Supporting Information

Sheltzer et al. 10.1073/pnas.1209227109

SI Materials and Methods

Yeast Expression Data. Gene expression data from disomic and *cdc* mutant yeast strains were downloaded from (1). These data include gene expression values from yeast grown in batch culture, in which cells were allowed to reach logarithmic growth. Under these conditions, the yeast disomes display an ESR. This dataset also includes expression values from yeast grown under phosphate limitation in chemostats. Under this growth regime, the doubling time between euploid and disomic strains was equalized, thereby masking the ESR.

Gene expression data for strains derived via triploid meiosis were acquired as previously described and were deposited in the Gene Expression Omnibus (accession number GSE35853) (1). In strain A15, we detected extra copies of chromosome X (2), which were not present in the published karyotype. Gene expression data from 23 aneuploid strains from the deletion collection were downloaded from http://hugheslab.ccbr.utoronto.ca/supplementary-data/rii/. A few of these strains are believed to have developed aneuploidies because the deletion directly affects the spindle assembly checkpoint or chromosome segregation. Other strains gained chromosomes that carry a paralog of the deleted gene. However, in the majority of cases, the causes of aneuploidy in these strains are unknown (3). The set of genes that constitute the budding yeast ESR were downloaded from http://genome-www.stanford.edu/yeast_stress/. Growth rate–responsive genes in yeast were acquired from (4).

For analyses involving *S. cerevisiae*, ORFs annotated as "dubious" and ORFs that were detected in fewer than 10 disomic strains were removed from consideration. In addition, YAR015W and YBR115C were excluded, as they were the sites of marker integration in the euploid strain; and YDR342C and YDR343C were excluded, as they were amplified at the DNA level in aneuploid strains (1). Three wild-type vs. wild-type replicates were used as a control. The percent of the genome that was aneuploid was calculated as the sum of the number of ORFs present on a chromosome that was gained or lost, divided by the total number of ORFs in a euploid cell of that ploidy. The intensity of the stress response in yeast was calculated as follows: for all genes x that are upregulated in the ESR and all genes y that are downregulated in the ESR, the stress-response intensity (SRI) of a given strain was calculated as

$$SRI = \frac{\sum_{i=1}^{m} x_i + \sum_{j=1}^{n} -y_j}{m+n}$$

As the expression levels of most ESR genes change in response to growth rate, stress response intensities are not bimodal, but instead vary along a continuum. Strains were considered to exhibit an ESR if their SRI were positive and their SRI significantly differed ($P < 10^{-5}$, Student t test) from that of an isogenic wild-type strain.

Gene expression data from aneuploid strains of fission yeast were downloaded from GEO (accession number GSE8782). Strain C16, which contained one additional copy of 163 ORFs, and strain Ch16+S28, which contained two additional copies of the same 163 ORFs and one additional copy of 63 ORFs, were used for analysis. As these two strains contain very similar duplicated regions, they were averaged for subsequent comparisons. *S. pombe* genes involved in the fission yeast ESR were downloaded from http://www.bahlerlab.info/projects/stress/.

***A. thaliana* Expression Data.** Gene expression data for plants trisomic for chromosome 5 were downloaded from http://bioinf.boku.ac.at/pub/trisomy2008/. Log2-FC values were calculated between aneuploid and euploid samples for all genes not on chromosome 5. Gene expression data from stressed plants were downloaded from http://www.weigelworld.org/resources/microarray/AtGenExpress/. Log2-FC values were calculated between treated and mock-treated samples at the 24 h time point.

Mammalian Expression Data. Gene expression data from trisomic mouse cell lines were downloaded from (5). Probes were updated using Release 32 of the NetAffx probe annotations, and the set of probes classified as "expressed" were used in this study. Gene expression data from stressed mouse cells were downloaded from GEO (accession number GSE18320). Gene expression data from human aneuploidies were downloaded from GEO (accession numbers GSE6283, GSE1397, GSE16176, and GSE25634). For all mammalian aneuploidies, nonspecific probe sets and probe sets mapping to the X or Y chromosome were excluded from consideration. Probe sets that mapped to the same gene were collapsed by averaging. Within each dataset, log2-FC values were calculated between aneuploid and euploid samples.

Data Analysis. Gene expression data were analyzed in Excel, MATLAB, and Python using custom scripts. For all correlative studies, only gene expression values from euploid autosomes were considered, and multiple replicates of individual strains were averaged for comparisons. Aneuploid chromosomes were excluded because of the possibility that dosage compensation mechanisms decreased the transcription of a select number of genes present on extra chromosomes, which would introduce biases into subsequent analyses. Correlation values reported in the text represent the Pearson coefficient between two samples. Every comparison found to be significant via the Pearson method was also found to be significant when the Spearman coefficient was calculated instead.

GO term enrichment analysis was performed using GProfiler with a Benjamini–Hochberg-corrected P value of 0.05 and a maximum P value of 10^{-3} (6). Enrichments were performed against the relevant background gene set, e.g., against all genes not on an aneuploid chromosome or against the set of one-to-one orthologs between species. Consistent with our previous methodology (1, 7), differentially expressed genes were identified using a ±1.3-FC cutoff in yeast and plants. We used a ±1.5-FC cutoff in multiple samples for mammalian expression data, a similarly low-stringency threshold that allowed us to detect expression changes in datasets that contained variable numbers of replicates. As a secondary method, we used the rank products algorithm, an FC cutoff-independent protocol that is particularly useful in identifying differentially expressed genes in datasets with small sample sizes (8–10). The rank products algorithm was implemented in TM4 using a $P < 0.05$ significance threshold for yeast and plant data and a $P < 0.005$ threshold for mammalian data (11). P values reported in the text are from GO term analysis using an FC cutoff.

For budding yeast, mice, and humans, orthologous genes were identified using GProfiler (6). For *Arabidopsis* and fission yeast, orthologous genes were identified using InParanoid (12). All interspecies comparisons were between genes in which a one-to-one orthology relationship existed. Clustering was performed in Gene Cluster 3.0 (13) and visualized in Java TreeView (14). For clustering and visualization, the expression values of genes present on aneuploid chromosomes were replaced by the average value of that gene on euploid chromosomes in other strains.

Permutation tests to confirm the significance of transcriptional similarities were performed in Python. For each strain or species, blank cells were fixed in place. Blank cells usually resulted from the exclusion of an aneuploid chromosome, and these cells were locked so as to keep the number of relevant comparisons constant. Next, nonempty cells were randomly shuffled, and after each shuffle the relevant parameter was scored. *P* values reported for permutation tests represent maximum probabilities based on the number of permutations performed.

1. Torres EM, et al. (2007) Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* 317:916–924.
2. Sheltzer JM, et al. (2011) Aneuploidy drives genomic instability in yeast. *Science* 333: 1026–1030.
3. Hughes TR, et al. (2000) Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat Genet* 25:333–337.
4. Brauer MJ, et al. (2008) Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol Biol Cell* 19:352–367.
5. Williams BR, et al. (2008) Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science* 322:703–709.
6. Reimand J, Arak T, Vilo J (2011) g:Profiler—A web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res* 39:W307-15.
7. Torres EM, et al. (2010) Identification of aneuploidy-tolerating mutations. *Cell* 143: 71–83.
8. Breitling R, Armengaud P, Amtmann A, Herzyk P (2004) Rank products: A simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett* 573:83–92.
9. Jeffery IB, Higgins DG, Culhane AC (2006) Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC Bioinformatics* 7:359.
10. Hong F, Breitling R (2008) A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics* 24: 374–382.
11. Saeed AI, et al. (2003) TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* 34:374–378.
12. Ostlund G, et al. (2010) InParanoid 7: New algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res* 38:D196–D203.
13. de Hoon MJL, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* 20:1453–1454.
14. Saldanha AJ (2004) Java Treeview—Extensible visualization of microarray data. *Bioinformatics* 20:3246–3248.

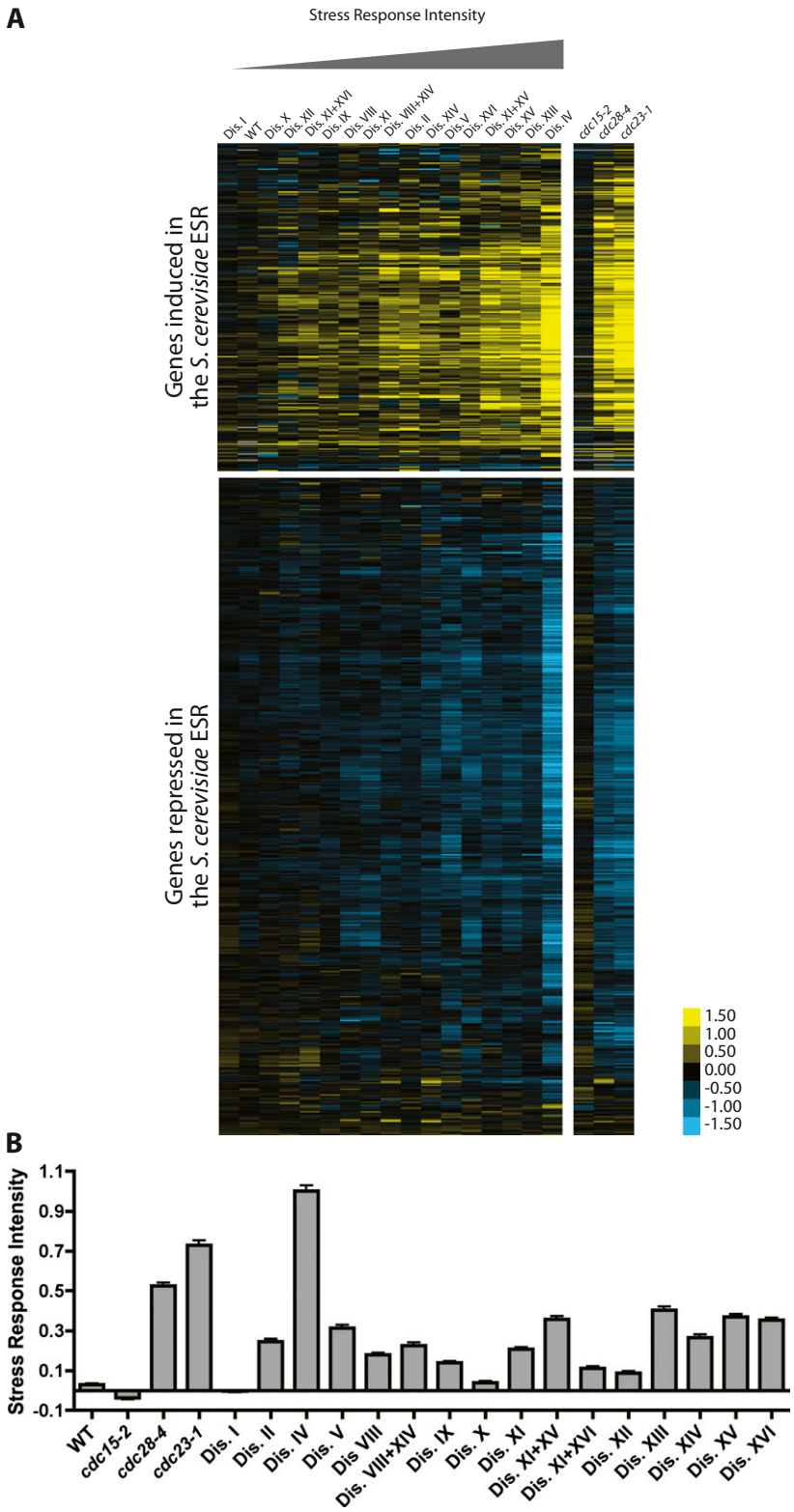

Fig. S1. The ESR in disomic strains of *S. cerevisiae*. (*A*) Strains were sorted according to their stress response intensity and genes annotated to the ESR were clustered. Genes that constitute the *S. cerevisiae* ESR were downloaded from http://genome-www.stanford.edu/yeast_stress/. The stress-response intensity in each disome was calculated by averaging the expression levels of genes upregulated in the ESR and then subtracting the average expression level of genes downregulated in the ESR. (*B*) The stress-response intensities of each disomic and *cdc* mutant strain are displayed.

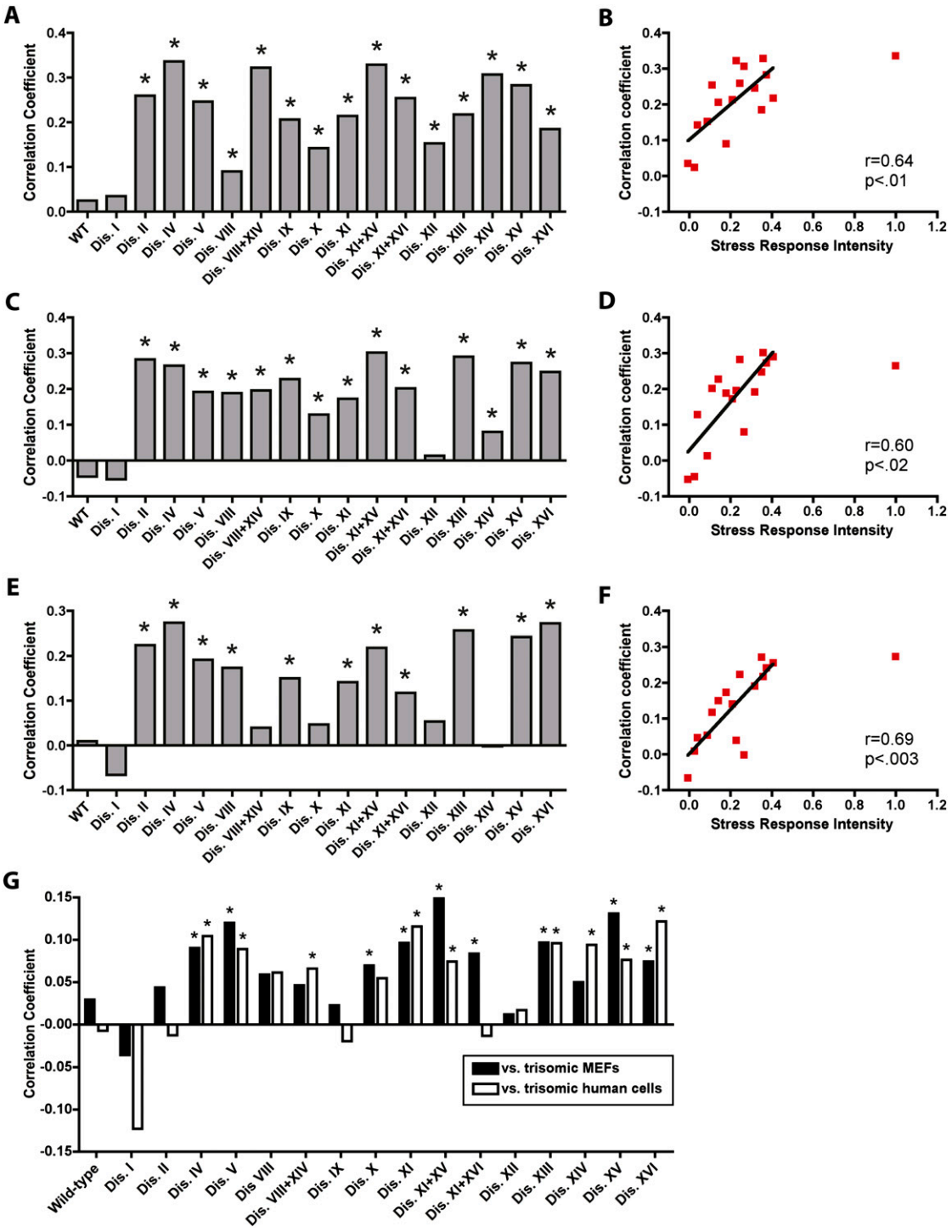

Fig. S2. Stress-related transcriptional similarities between aneuploid budding yeast, fission yeast, plants, and mammalian cells. The correlation coefficients between each disomic strain and (*A*) aneuploid strains from the yeast deletion collection, (*C*) aneuploid strains of *S. pombe*, (*E*) trisomic *A. thaliana*, and (*G*) and trisomic mammalian cells are displayed. The stress-response intensity of each disomic strain is plotted against its correlation coefficient with (*B*) aneuploid strains from the yeast deletion collection, (*D*) aneuploid strains of *S. pombe*, and (*F*) trisomic *A. thaliana*. We found that disomic strains that did not exhibit a stress response (e.g., disome I and XII) exhibited small or insignificant correlations with aneuploid cells in other species, while disomic strains that exhibited strong stress responses (e.g., disome IV and disome XIII) tended to exhibit high interspecies correlations. This is consistent with our hypothesis that stress-related transcription underlies the significant genomewide similarities observed between aneuploid cells in different species. A linear regression excluding disome IV (SRI~1) is plotted against the data in each graph.

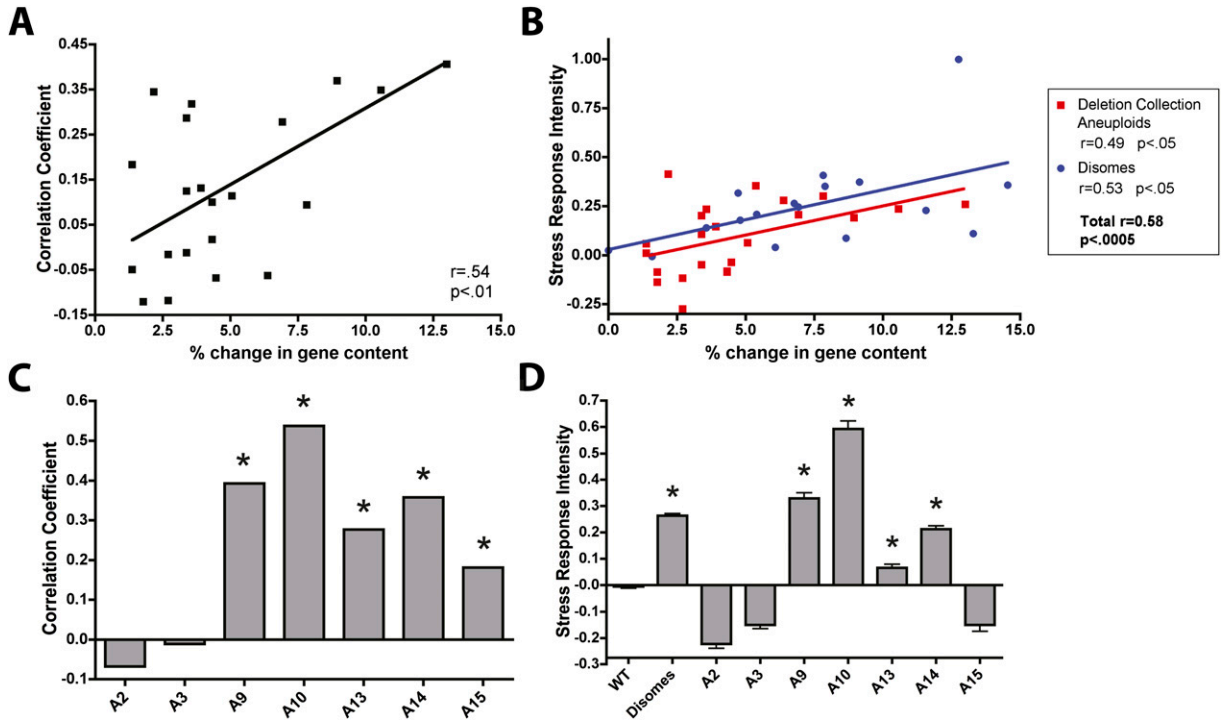

Fig. S3. Effects of aneuploidy in deletion collection strains and triploid meiotic products. (*A*) The correlation coefficients between individual aneuploid deletion collection strains and the disomes were plotted against the percent of the genome that was aneuploid in each strain (see *Materials and Methods*). A linear regression is plotted against the data. (*B*) The stress response intensities of disomic strains and aneuploid deletion collection strains were plotted against the percent of the genome that was aneuploid in each strain. Red and blue lines represent linear regressions plotted against the data. (*C*) The correlation coefficients between the disomic strains and aneuploid products of triploid meiosis were plotted. Karyotypes of these strains are: A2: 1N+II,XII; A3: 1N+I,II,XII; A9: 1N+II,XIII,XVI; A10: 1N+III,XI,XII,XV; A13: 1N+1,II,VIII,XI,XIII; A14: 1N+IX,XVI; and A15: 1N+II,III,VII,IX,X,XI,XII. (*D*) The stress-response intensities of aneuploid products of triploid meiosis were calculated. An asterisk indicates a statistically significant increase in the stress response relative to wild-type ($P < 10^{-5}$, Student *t* test).

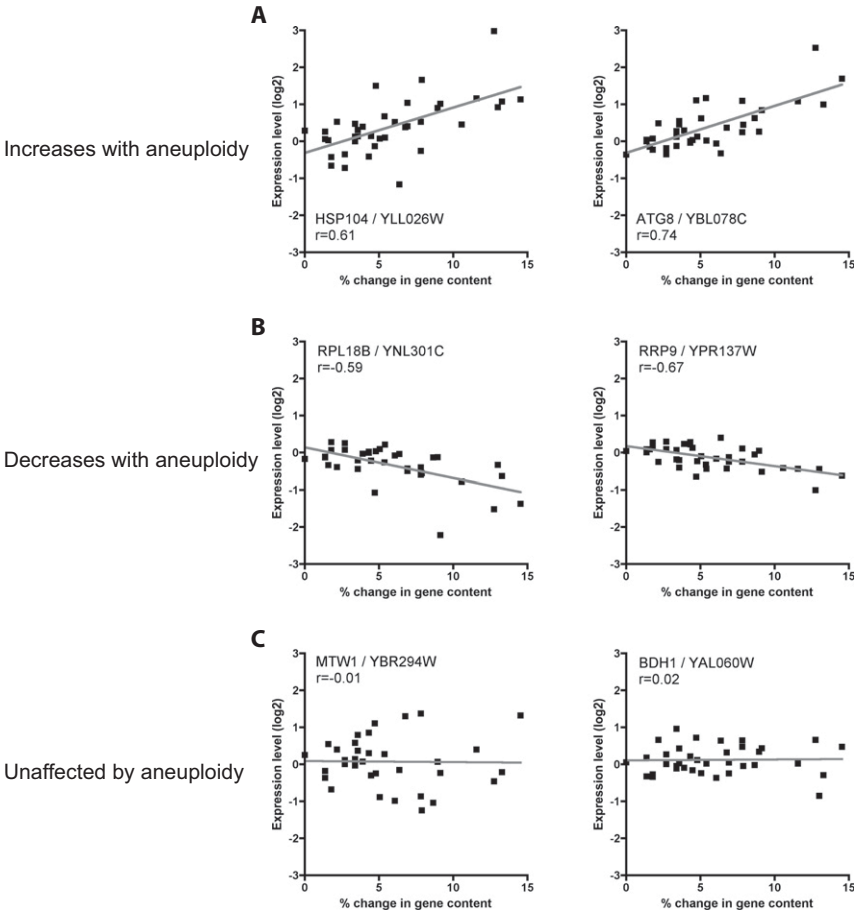

Fig. S4. Aneuploidy-responsive genes in *S. cerevisiae*. Yeast disomes and aneuploid strains from the deletion collection were sorted according to the number of genes present on aneuploid chromosomes in each strain, normalized to the baseline ploidy of each strain. Then, a correlation coefficient was calculated between the degree of aneuploidy (measured as the percent change in gene content) and the transcript level of each gene. A cutoff PCC value of ± 0.5, corresponding to a *P* value less than 0.002, was used to identify aneuploidy-responsive genes. GO term enrichment analysis of these gene sets are presented in Table S1, and examples of the different gene classes identified are displayed here. (*A*) The expression of *HSP104* (*Left*) and *ATG8* (*Right*) tend to increase as the degree of aneuploidy increases. (*B*) The expression of *RPL18B* (*Left*) and *RRP9* (*Right*) tends to decrease as the degree of aneuploidy increases. (*C*) The expression of *MTW1* (*Left*) and *BDH1* (*Right*) are not significantly affected by increasing degrees of aneuploidy.

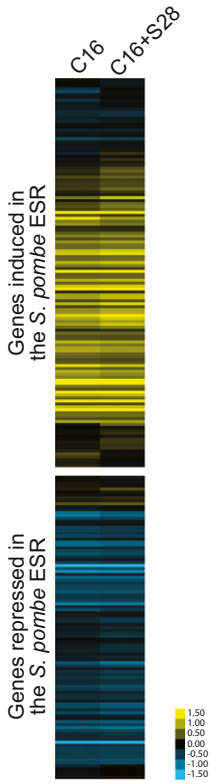

Fig. S5. Aneuploid strains of *S. pombe* display a stress response. Genes that constitute the *S. pombe* ESR were downloaded from http://www.bahlerlab.info/projects/stress/ and the expression values of these genes in the two aneuploid strains were clustered. Of 132 genes that are upregulated in the ESR, 106 are upregulated when the aneuploid strains are averaged. Of 104 genes that are downregulated in the ESR, 97 are downregulated when the aneuploid strains are averaged.

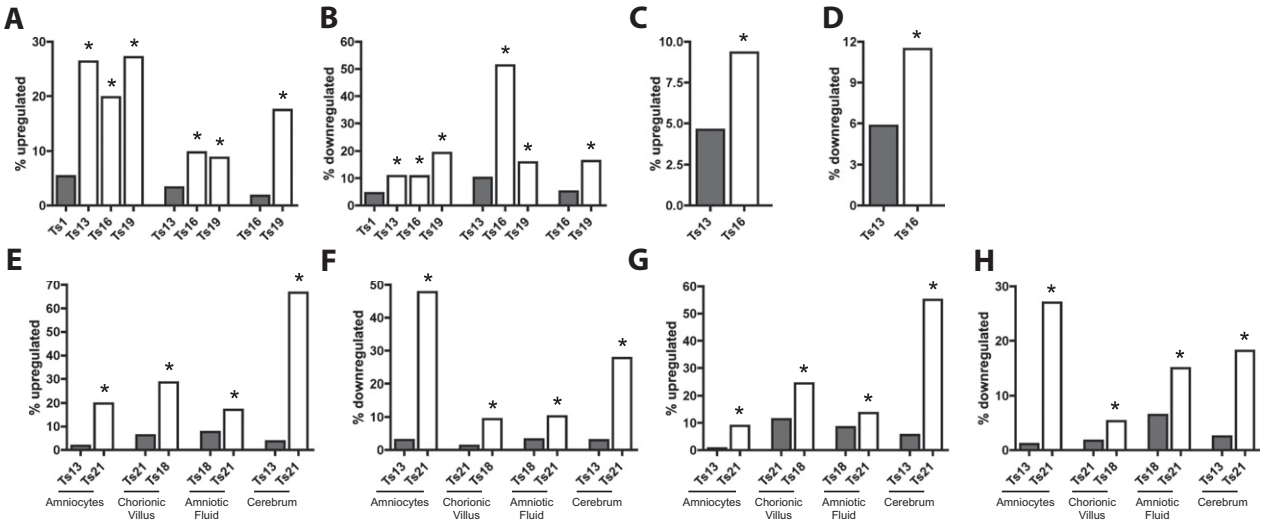

Fig. S6. Expression similarity in trisomic MEFs and human cells. (*A* and *B*) Genes up- or downregulated in one trisomic MEF line are significantly more likely to exhibit a similar change in another trisomy. Gray bars indicate the percentage of all genes up- or downregulated at a 2-FC cutoff in the indicated cell line. White bars indicate the percentage of genes up- or downregulated in that trisomy that are also up- or downregulated in the trisomy represented with a gray bar. (*C* and *D*) Analysis as above, but differentially expressed genes were identified among replicates of Ts13 and Ts16 using a 1.5-FC and $P < 0.05$ (Student *t* test) cutoff. Note that only single replicates of Ts1 and Ts19 were analyzed in (5), precluding the use of a *t* test to analyze these cell lines. (*E* and *F*) Analysis as above using data from human trisomies. Differentially expressed genes were identified using a 2-FC cutoff. (*G* and *H*) Analysis as above, but differentially expressed genes were identified using a 1.5-FC and $P < 0.05$ (Student *t* test) cutoff. Asterisks indicate statistical significance ($P < 0.05$, hypergeometric test).

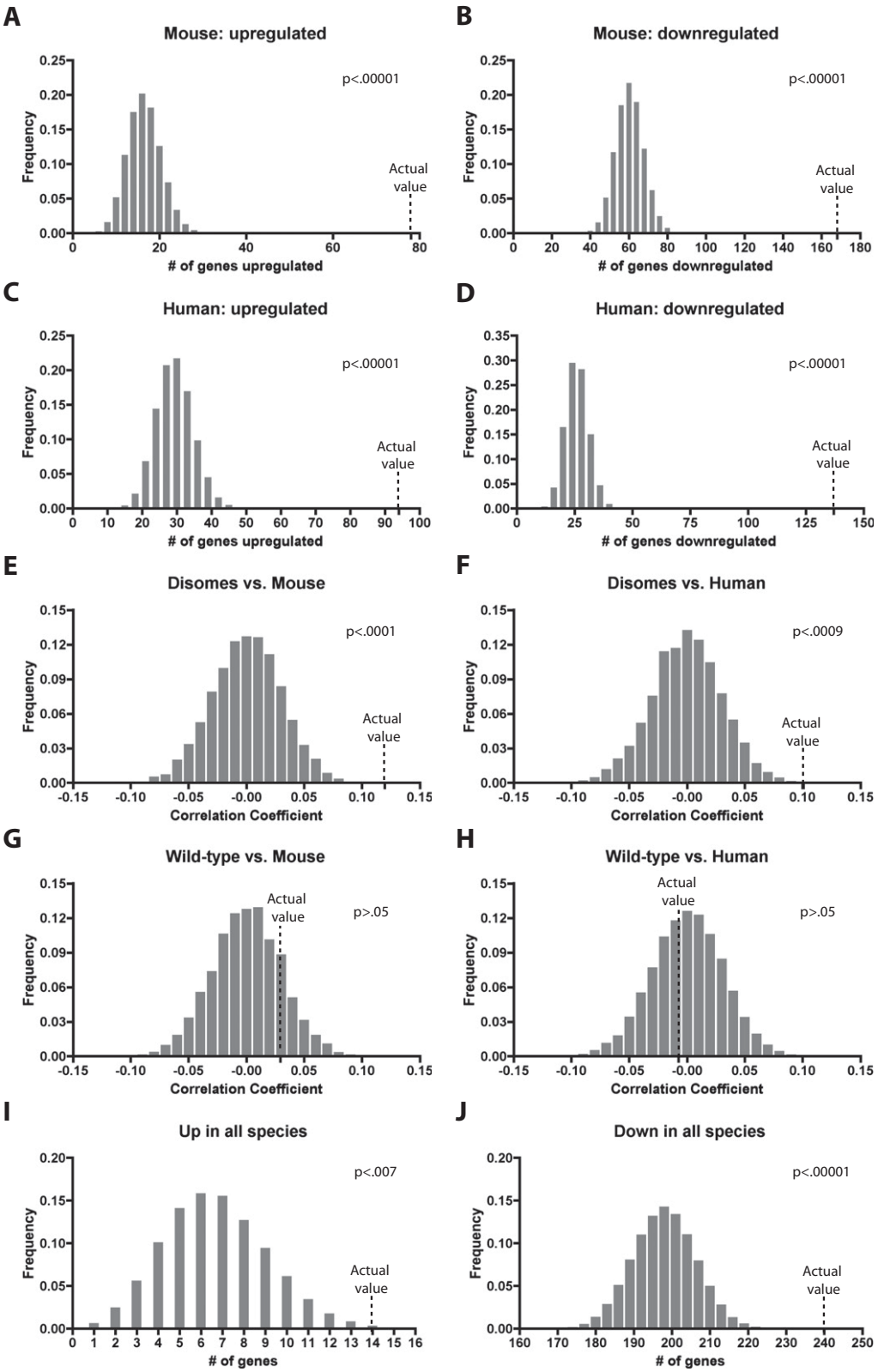

Fig. S7. Permutation testing of similarity significance. (*A*–*D*) To confirm that the number of similarly up- or downregulated genes was significant across mammalian aneuploidies, we randomized gene expression values and then calculated the number of similarly affected genes that occurred by chance. (*A* and *B*) 100,000 random permutations of gene expression data from trisomic MEFs. Bar graphs indicate the number of genes that are (*A*) upregulated or (*B*) downregulated in three or more cell lines. (*C* and *D*) 100,000 random permutations of gene expression data from human trisomies. Bar graphs indicate the number of genes that are (*C*) upregulated or (*D*) downregulated in four or more sample types. (*E*–*H*) To confirm the significance of Pearson correlation comparisons between mammalian cells and other eukaryotes, we randomized gene expression values and then calculated the PCC for each randomized sample. (*E* and *F*) 10,000 random permutations of disomic yeast gene expression values compared with (*E*) mouse and (*F*) human values. (*G* and *H*) 10,000

Legend continued on following page

random permutations of wild-type yeast gene expression values compared with (*G*) mouse and (*H*) human values. (*I* and *J*) To confirm the significance of the codirectional changes among 1:1 orthologs observed in aneuploid cells, we randomized gene expression values and calculated the number of co-directional changes observed in each randomized sample. The number of (*I*) upregulated and (*J*) downregulated genes observed among 100,000 random permutation are displayed.

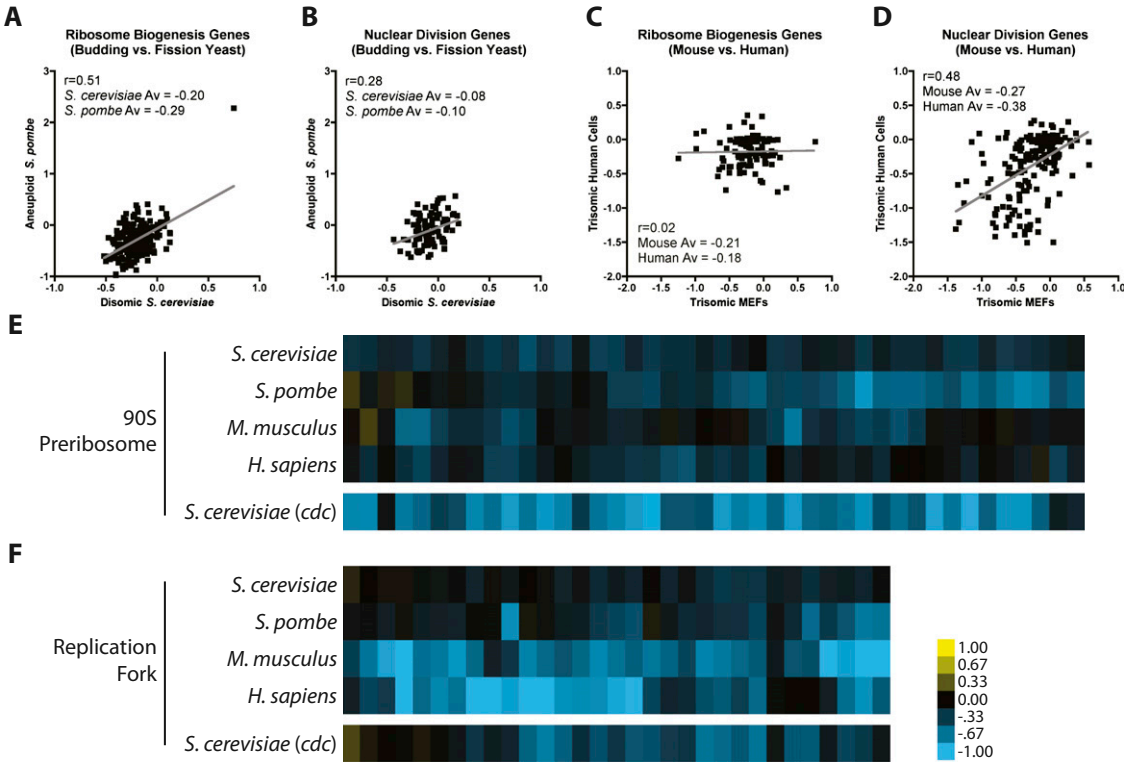

Fig. S8. Cellular processes perturbed across species by stress and aneuploidy. Scatter plots of genes annotated to the indicated GO terms comparing expression levels in either (*A* and *B*) aneuploid budding and fission yeast or (*C* and *D*) trisomic mouse and human cells. Although the average transcripts in each category are expressed at less than euploid levels, cell-cycle transcripts (e.g., nuclear division genes) are expressed at lower levels in mammalian cells, and ribosomal and translational transcripts (e.g., ribosome biogenesis genes) are expressed at lower levels in fungi. Additionally, ribosome biogenesis transcript levels in trisomic MEFs and human cells are uncorrelated with one another, while nuclear division genes [and cell cycle genes generally (Table S14)] show a strong correlation between species. (*E* and *F*) Heat maps of single-ortholog genes in aneuploid budding yeast, fission yeast, mouse, and human cells. Genes that are downregulated by aneuploidy are also down-regulated in ESR-exhibiting yeast *cdc* mutants (the average of *cdc28-4* and *cdc23-1*, *Bottom*).

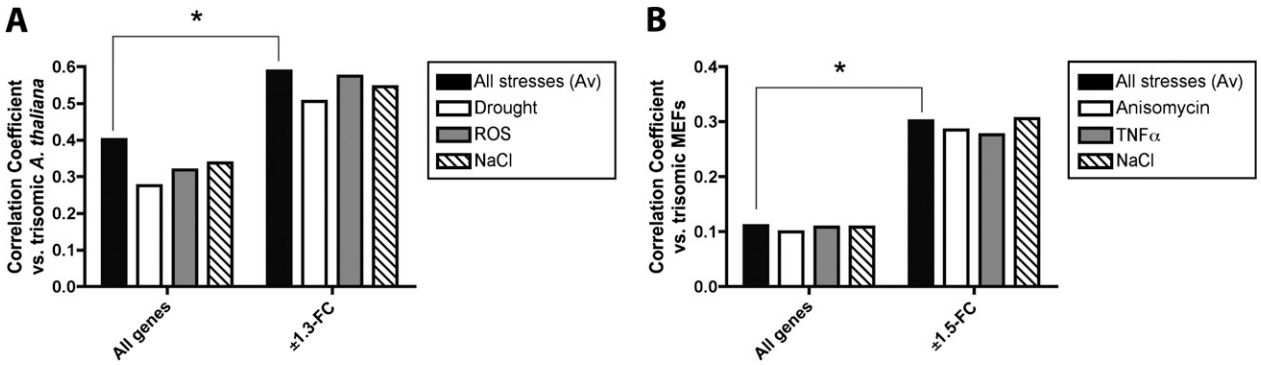

Fig. S9. Similarities between stressed and aneuploid *A. thaliana* and MEFs. (*A* and *B*) The Pearson correlation coefficient between (*A*) trisomic and stressed *A. thaliana* and (*B*) trisomic and stressed MEFs are displayed. When stress-response genes were defined in these species by examining only those genes that changed ± a certain threshold under all stress conditions, the correlation coefficients significantly increased ($P < 0.05$).

Other Supporting Information Files

Tables S1-S18 (XLSX)