File S1

Automated analysis of genome-wide screen images

Methods: An ImageJ script was developed to generate counts of MSL1 foci and Hoescht-stained nuclei. Each image was assigned a MSL1 fragmentation score, equal to the total number of MSL1 foci divided by the total number of nuclei. The mean and standard deviation of all scores within each plate were calculated using InStat software. Each image was also assigned a statistically normalized score, comprising the difference between the image's fragmentation score and the respective plate mean score, divided by the plate's standard deviation. Images with normalized scores of at least one standard deviation above the mean were selected as hits. Selected positives were manually revisited for qualitative verification of supernumerary MSL1 phenotype.

Results and Discussion: Our primary analysis was based on qualitative scoring of screen images by eye, which likely biases our data toward dsRNAs with strong effects. We chose this method due to the variation among our images, including some images in which non-specific binding of the MSL antibody produced large bright artifacts, or in which MSL staining was slightly out of the plane of focus. For most images, this variation was easily compensated by the human eye, but was more difficult to analyze computationally at high stringency. However, we reasoned that our image data could be further explored using a computational approach at a lower stringency combined with a secondary verification by visual inspection, which could permit us to uncover additional dsRNAs that produce the supernumerary MSL phenotype. To accomplish this, we first developed an ImageJ script to generate counts of the average number of MSL1 foci per nucleus for each image analyzed. We next established a relaxed cutoff wherein any image with a count that surpassed one standard deviation above the average count from a given 384-well plate was selected as a potential supernumerary MSL1 hit (see Materials and Methods). Using this method, we screened approximately 90% of our original set of images, and scored 1,152 images as putative hits. Using manual inspection, we qualitatively confirmed the supernumerary MSL1 phenotype in 799 (69.4%) images. The remaining images were either out of focus, had high-intensity artifacts that caused the script to mistakenly select them as supernumerary MSL1 hits, or displayed what appeared to be normal MSL1 localization.

The 799 images scored as hits corresponded to a total of 374 annotated genes (Table S5). Although the set of genes identified by our computational screening expanded our hits considerably, it showed little overlap with those identified by our original qualitative analysis, with only three genes in common (*Myb*, *feo*, and *scra*). For many of the screen hits from our original analysis, it is likely that qualitative differences among the images, particularly the

brightness and size of staining bodies for both MSL1 foci and nuclei, prevented consistent identification and segmentation by our imaging script.

Although we did not see significant enrichment of GO terms among the 374 gene hits from our computational scoring, we did identify several hits with the GO term “cytokinesis”, including *Syx5*, *stmA*, *rok*, *sqh*, *shi*, and *Rho1* (in addition to *Myb*, *feo* and *scra* that were also among our SN hits), and the centromere protein *cid*. In addition, RNAi disruption of two other hits from our computation scoring, *CG30020* and *Nek2*, cause significant chromosomal missegregation during mitosis (R. D. and B. G. M., unpublished data), consistent with roles for these genes in cell cycle processes.