Supporting Information

Lemos *et al.* 10.1073/pnas.0805160105

SI Text

Number of Differently Expressed Genes Among Chromosome Substitution Lines: A Simple Regulatory Model of Gene Expression. Chromosome substitution lines are produced by the replacement of one chromosome in an isogenic background; it results in lineages that are genetically diverse for the substituted chromosome while being completely homogeneous for all remaining chromosomes and other genetic elements. Let us assume that the substituted chromosome contains L regulatory loci. Each such locus controls the expression of an independent set of n_{cis} and/or n_{trans} genes located in *cis* and in *trans*, respectively. Gene promoters are classic examples of *cis*-regulatory loci; these DNA sequences usually control the expression of one gene situated downstream in the same DNA molecule. On the other hand, *trans*-acting loci, such as transcriptional factors, activate or repress the expression of one or several genes, regardless of their chromosomal locations (same or different chromosome). In addition, some loci control simultaneously gene expression in *cis* and in *trans*: for instance, a mutation in the promoter of a transcriptional factor increases its level of expression and therefore changes the activation of genes it is regulating. The substituted chromosome contains a proportion $p_{cistrans}$ of these loci with both *cis* and *trans* regulation and, p_{cis} and p_{trans} loci with pure *cis* or *trans* regulation, respectively ($p_{cistrans} + p_{cis} + p_{trans} = 1$). Let us also assume that two alleles at a same locus can only have additive and dominant/recessive interactions (no transvection) and two alleles at different loci cannot interact (no epistasis). Under these assumptions, the number of genes differently expressed between two substitution lines depends on the genetic identities I_{xy} between the substituted chromosomes x and y. The total identity (I_{xy}) between two chromosomes is comprised of identities at *cis*-regulatory loci ($I_{xy,\ cis}$), *trans*-regulatory loci ($I_{xy,\ trans}$), and *cis-trans*-regulatory loci ($I_{xy,\ cistrans}$). Finally, genetic divergence between two chromosomes x and y is defined as $v_{xy} = 1 - I_{xy}$. A diagram of the model is shown in SI Fig. S3.

Comparison between two homozygous lines:

The total number of genes differently expressed between two homozygous substitution lines is

$$G_{xx,yy} = L[(v_{xy,cis}p_{cis} + v_{xy,cistrans}p_{cistrans})n_{cis} + (v_{xy,trans}p_{trans} + v_{xy,cistrans}p_{cistrans})n_{trans}], \quad [1]$$

where $v_{xy,cis}$, $v_{xy,trans}$ and $v_{xy,cistrans}$ are the genetic divergences between two substituted chromosomes x and y, at loci regulating genes *in cis*, *in trans* or both, respectively.

The substituted chromosome contains all differently expressed genes regulated in *cis* and a fraction of the ones regulated *in trans*. Calling k this fraction, the number of differently expressed genes carried by the substituted (G^{sub}) chromosome is

$$G^{sub}_{xx,yy} = L[(v_{xy,cis}p_{cis} + v_{xy,cistrans}p_{cistrans})n_{cis} + k(v_{xy,trans}p_{trans} + v_{xy,cistrans}p_{cistrans})n_{trans}], \quad [2]$$

and consequently

$$G^{back}_{xx,yy} = L(1-k)(v_{xy,trans}p_{trans} + v_{xy,cistrans}p_{cistrans})n_{trans}, \quad [3]$$

for the background (G^{back}) chromosomes.

Comparison between one homozygous and one heterozygous line:

In the case of one homozygous line xx and one heterozygous line xy, the difference in gene expression can be masked by dominant alleles carried by the chromosome x. Consequently, the total number of differently expressed genes is

$$G_{xx,xy} = (1 - d_{x,xy})G_{xx,yy}, \quad [4]$$

where $d_{x,xy}$, measures the proportion of loci where the allele on chromosome x is dominant over the allele carried by the chromosome y.

Comparison between two heterozygous lines: In the same way, the number of differently expressed genes between two heterozygous lines, xz and yz, depends simply on the genetic divergence v_{xy} and thus for the substituted chromosome:

$$G_{xz,yz} = (1 - d_{z,xy})G_{xx,yy}, \quad [5]$$

In this case, $d_{z,xy}$ measures the proportion of loci where the allele on chromosome z is dominant over alleles carried by the chromosomes x and y.

Estimation of the proportion of dominant loci:

Using Eqs. **4** and **5**, we can estimate the proportion of loci with dominant alleles as

$$d_{z,xy} = 1 - \frac{G^{sub}_{xz,yz}}{G^{sub}_{xx,yy}} = 1 - \frac{G^{back}_{xz,yz}}{G^{back}_{xx,yy}}. \quad [6]$$

Observed number of *cis*- and *trans*-regulated genes:

From Eq. **3**, we can estimate the observed number of *trans*-regulated genes as

$$G^{trans}_{xx,yy} = L(v_{xy,trans}p_{trans} + v_{xy,cistrans}p_{cistrans})n_{trans} = \frac{G^{back}_{xx,yy}}{(1-k)}. \quad [7]$$

By combining Eqs. **2** and **6**, the observed number of *cis*-regulated genes is:

$$G^{cis}_{xx,yy} = L(v_{xy,cis}p_{cis} + v_{xy,cistrans}p_{cistrans})n_{cis} = \frac{(1-k)G^{sub}_{xx,yy} - kG^{back}_{xx,yy}}{(1-k)}. \quad [8]$$

For comparisons including heterozygous lines, the observed numbers of *trans*- and *cis*-regulated genes are:

$$G^{trans}_{xw,yw} = (1 - d_{w,xy})G^{trans}_{xx,yy} = \frac{G^{back}_{xw,yw}}{(1-k)}, \quad [9]$$

and

$$G^{cis}_{xw,yw} = (1 - d_{w,xy})G^{cis}_{xx,yy} = \frac{(1-k)G^{sub}_{xw,yw} - kG^{back}_{xw,yw}}{(1-k)}, \quad [10]$$

respectively (see Fig. S4).

From Eqs. **9** and **10**, we can estimate the ratio $G^{trans}_{xw,yw}/G^{cis}_{xw,yw}$ as

$$\frac{L(v_{xy,trans}p_{trans} + v_{xy,cistrans}p_{cistrans})n_{trans}}{L(v_{xy,cis}p_{cis} + v_{xy,cistrans}p_{cistrans})n_{cis}} = \frac{G^{back}_{xw,yw}}{(1-k)G^{sub}_{xw,yw} - kG^{back}_{xw,yw}}. \quad [11]$$

Further, we can rewrite Eq. **11** as follows:

$$\frac{L_{trans}n_{trans}}{L_{cis}n_{cis}} = \frac{G^{\text{back}}_{xw,yw}}{(1-k)G^{\text{sub}}_{xw,yw} - kG^{\text{back}}_{xw,yw}}, \qquad [12]$$

where L_{cis} and L_{trans} are the numbers of polymorphic loci regulating gene expression in *cis* and in *trans*, respectively (Fig. S5). Fig. S5 uses the number of differential expression estimated from the data and may be illustrative about the sizes of the effects of *trans*-regulatory loci relative to *cis*-regulatory loci. For instance, under the assumptions that the average number of genes affected by a *cis*-regulatory locus is one ($n_{cis} = 1$) and that the number of polymorphic *cis*-regulatory loci is between 8 and 16 times larger than the number of *trans*-regulatory loci, one might estimate that the average number of *trans*-effects of a *trans*-regulatory loci (n_{trans}) is between 10 and 20 genes.

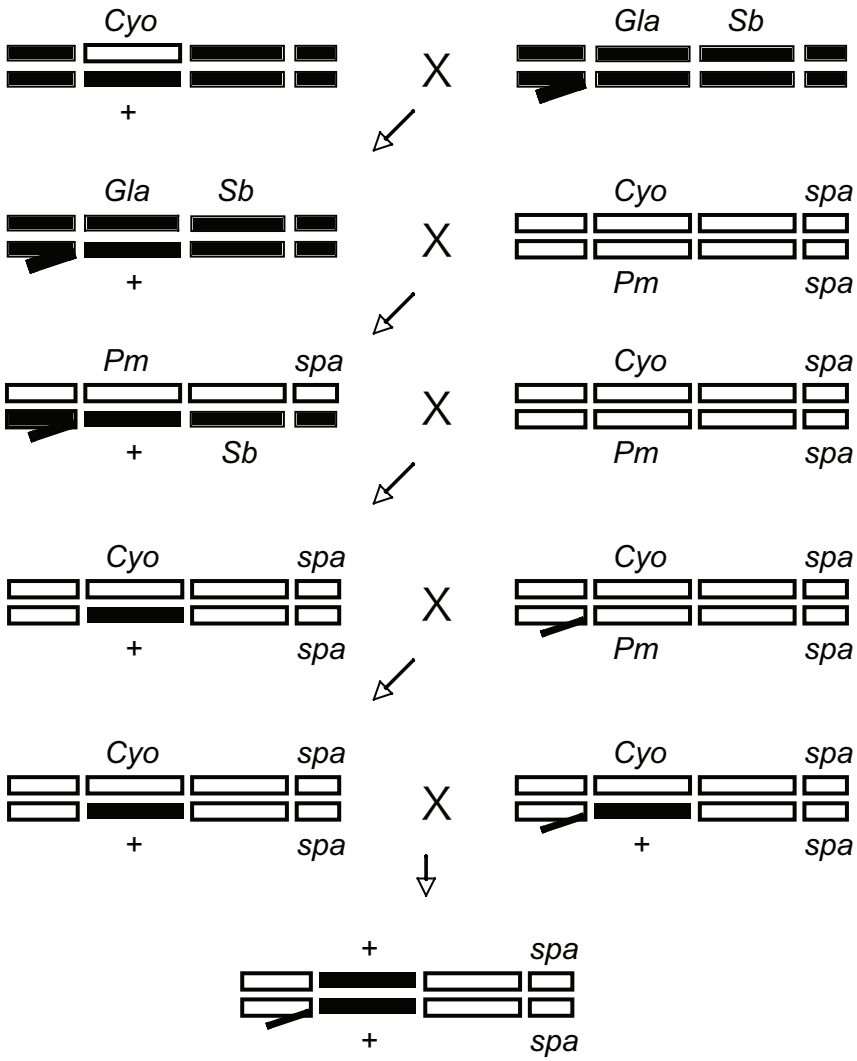

Fig. S1. Cross scheme for obtaining stable second-chromosome substitution lines.

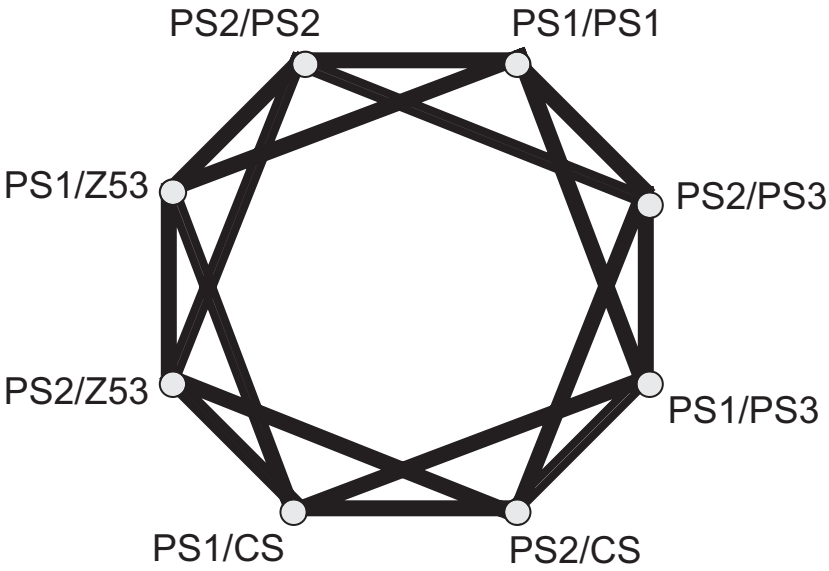

Fig. S2. Experimental design for collecting microarray gene expression data. Each line represents two hybridizations in which the Cy3 and Cy5 dyes were swapped, for a total of eight replicates per genotype.

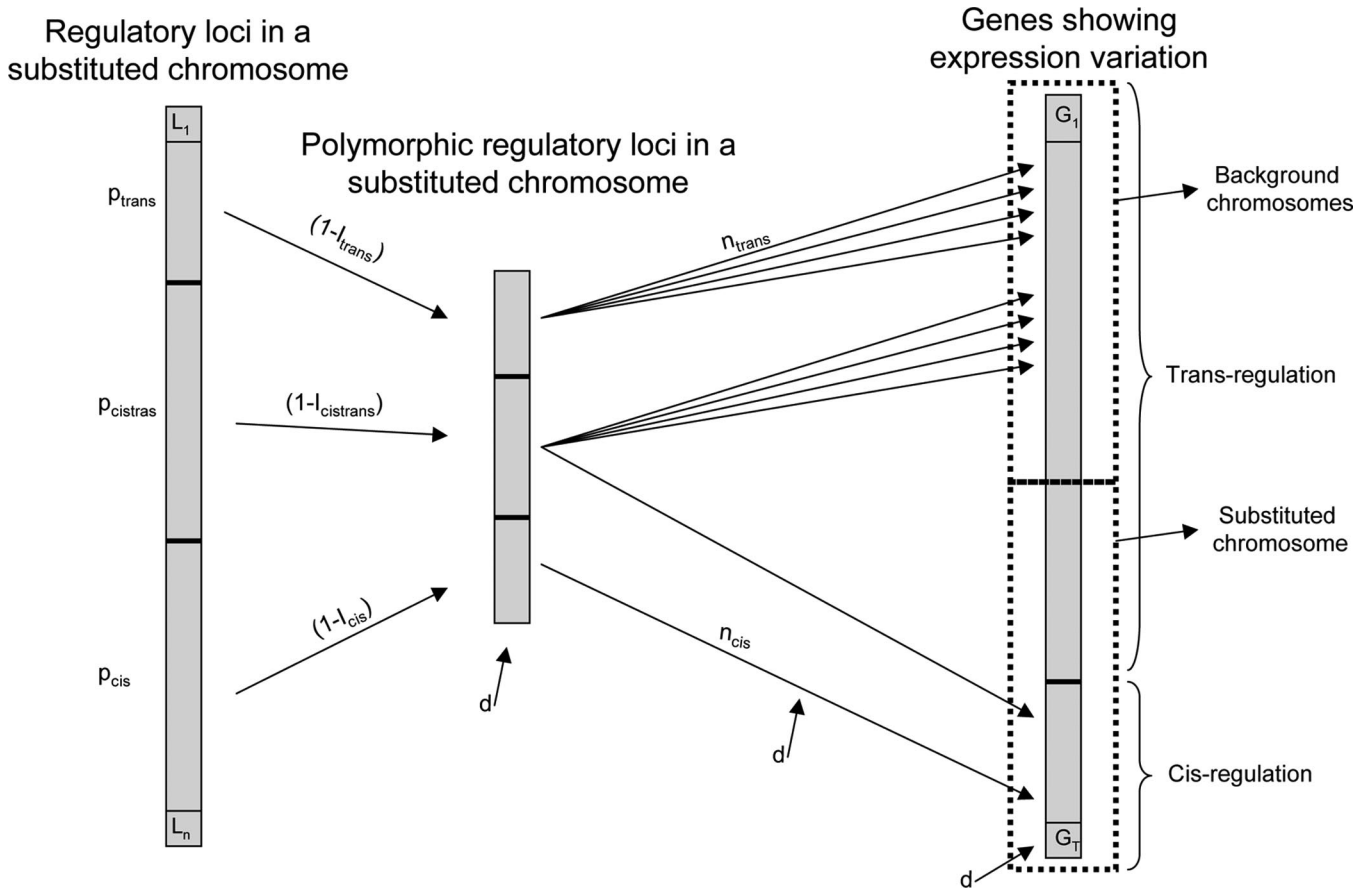

Fig. S3. Diagram of the model. (1-*Ixy*) denotes the divergence between two chromosome in the set of regulatory loci with *cis*-effects (Ixy, *cis*), *trans*-effects (*Ixy*, *trans*), and *cis*- and *trans*-effects (*Ixy*, cistrans). *p* denotes the fraction of all loci with regulatory effects *in cis* (pcis), *trans* (ptrans), and *cis* and *trans* (pcistrans). *n* denotes the number of genes affected per unit of change in regulatory loci with *cis*-effects (ncis) and *trans*-effects (ntrans). Finally, *d* denotes the fraction of loci showing dominance effects. In the framework of our model dominance as observed at the level of steady state, mRNA abundances for a gene may be the outcome of at least two distinct mechanisms (noted with an arrow in the diagram). First, dominance may arise at the level of the loci harboring heterozygous alleles, where one allele is expressed at a higher strength than the other (allelic dominance). Second, dominance may arise from properties of the regulatory networks and may, for instance, affect *cis*- and *trans*-variation differently (regulatory dominance).

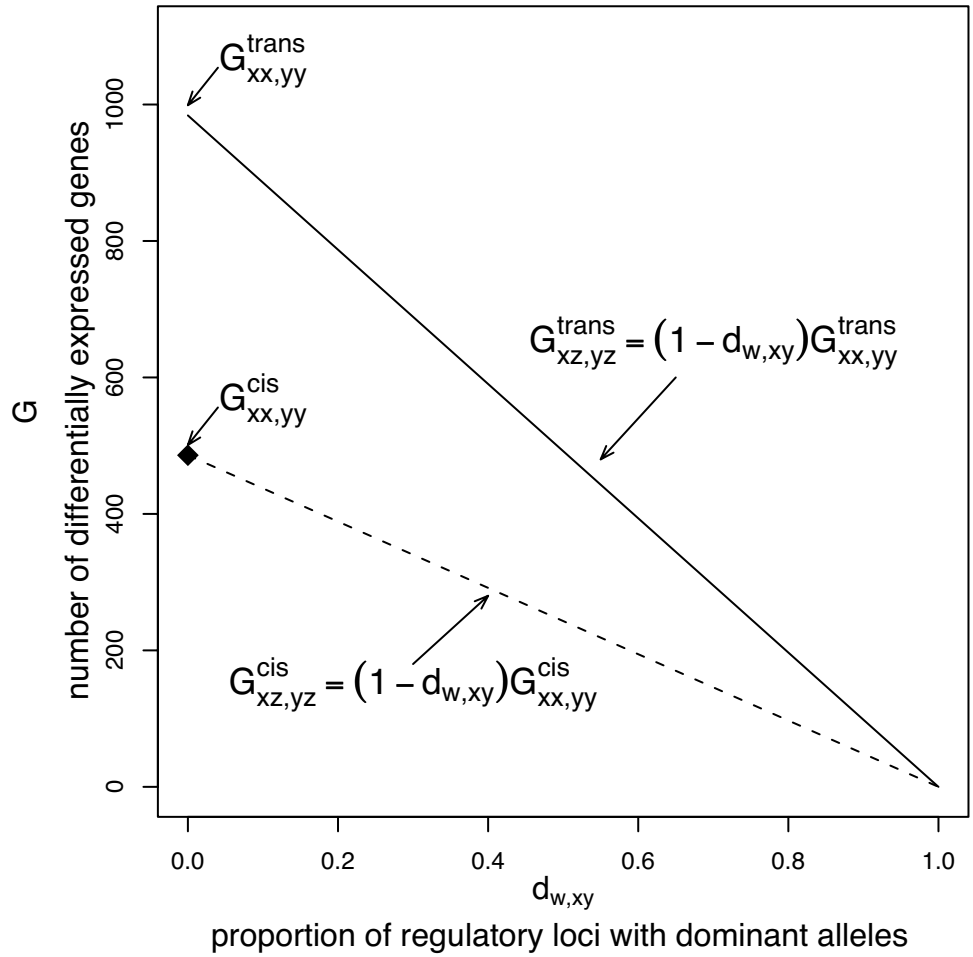

Fig. S4. Number of differentially expressed genes in homozygous and heterozygous comparisons. The filled circle and filled square show the number of genes differentially expressed in the homozygous-homozygous contrast. The lines show the expected reduction in differential expression due dominance effects in contrasts involving heterozygous. The parameters used are L = 3,000, vcis = 0.01, vtrans = 0.01, vcistrans = 0.01, pcis = 0.2, ptrans = 0.4, pcistrans = 0.4, k = 2/5, ncis = 1, and ntrans = 2.

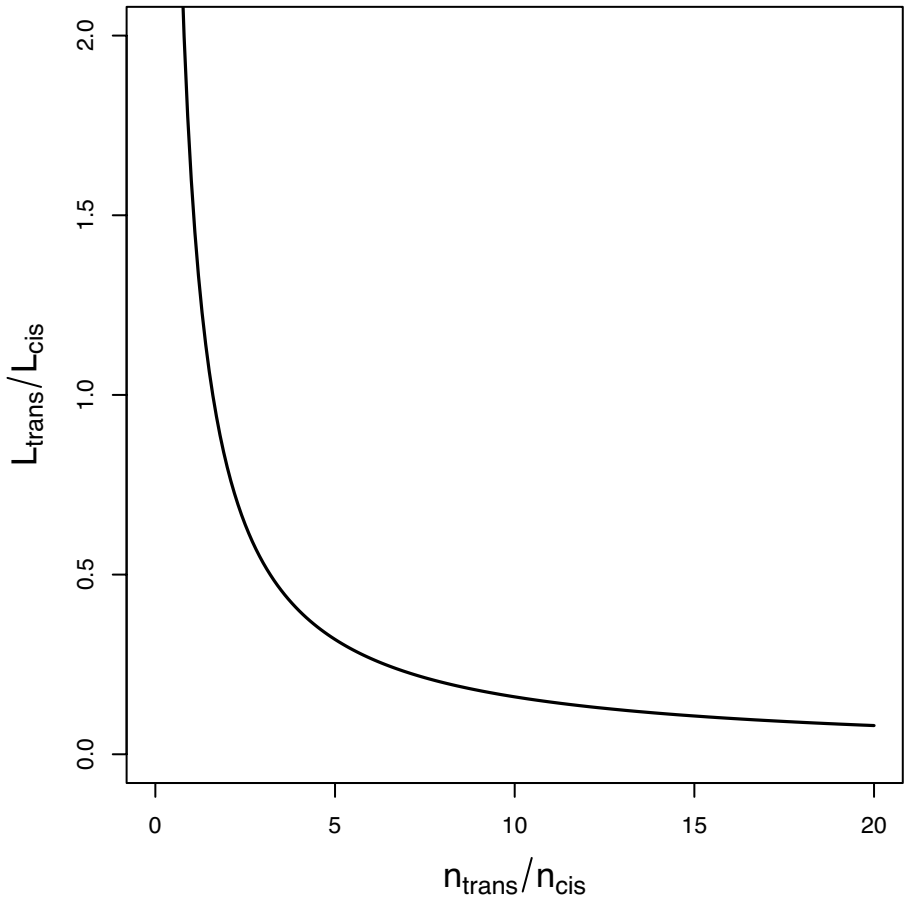

Fig. S5. Ratio of polymorphic *trans*-regulatory (Ltrans) and *cis*-regulatory (Lcis) loci as a function of the ratio of the number of *trans*-regulated genes per *trans*-regulatory locus (ntrans) over the number of *cis*-regulated genes per *cis*-regulatory locus (ncis) (see Eq. **12**). The parameters used are k = 2/5, Gback, xw,yw = 155, and Gsub, xw,yw = 265 (these values are estimated from the data; see homozygous-heterozygous comparisons in the text).

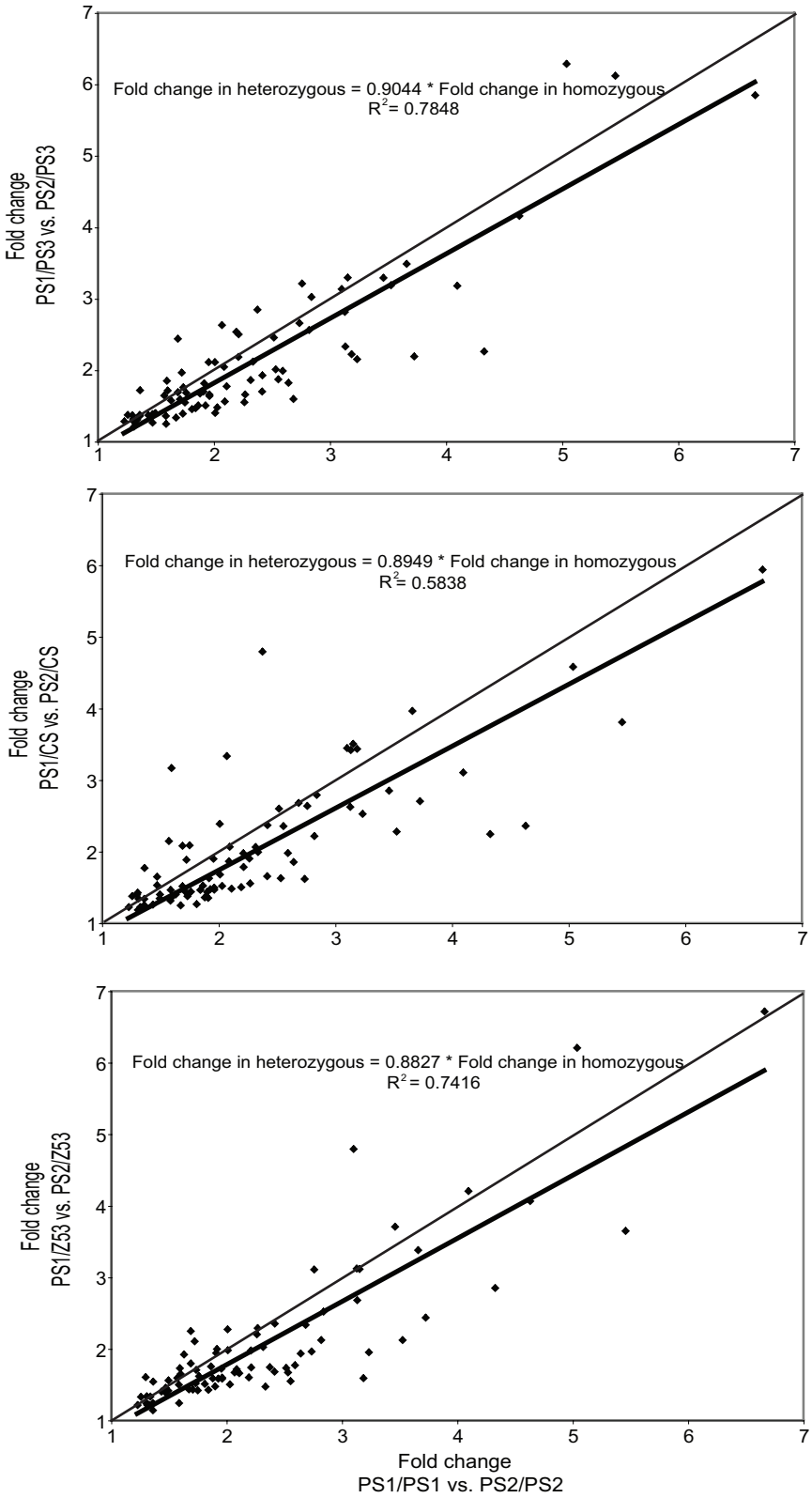

Fig. S6. Fold change for ''phenotypically dominant'' differences in homozygous-homozygous contrasts and in heterozygous-heterozygous contrasts. Thick line represents the regression slope.

Table S1. Inference of allelic effects in a heterozygous vs. heterozygous comparison of the kind AC vs BC, where A, B, and C are whole chromosomes

PS1 PS3 vs PS2 PS3	PS3 allele Dominant	PS3 allele Recessive	PS3 allele Additive
○ ○ vs ● ○	Equal expression	Differential expression	Differential expression
○ ● vs ● ●	Equal expression	Differential expression	Differential expression

Dots represent alleles, which can take two states (white or black) and are different between chromosomes A and B. If the allele in C (which in our experiment comes from PS3, CS, or Z53) is dominant over one of the alleles in A (PS1) or B (PS2), the expression difference detected in the homozygous (AA) vs. homozygous (BB) comparison will be masked in the heterozygous (AC) vs. heterozygous (BC) comparison. If the allele in C is recessive or if the two allelic states have additive effects, then the expression difference between homozygous will not be masked in the heterozygous (AC) vs. heterozygous (BC). Hence, gene expression differences between PS1/PS1 vs. PS2/PS2 that are not present in the PS1/PS3 vs. PS2/PS3 arise because the allele present in either PS1 or PS2 is recessive; while the PS3 allele is dominant. Conversely, if the allele in PS3 takes the recessive state then the difference originally identified in the PS1/PS1 vs PS2/PS2 contrast will be maintained in the PS1/PS3 vs. PS2/PS3 contrast. This differential expression will also be maintained when the two alleles are additive.

Table S2. Fold change and chromosome distribution of gene expression differences that are masked (phenotypically recessive) and not masked (phenotypically dominant) in heterozygous

Mode	Chromosome	Fold Change			Mean Fold Change (a)	Total
Phenotypically recessive		<2	2–3	>3		
	X	60 (98%)	1		1.3	61
	3	213 (95%)	11 (5%)	1	1.3	225
	2	303 (94%)	16 (5%)	4	1.5	323
Phenotypically dominant	X and 3	5 (55%)	4 (45%)		1.9	9
	2	42 (47%)	27 (30%)	20 (22%)	2.7 (b)	89

(a) Mean fold change between genotypes *PS1*/*PS1* and *PS2*/*PS2*.
(b) After removal of an outlier; otherwise mean fold change = 3.9.