Supplementary Material

Hamada, Park, Gordadze & Kuroda

This document contains additional information for the manuscript *Global regulation of X chromosomal genes by the MSL complex in Drosophila melanogaster*. The Affymetrix raw data (.cel files) and an Excel file containing the expression values, ratios, and other information are available at `http://chip.tch.harvard.edu/~ppark/KurodaLab`. All analysis was carried out using the statistical language R.

1 Data Processing

Affymetrix Drosophila Genome 2.0 Array with 18880 probe sets was used. The data were processed using GeneChip Operating Software (GCOS) 1.1 from Affymetrix, Inc. The .cel files for the probe level data were inspected visually and no obvious artifacts were observed. No saturations were detected on the high intensity signals. The average Pearson correlation coefficients, which are independent of linear normalization, among the biological triplicates were .989 for GFP controls and .980 for RNAi experiments. The overall data quality is excellent. Three independent biological experiments were performed on different days, for the total of six arrays.

Calculation of the fold changes

There are several ways to calculate the ratio of the transcripts between the RNAi and GFP samples for each gene. For example, one can take the average of the three RNAi arrays and the average of the three GFP arrays and divide one by the other. However, this is not optimal because each experiment was performed separately and there is a natural pairing between the RNAi and GFP arrays from the same experiment. Instead, it is better to compute the ratio from each experiment first and take median of the three ratios. This way, it also allows us to filter out the less reliable ratios by seeing how consistent the ratios are among the three replicates. Within each experiment, one can obtain the ratio simply by dividing the expression level in RNAi by that of GFP. However, the accuracy of the ratio can be further improved by using the probe level data instead of the probe set level data. (Affymetrix estimates the abundance of each transcript by averaging over multiple probes; 'probe set' level refers to the averaged values) By comparing each of the corresponding individual probes in the two arrays first and then averaging the ratios in a robust way, a more accurate estimate can be derived. As the sample size increases in a two-group comparison, the advantage gained by considering the probe level data diminishes, but for a comparison between two single arrays, this method is superior.

All ratios from each of the three experiments are plotted in Figure A. We take log (base 2) so that up-regulation and down-regulation are symmetric on the plot. 1 on the x-axis indicates two-fold up-regulation and -1 indicates two-fold down-regulation. We see that in all three experiments, the X chromosomal genes (red line) are down-regulated in the RNAi experiment compared to the autosomal genes (black line).

Filtering

Because many of the ratios in Figure A are derived from non-expressed genes or are the result of bad probes that give unreliable ratios, we applied two different techniques described below to filter out those cases. These two filtering steps significantly reduced the false positive rate for the genes identified as up- or down-regulated. The distributions after filtering are shown in Figure B, which

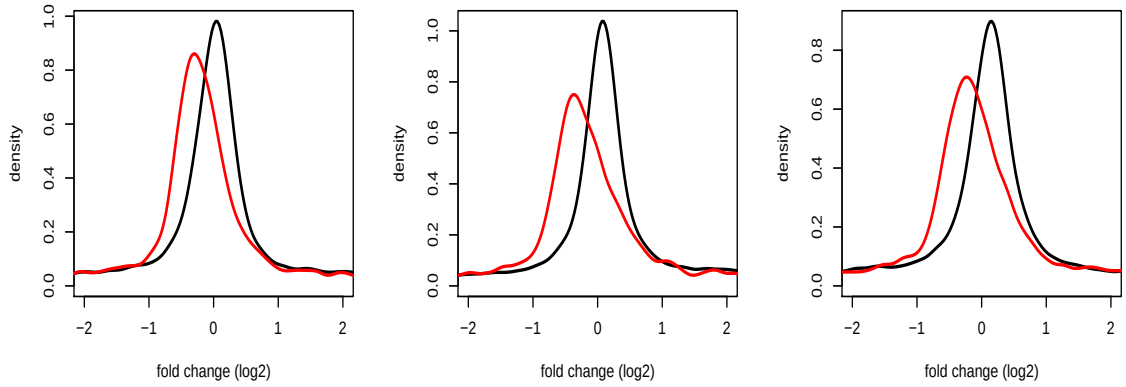

Figure A: Distribution of ratios from three independent experiments: the red and black lines in each experiment denote the probability density function for the log ratios of the genes on the X chromosomes and autosomes, respectively. All log-ratios (18369 genes that were mapped to their chromosomal location) without any filtering are plotted here.

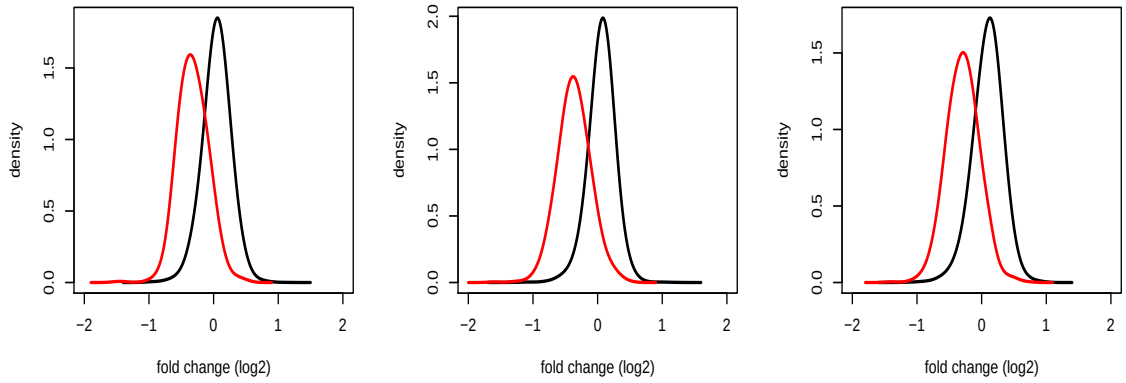

Figure B: The effect of filtering on the log ratios: as before, the red and black lines denote the probability density function for the log ratios of the genes on the X chromosomes and autosomes, respectively. The separation is sharper after removing non-expressed genes (based on Present/Absent calls) and those with highly variable ratios in the replicates. 5436 log ratios are plotted here.

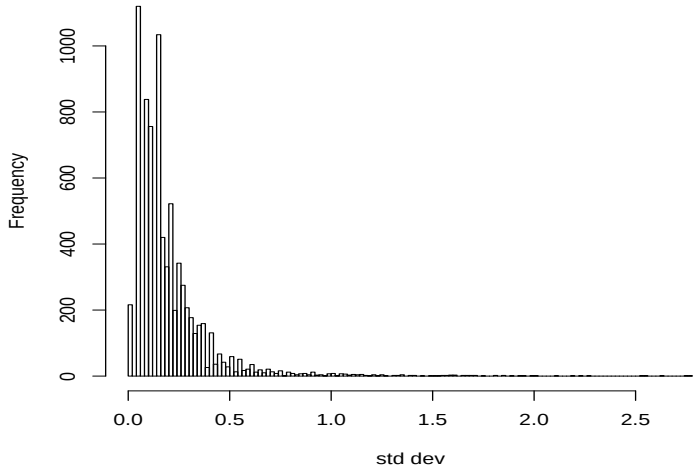

Figure C: Histogram of standard deviations for log ratios in the triplicates

indicates a clearer separation between the ratios of the X and the autosomal genes. (The median of the log ratios among the experiments are plotted in Figure 3A of the paper.)

A popular method of filtering is to set a threshold value and eliminate those genes whose expression levels are below the threshold in a given fraction of samples. An alternative is to use the Affymetrix Absent/Present/Marginal calls, which are assigned to each measurement based on the probe level data. (Those that are called Absent are generally non-expressed or low expressed genes) With 6 arrays in this study, we filtered out those genes that received the Absent call in more than 3 arrays. We picked 3 as the threshold here because we would like to keep a gene if it is present in at least one of the two conditions (either experiment or control), and this requires that there are at least 3 Present calls. This reduced the number of probe sets to 7923.

Among the genes that passed the filter, many still had highly variable ratios in the three replicates. For example, probe set `1631475_at` has log2 ratios .1, 4.9, and .1 (fold: 1.1,30,1.1); `1626270_s_at` has -2.2, .1, and 2.9 (fold: 0.28,1.1,7.5). Since any estimate based on these fluctuating ratios are likely to be unreliable, a second filtering step was implemented. We first calculated the standard deviation (in the log scale) for each gene, and remove those with high standard deviation. As can be seen in the histogram of the standard deviations (Figure C), there is a wide range of variability, though not a clear threshold for deciding on a 'high' value. We chose .25 in our analysis, as this left enough genes (5436) while being somewhat stringent. (If the log ratios are, e.g., 1, 1.25, 1.5, this has the standard deviation .25). Other thresholds also give similar results.

Other Algorithms for Computing Expression Values

We have also carried out the analysis using other algorithms for computing expression values. The accurate pre-processing of the Affymetrix raw data to derive the estimates of the expression values is a subject of its own and much effort has been dedicated to this topic in the recent years. The image file must be processed to account for spatial variation, correct background subtraction must be performed, probe level data for PM and MM must be calculated, and these values have to be properly combined in a robust way. The Affymetrix software GCOS combines the PM-MM values using a weight function to put less weight on the outlier values (See the GCOS manual). Other popular methods include model-based expression intensity, implemented in the dChip software, (Li and Wong, *PNAS* **80**:192, 2001) and Robust Multi-chip Analysis (Irizarry et al, *Nucleic Acids Research* **31**:e15, 2003) that does not use MM probes, which can often introduce noise especially

	Original	Percent	Filtered	Percent
2L	3189	17.4%	936	17.2%
2R	3479	18.9%	1097	20.2%
3L	3574	19.5%	1047	19.3%
3R	4651	25.3%	1363	25.1%
4	106	0.6%	41	0.8%
X	2971	16.2%	897	16.5%
U,2h,3h,Xh,Yh	399	2.2%	203	1.0%
Total	18369	100%	5436	100%

Table 1: Chromosomal location of the probe sets

for the low expressed genes. Normalization is also an issue that has been studied extensively. In general, a robust standardization at the probe set level, for example by a trimmed mean, appears to be sufficient in many cases. Saturation and nonlinear effects used to be a serious problem but they are no longer a problem in most datasets. We have explored the effect of these other methods for pre-processing the raw data, but nearly identical results are obtained.

The probability density functions in the figures were estimated using the binned kernel density estimate, as implemented in the KernSmooth package in the statistical language R. This is a standard smoothing technique in which the estimate at each point accounts for the value of the neighboring points, with a weight that decreases as the distance from the point increases. The Gaussian kernel is used here, which is a weight function that looks like a normal distribution. Just as in generating histograms, a larger bandwidth of the kernel will result in smooth estimates with few local features whereas a smaller bandwidth might result in spurious features.

2 Chromosomal Location

The Affymetrix probes were designed based on Flybase version 3.1. Of the 18769 probe sets on the array (after 183 control probes are removed), 18369 were found to align to a location on the genome (Affymetrix annotation, November 04). Table 1 shows the distribution of the probes on the chromosomes.

At different thresholds for fold changes, the same pattern is observed as in Figure 4 of the paper. For example, at 1.2 fold threshold, more genes are marked, as shown in Figure D (Figure 4 of the paper shows the result at 1.4 threshold). Because the resolution of the any file format is lower than that of the base pairs along the chromosomes, the banding pattern should be interpreted with some caution.

To determine at what point the difference between the X and the autosomes is most pronounced, the percentage of down-regulated genes is plotted at each threshold in Figure E. The solid line represents the X chromosome and the dotted line represents the autosomes. At every threshold, the percentage is much greater for the X chromosome. As mentioned in the paper, at 1.4 fold decrease, there are 261 out of 897 (29.1%) on the X and 56 out of 4484 (1.2%) on the automsomes, which translates to a 23-fold change in percentages. At a higher thresholdi, the number of down-regulated genes starts to decrease and so the percentages are not as reliable. This plot is equivalent to summing up the area under the curve to the left of a x-axis value in Figure 3A of the paper.

3 Normalization-independent analysis

To demonstrate the lack of dependence on normalization, we have evaluated the distribution of the p-values that measure reliable detection of transcripts *within* each array. (This part is done on the unfiltered list of 17970 genes that were mapped to 2L, 2R, 3L, 3R, 4, or X) The detection p-values are computed in the following way. For each probe pair, the intensity difference relative to its total intensity (PM-MM)/(PM+MM) is computed. The Wilcoxon Signed Rank test is then used to rank

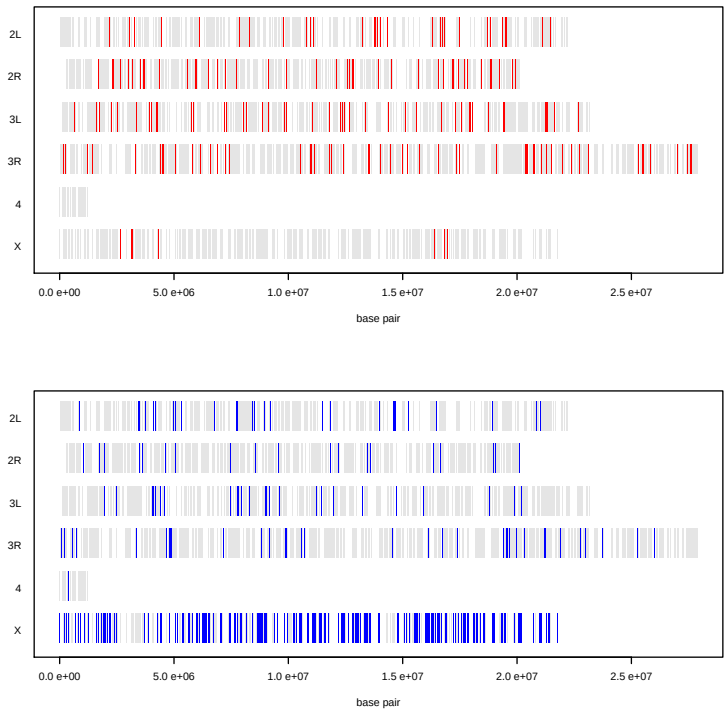

Figure D: Up- and down-regulated genes on the chromosomes: the top and bottom panel show the location of the genes with greater than 1.2 fold up- and down-regulation, respectively.

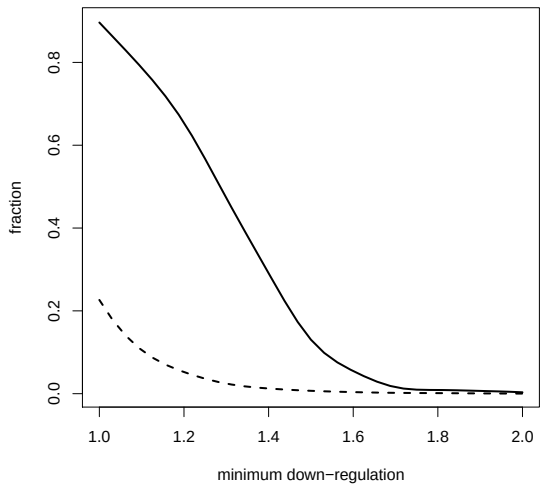

Figure E: The percentage of down-regulated genes on the X and the autosomes

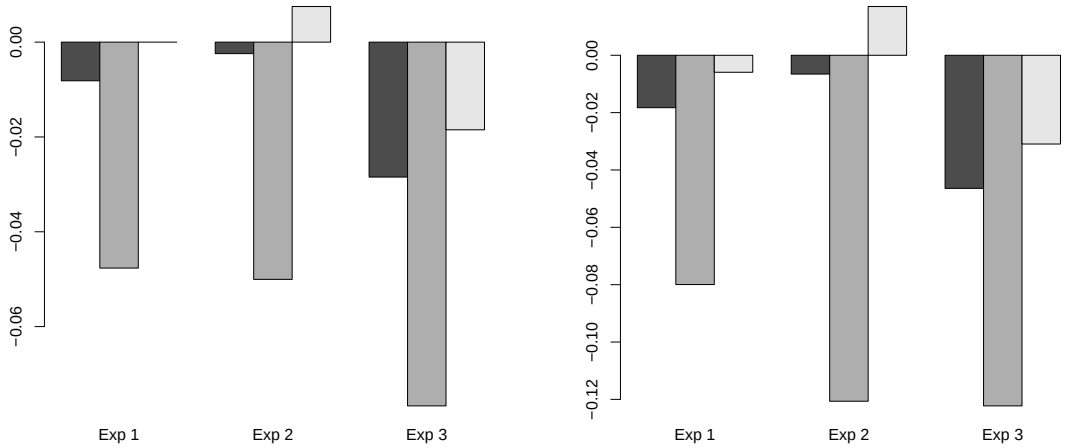

Figure F: The changes in the number of reliably detected genes in the RNAi experiment compared to the GFP control. In each experiment, the three bars correspond to all, X, and autosomal genes. On the left panel, .01 was used as the threshold for determining reliable detection. On the right panel, .001 was used. In all cases, there is a clear decrease in the detection of genes on the X.

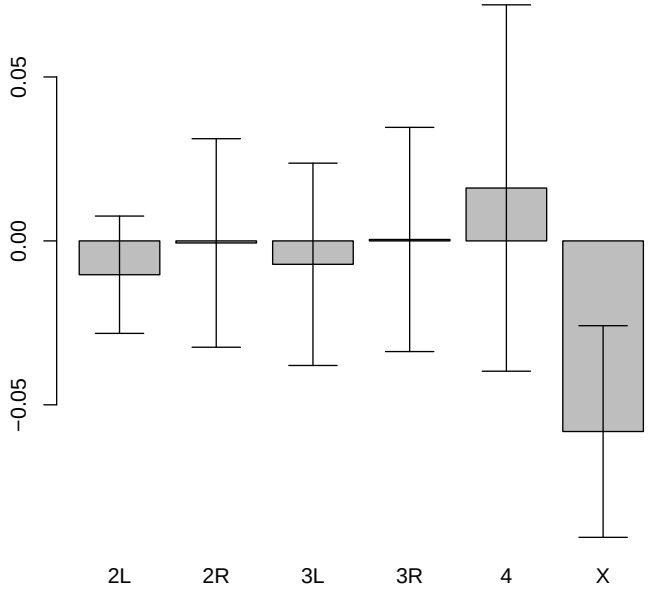

Figure G: The number of reliably detected genes for each chromosome. Again, there is a significant decrease in the X but not in other chromosomes.

the deviation of each value to a specified threshold (we used the default values) and to determine whether the pattern of deviations shows larger relative intensities of the PM probes. A low p-value means that the transcript was detected. The Affymetrix Present and Absent calls are computed in the same way but the calls themselves are based on a high threshold default, which makes it difficult to uncover the effect we would like to observe. Here, we work with the p-values directly.

In Figure 3B of the paper, we showed the results of the three experiments using the threshold p-value of .01. In Figure F, we show the results using the p-value of .001 on the right panel to show that the results are not sensitive to the threshold value (Figure 3B of the paper is reproduced on the left for comparison). Again, in each experiment, the three bars correspond to all, X, and autosomal genes. The magnitude of the change as indicated by y-axis increase for more stringent threshold, but the dominant feature in both cases is clearly the strong decrease in the number of reliably detected genes on the X. In Figure G, the results were divided by each chromosome, along with their error bars calculated from the three experiments.

In all cases, the pattern is the same: the number of reliably detected genes on the X chromosome goes down in the RNAi experiment compared to the GFP control, while it remains basically unchanged for the autosomes. No normalization was involved here because we did not directly compare the expression levels between arrays.