlén, M., and Lundeberg, J. (2000). Single-nucleotide polymorphism analysis by pyrosequencing. Anal. Biochem. *280*, 103–110.

Anaka, M., Lynn, A., McGinn, P., and Lloyd, V.K. (2009). Genomic imprinting in *Drosophila* has properties of both mammalian and insect imprinting. Dev. Genes Evol. *219*, 59–66.

Babak, T., Deveale, B., Armour, C., Raymond, C., Cleary, M.A., van der Kooy, D., Johnson, J.M., and Lim, L.P. (2008). Global survey of genomic imprinting by transcriptome sequencing. Curr. Biol. *18*, 1735–1741.

Begun, D.J., and Aquadro, C.F. (1993). African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature *365*, 548–550.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. *57*, 289–300.

Buiting, K., Saitoh, S., Gross, S., Dittrich, B., Schwartz, S., Nicholls, R.D., and Horsthemke, B. (1995). Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. Nat. Genet. *9*, 395–400.

Caspary, T., Cleary, M.A., Baker, C.C., Guan, X.J., and Tilghman, S.M. (1998). Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. Mol. Cell. Biol. *18*, 3466–3474.

Crouse, H.V. (1960). The controlling element in sex chromosome behavior in *Sciara*. Genetics *45*, 1429–1443.

Degner, J.F., Marioni, J.C., Pai, A.A., Pickrell, J.K., Nkadori, E., Gilad, Y., and Pritchard, J.K. (2009). Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. Bioinformatics *25*, 3207–3212.

DeVeale, B., van der Kooy, D., and Babak, T. (2012). Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. PLoS Genet. *8*, e1002600.

Emerson, J.J., Hsieh, L., Sung, H., Wang, T., Huang, C., Lu, H.H., Lu, M.J., Wu, S., and Li, W. (2010). Natural selection on cis and trans regulation in yeasts. Genome Res. *20*, 826–836.

Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. Nat. Rev. Genet. *12*, 565–575.

Ferree, P.M., and Barbash, D.A. (2009). Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. PLoS Biol. *7*, e1000234.

Fontanillas, P., Landry, C.R., Wittkopp, P.J., Russ, C., Gruber, J.D., Nusbaum, C., and Hartl, D.L. (2010). Key considerations for measuring allelic expression on a genomic scale using high-throughput sequencing. Mol. Ecol. *19* (*Suppl 1*), 212–227.

Fuyama, Y. (1984). Gynogenesis in *Drosophila melanogaster*. Jpn. J. Genet. *59*, 91–96.

Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. PLoS ONE *6*, e23687.

Gibson, G., Riley-Berger, R., Harshman, L.G., Kopp, A., Vacha, S., Nuzhdin, S.V., and Wayne, M. (2004). Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. Genetics *167*, 1791–1799.

Golic, K.G., Golic, M.M., and Pimpinelli, S. (1998). Imprinted control of gene activity in *Drosophila*. Curr. Biol. *8*, 1273–1276.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., et al. (2011). The developmental transcriptome of *Drosophila melanogaster*. Nature *471*, 473–479.

Graze, R.M., Novelo, L.L., Amin, V., Fear, J.M., Casella, G., Nuzhdin, S.V., and McIntyre, L.M. (2012). Allelic imbalance in *Drosophila* hybrid heads: exons, isoforms, and evolution. Mol. Biol. Evol. *29*, 1521–1532. Published online January 7, 2012. 10.1093/molbev/msr2318.

Gregg, C., Zhang, J., Butler, J.E., Haig, D., and Dulac, C. (2010a). Sex-specific parent-of-origin allelic expression in the mouse brain. Science *329*, 682–685.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010b). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science *329*, 643–648.

Haller, B.S., and Woodruff, R.C. (2000). Varied expression of a Y-linked P[w+] insert due to imprinting in *Drosophila melanogaster*. Genome *43*, 285–292.

Hollocher, H., Ting, C.T., Wu, M.L., and Wu, C.I. (1997). Incipient speciation by sexual isolation in *Drosophila melanogaster*: extensive genetic divergence without reinforcement. Genetics *147*, 1191–1201.

Hsieh, T.F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., and Fischer, R.L. (2011). Regulation of imprinted gene expression in *Arabidopsis* endosperm. Proc. Natl. Acad. Sci. USA *108*, 1755–1762.

Joanis, V., and Lloyd, V.K. (2002). Genomic imprinting in *Drosophila* is maintained by the products of Suppressor of variegation and trithorax group, but not Polycomb group, genes. Mol. Genet. Genomics *268*, 103–112.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Komma, D.J., and Endow, S.A. (1995). Haploidy and androgenesis in *Drosophila*. Proc. Natl. Acad. Sci. USA *92*, 11884–11888.

Landry, C.R., Wittkopp, P.J., Taubes, C.H., Ranz, J.M., Clark, A.G., and Hartl, D.L. (2005). Compensatory *cis-trans* evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. Genetics *171*, 1813–1822.

Lewis, A., Mitsuya, K., Umlauf, D., Smith, P., Dean, W., Walter, J., Higgins, M., Feil, R., and Reik, W. (2004). Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. Nat. Genet. *36*, 1291–1295.

Lloyd, V.K., Sinclair, D.A., and Grigliatti, T.A. (1999). Genomic imprinting and position-effect variegation in *Drosophila melanogaster*. Genetics *151*, 1503–1516.

Lopes, S., Lewis, A., Hajkova, P., Dean, W., Oswald, J., Forné, T., Murrell, A., Constância, M., Bartolomei, M., Walter, J., and Reik, W. (2003). Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions. Hum. Mol. Genet. *12*, 295–305.

Luo, M., Taylor, J.M., Spriggs, A., Zhang, H., Wu, X., Russell, S., Singh, M., and Koltunow, A. (2011). A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. PLoS Genet. *7*, e1002125.

MacDonald, W.A., Menon, D., Bartlett, N.J., Sperry, G.E., Rasheva, V., Meller, V., and Lloyd, V.K. (2010). The *Drosophila* homolog of the mammalian imprint regulator, CTCF, maintains the maternal genomic imprint in *Drosophila melanogaster*. BMC Biol. *8*, 105.

Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S., Han, Y., Magwire, M.M., Cridland, J.M., et al. (2012). The *Drosophila melanogaster* Genetic Reference Panel. Nature *482*, 173–178.

Maggert, K.A., and Golic, K.G. (2002). The Y chromosome of *Drosophila melanogaster* exhibits chromosome-wide imprinting. Genetics *162*, 1245–1258.

Mancini-Dinardo, D., Steele, S.J.S., Levorse, J.M., Ingram, R.S., and Tilghman, S.M. (2006). Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev. *20*, 1268–1282.

McGrath, J., and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell *37*, 179–183.

McManus, C.J., Coolon, J.D., Duff, M.O., Eipper-Mains, J., Graveley, B.R., and Wittkopp, P.J. (2010). Regulatory divergence in *Drosophila* revealed by mRNA-seq. Genome Res. *20*, 816–825.

Menon, D.U., and Meller, V.H. (2009). Imprinting of the Y chromosome influences dosage compensation in *roX1 roX2 Drosophila melanogaster*. Genetics *183*, 811–820.

Menon, D.U., and Meller, V.H. (2010). Germ line imprinting in *Drosophila*: epigenetics in search of function. Fly (Austin) *4*, 48–52.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Sawamura, K., Yamamoto, M.T., and Watanabe, T.K. (1993). Hybrid lethal systems in the *Drosophila melanogaster* species complex. II. The Zygotic hybrid rescue (Zhr) gene of *D. melanogaster*. Genetics *133*, 307–313.

Shibata, Y., Kumar, P., Layer, R., Willcox, S., Gagan, J.R., Griffith, J.D., and Dutta, A. (2012). Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. Science *336*, 82–86.

Surani, M.A.H., Barton, S.C., and Norris, M.L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature *308*, 548–550.

Wang, X., Sun, Q., McGrath, S.D., Mardis, E.R., Soloway, P.D., and Clark, A.G. (2008). Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. PLoS ONE *3*, e3839.

Wang, X., Soloway, P.D., and Clark, A.G. (2011). A survey for novel imprinted genes in the mouse placenta by mRNA-seq. Genetics *189*, 109–122.

Waters, A.J., Makarevitch, I., Eichten, S.R., Swanson-Wagner, R.A., Yeh, C.T., Xu, W., Schnable, P.S., Vaughn, M.W., Gehring, M., and Springer, N.M. (2011). Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. Plant Cell *23*, 4221–4233.

Weksberg, R., Shen, D.R., Fei, Y.L., Song, Q.L., and Squire, J. (1993). Disruption of insulin-like growth factor 2 imprinting in Beckwith-Wiedemann syndrome. Nat. Genet. *5*, 143–150.

Wittkopp, P.J. (2011). Using pyrosequencing to measure allele-specific mRNA abundance and infer the effects of *cis*- and *trans*-regulatory differences. In Molecular Methods for Evolutionary Genetics, V. Orgogozo and M. Rockman, eds. (New York: Humana Press), pp. 297–317.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2004). Evolutionary changes in *cis* and *trans* gene regulation. Nature *430*, 85–88.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2006). Parent-of-origin effects on mRNA expression in *Drosophila melanogaster* not caused by genomic imprinting. Genetics *173*, 1817–1821.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2008). Regulatory changes underlying expression differences within and between *Drosophila* species. Nat. Genet. *40*, 346–350.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T., Spillane, C., Nordborg, M., Rehmsmeier, M., and Köhler, C. (2011). High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. PLoS Genet. *7*, e1002126.

Wood, A.J., and Oakey, R.J. (2006). Genomic imprinting in mammals: emerging themes and established theories. PLoS Genet. *2*, e147.

Wu, C.I., Hollocher, H., Begun, D.J., Aquadro, C.F., Xu, Y., and Wu, M.L. (1995). Sexual isolation in *Drosophila melanogaster*: a possible case of incipient speciation. Proc. Natl. Acad. Sci. USA *92*, 2519–2523.

Cell PRESS
Open ACCESS

EXTENDED EXPERIMENTAL PROCEDURES

Resequencing of zhr and z30 Genomes and Genome Assembly

Genomic DNA sequence reads were aligned to the *D. melanogaster* genome assembly (dm3; Adams et al., 2000; Celniker et al., 2002) using BWA (Li and Durbin, 2010; version 0.5.6). Each read was aligned separately using default parameters, and sam format files were generated using the bwa sampe command. For zhr, an additional sam file was prepared from single-read Illumina data. Alignment files were converted to bam format and vcf files describing snps and indels were created using the samtools package (Li et al., 2009; version 0.1.7a; modules view, sort, and pileup). SNP and indel calls were filtered using the samtools.pl varFilter command (as described at http://samtools.sourceforge.net/cns0.shtml) to retain SNPs and indels with PHRED scale quality scores of 20 or higher. At some positions, SAMtools identified heterozygous sites. This creates a complication for comparative RNA-seq, as the heterozygous genotype of one strain can partially overlap with the other strain. For example, if resequencing identified an "R" (either A or G) base at a coordinate in zhr and a "G" in z30, RNA-seq reads originating from the z30 could be mapped to both strains, while "A" containing reads from zhr would be strain specific. In order to avoid using regions of partial overlap in allele-specific RNA-seq assignments, we changed both SNP calls to the most ambiguous genotype possible using a custom perl script (snp_compare_filter.pl), effectively making these sites uninformative for allele assignment.

Strain-specific genome sequences were produced using a custom Perl script (snp_adder.pl). This script sequentially rewrites the *D. melanogaster* genome with corrected SNP calls and indels. The positions of insertions and deletions were recorded in custom liftover chain files during the rewriting process. These chain files allow the conversion of genomic features between strain and reference genomes using the UCSC genome browser liftover tool (http://genome.ucsc.edu; Kent et al., 2002). Heterozygous indel sites (insertion in one allele in one strain) were tracked in separate genome files (mixed indel 1 and mixed indel 2). The genomes and chain files are available upon request.

Library Preparation

cDNA libraries were prepared as in McManus et al. (2010). Briefly, 10 μg of total RNA from each sample was treated with DNase 1 (Invitrogen) followed by poly(A)$^+$ selection using Dynal magnetic beads (Invitrogen) following manufacturer recommendations. Poly(A)$^+$ RNA was then fragmented using RNA fragmentation reagent (Ambion) before cDNA synthesis. Double-stranded cDNA was primed using random hexamers and Superscript II reverse transcriptase (Invitrogen). cDNA was run on a 2% agarose gel and the region corresponding to ~300bp fragments was gel extracted. This size-selected double-stranded cDNA was used in the Paired-End Genomic DNA Library Preparation Kit (Illumina) according to manufacturer's recommendations. Genomic DNA libraries were prepared from pools of 20 female flies of each strain (zhr and z30) and genomic DNA was extracted using the DNeasy Blood & Tissue Kit (QIAGEN). 10 μg of gDNA was used to make gDNA sequencing libraries using the Paired-End Genomic DNA Library Preparation Kit (Illumina) according to manufacturer's recommendations. The cDNA libraries (ZM and MZ) as well as the gDNA libraries (zhr and z30) were subjected to paired-end sequencing on an Illumina Genome Analyzer IIx on one lane each for 76 cycles per read. The zhr gDNA sample was also sequenced from a single end on additional lane for 76 cycles per read. Images were analyzed using the Firecrest and Bustard modules to generate sequence and quality scores for each read.

Quantifying Allele-Specific Expression from Sequencing Reads

To quantify gene expression from the Illumina sequencing output we aligned each mate of the paired-end RNA-seq reads separately to the newly built *D. melanogaster* genomes (zhr and z30) using the MOSAIK aligner (version 1.0.1384, http://bioinformatics.bc.edu/marthlab/Mosaik). We used the following command line options for the alignment: –hs 13 –mm 0 –p24 –mph 100 –act 20. The 13 base hash size (–hs 13) option allowed >99% of ambiguous base containing regions to be seeded for alignment by MOSAIK. Only uniquely aligning reads with no mismatches were retained for analysis. After the initial 76 bp reads were aligned to both reference genomes, those reads that did not map to either were trimmed 13 bases from the 3′ end using a custom Perl script (fastq_trimmer.pl) and again aligned with MOSAIK. This was repeated three times (sequence lengths 76bp, 63bp, 50bp, 37bp). Any sequences that did not uniquely align after the final iteration were discarded.

Using the chain files created in the genome assembly process, we converted the respective genome coordinates (in zhr or z30 space) to the sequenced dm3 coordinates using the liftOver utility from the UCSC Genome Browser (Kent et al., 2002) (http://genome.ucsc.edu) and a custom Perl script (convert.pl). Sequence reads were then filtered based on their alignment to a previously identified set of constitutively expressed exons within the *D. melanogaster* genome (McManus et al., 2010) using intersectBed module of BedTools, with those reads not aligning to these regions discarded. Additionally, regions in the constitutive exon set found to overlap were removed using intersectBed module of BedTools and custom scripts. Constitutively expressed exon filtering was performed to reduce biases associated with isoform specific differences. The filtered set of sequencing reads was used for quantification of allele-specific gene expression. Reads were assigned to the zhr or z30 allele based on reported alignments using a custom Perl script (classify.pl). Because each paired-end read represents a single transcript, we only incremented gene counts once for each paired-end read (or once if only one end of the read mapped). For many genes, the number of reads aligning and contributing to quantification of gene expression exceeded the number of mappable positions, which means that identical sequencing reads were identified and included in our final quantification to avoid imparting maximum expression level thresholds to genes based on their length.

To correct for mappability differences between the two genomes that could lead to biases, we determined the total number of informative reads that aligned allele-specifically to the zhr and the z30 genomes for all genes in each F_1 hybrid sample (ZM and MZ) with the expectation of equal representation. Because the zhr alleles were slightly more abundant across the whole genome in both MZ and ZM, we down-sampled the zhr allelic counts globally by multiplying by 0.9706 in ZM and by 0.9736 in MZ followed by rounding to the nearest integer. To make comparisons between the reciprocal crosses we corrected for differences in sequencing depth between the two sequencing efforts. The ZM library had 31,432,754 reads and the MZ library had 31,439,998 reads. To correct for this minor difference, we multiplied the MZ counts by 0.9997 followed by rounding to generate integer read counts. For each gene, allele-specific expression levels are reported as the number of sequences that map to either the zhr or the z30 allele (Table S2) with no correction for gene length because all comparisons were made between alleles of equal size in the two strains. Because genes with low counts are more likely to be influenced by sampling error, we removed all genes from analyses that had less that 20 allele-specific reads used for expression quantification, retaining those that satisfy (allele 1 + allele 2 $\geq$ 20) for statistical analysis.

We performed Fisher's exact tests (FETs) using allelic expression counts (zhr and z30) from MZ and ZM to test for unequal allelic expression between progeny from reciprocal crosses. To correct for the multiple comparisons made (FETs), we used a false discovery rate of 5% (Benjamini and Hochberg, 1995). Statistical analyses were performed in R (version 2.12.2, CRAN).

SUPPLEMENTAL REFERENCES

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. (2000). The genome sequence of *Drosophila melanogaster.* Science *287*, 2185–2195.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Series B Stat. Methodol. *57*, 289–300.

Celniker SE, Wheeler DA, Kronmiller B, et al. (2002) Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biol* **3**, RESEARCH0079.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Li, H., Handsaker, B., Wysoker, A., et al. (2009). The Sequence alignment/map (SAM) format and SAMtools. Bioinfomratics *25*, 2078–2079.

McManus, C.J., Coolon, J.D., Duff, M.O., Eipper-Mains, J., Graveley, B.R., and Wittkopp, P.J. (2010). Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res. *20*, 816–825.

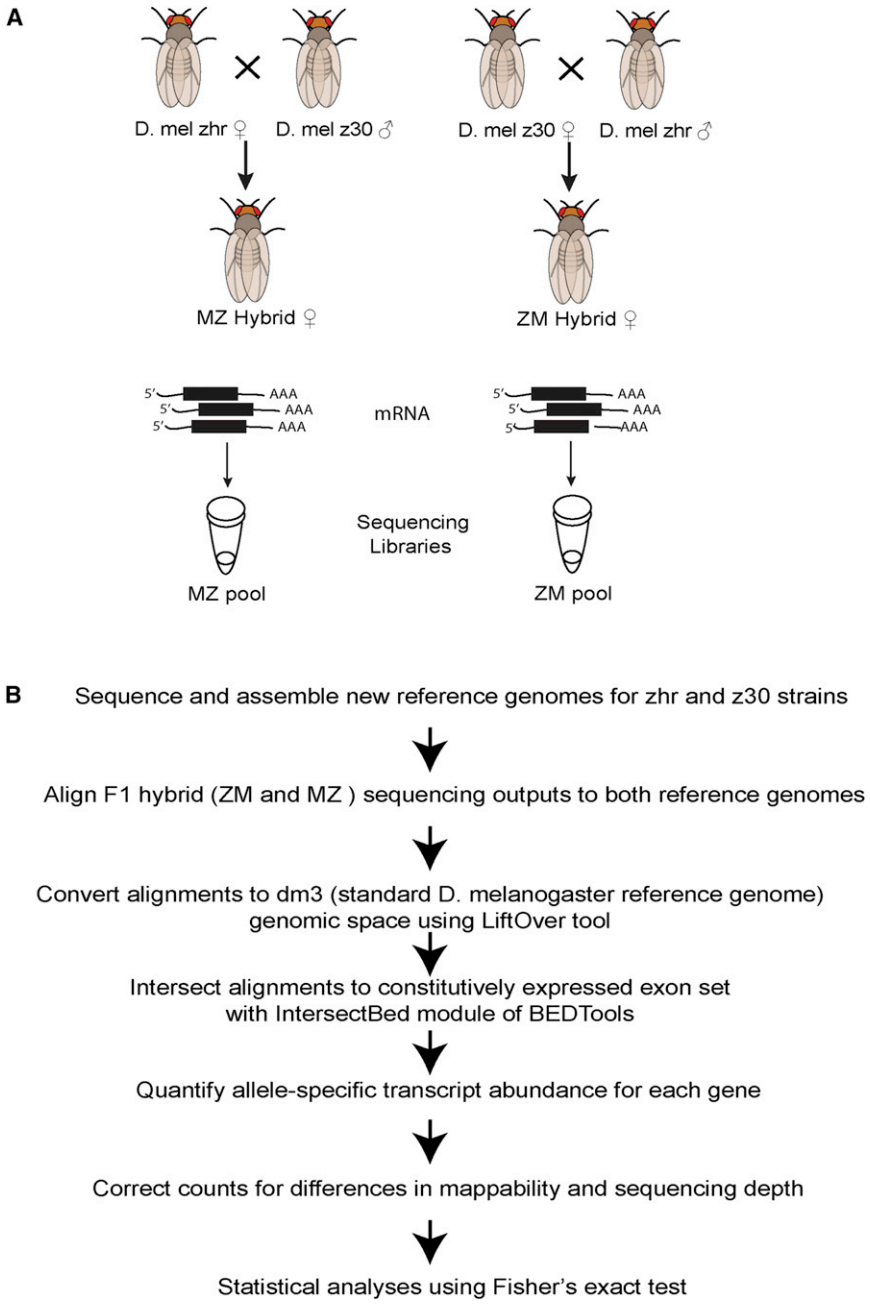

Figure S1. Experimental Method for Investigating Imprinting, Related to Figures 1–4

(A) Reciprocal crosses between M-type (zhr) and Z-type (z30) strains of *D. melanogaster* were performed to generate MZ (zhr females X z30 males) and ZM (z30 females X zhr males) F1 progeny. Pools of 20 female progeny were used for isolation of RNA and DNA (see Experimental Procedures).

(B) A flowchart of the steps used to transform Illumina sequencing reads into allele-specific gene-expression counts for MZ and ZM cDNA libraries (see Extended Experimental Procedures) is shown.

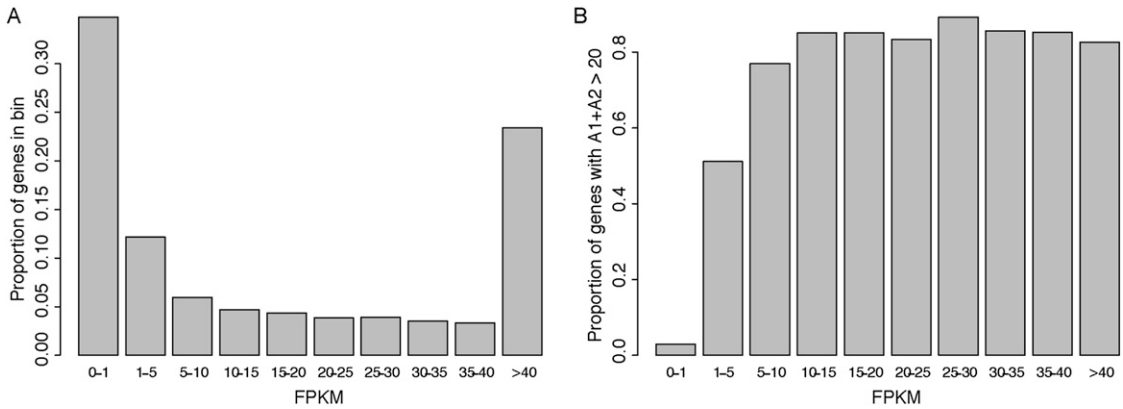

Figure S2. Coverage of Genes Tested for Imprinting, Related to Figure 1

(A) Genes were binned based on their total expression level in fragments per kilobase per million mapped reads. Total expression is plotted on the *x* axis and the proportion of genes in each bin is indicated on the *y* axis.

(B) Total expression (FPKM) is plotted on the *x* axis, and the proportion of genes in each total expression bin tested for imprinting is shown on the *y* axis.

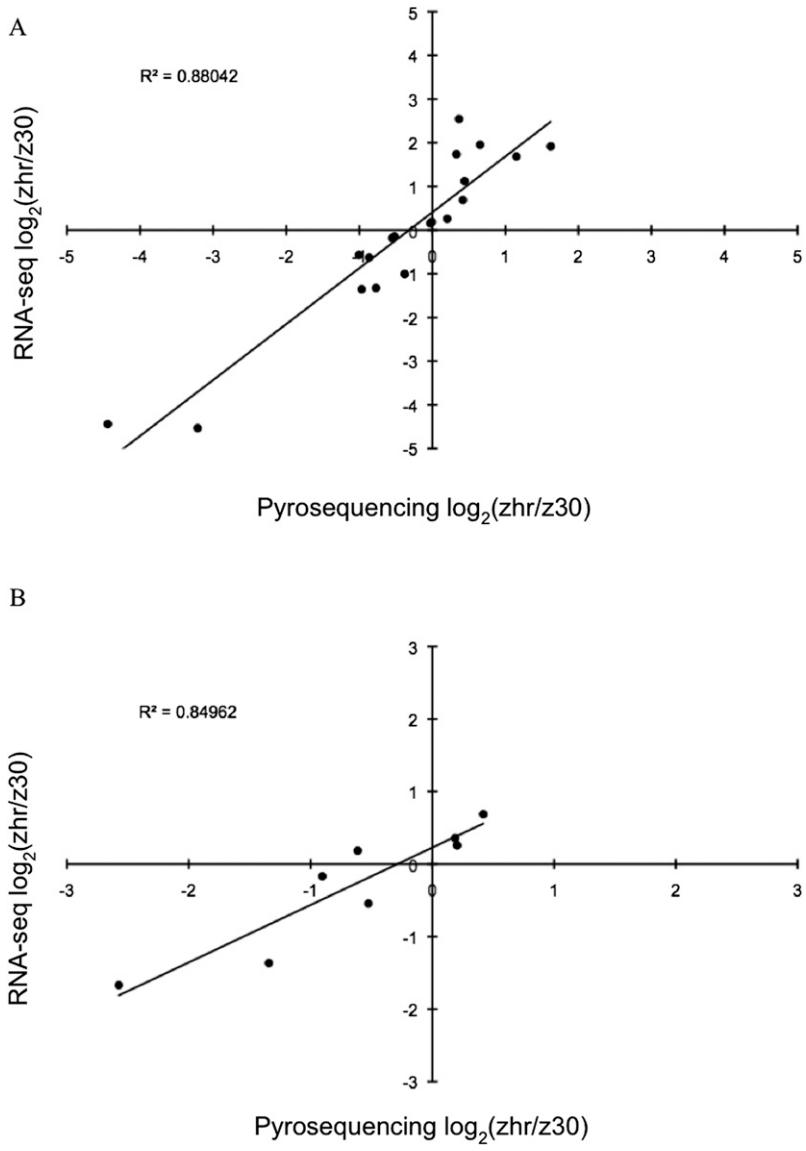

Figure S3. Pyrosequencing Validation of RNA-seq Data, Related to Figure 1

(A and B) Log_2 transformed allelic expression ratios (zhr/z30) from pyrosequencing on the *x* axis and log_2(zhr/z30) allelic expression ratio from RNA-seq on the *y* axis. Two points represent each gene, one for allelic expression measures from ZM and one for those from MZ. Data from randomly selected genes (A) and from PIGs (B) are shown. Type 2 regressions were performed, and correlation coefficients (R^2) for (A) and (B) are shown.

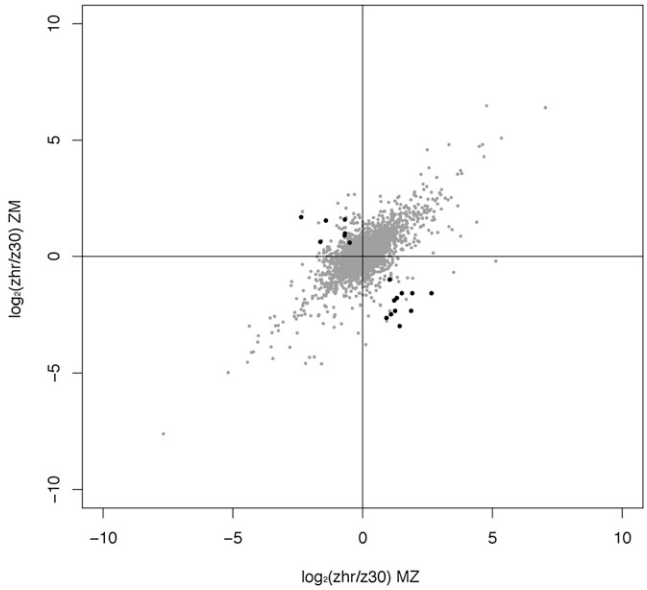

Figure S4. RNA-seq Data Plotted with Filtered PIGs Highlighted, Related to Figure 4

Log_2-transformed allelic expression ratios (zhr/z30) from MZ on the *x* axis and from ZM on the *y* axis. Each point represents one gene. Points are color-coded: black points indicate genes that were identified with significant allelic-expression differences between MZ and ZM and met a more strict definition of imprinting-like expression pattern where either the maternally or paternally inherited allele accounts for the majority of the expression of the imprinted gene; gray points represent all other genes quantified with RNA-seq.