Supplementary Information

Supplement to “Genome-wide allele- and strand-specific expression profiling”, Julien Gagneur, Himanshu Sinha, Fabiana Perocchi, Richard Bourgon, Wolfgang Huber and Lars M. Steinmetz

Web site

Further information is provided on the website http://steinmetzlab.embl.de/allelic including a query interface for the expression levels of all transcripts and the fits on the mixture series.

Microarray data

All microarray data are accessible at ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae/). The cDNA hybridizations are available under the accession number E-TABM-569 and the array design under A-AFFY-116. We used the following genomic DNA hybridizations of Mancera *et al*. (2008) (E-TABM-470):

recombination_060501_S96, recombination_060501_YJM789,
recombination_060502_S96, recombination_060502_YJM789,
recombination_060503_S96, recombination_060503_YJM789,
recombination_060504_S96, recombination_060504_YJM789

Fit of the model without parental cDNA

The model can be fit using cDNA of the hybrid only. Relative ADE coefficients obtained from this restricted dataset highly correlates with the original fit that includes parental cDNA (Supplementary Figure S5, Pearson’s correlation coefficient 0.932). A slightly larger dispersion of the allelic differences can be noticed in the hybrid-only analysis versus the full dataset. This is likely due to the smaller amount of hybridizations used to infer probe affinity, which yields noisier allelic level estimates and a larger variability of allelic expression differences. This also indirectly shows that not only the gDNA samples contribute to the estimation of the probe affinities but also the cDNA samples.

Sample-to-sample variation

Having biological replicates also enabled assessing sample-to-sample variation in ADE measurements, by performing the analysis on each replicate separately. Differences between replicates are to some extent attributable to noise, but may also reflect true differences in allele-specific expression in distinct samples. Relative ADE coefficient, which measures the degree of ADE (see Methods) inferred from each sample strongly correlated with one another (Pearson's correlation ranges between 0.748 and 0.826 for the expressed transcripts with 8 CSPs or more, Supplementary Figure S6). Importantly, very few transcripts had a strong relative ADE coefficient in one sample but not in the others, indicating that sample-specific ADE was rare. Therefore we proceeded with allelic expression inferred from the combined analysis of all three biological replicates.

Variance scaling

The probe intensity variance is modeled as a second order polynomial function of the intensity (Equation 2, Methods). This implies specific asymptotic behavior of the standard deviation. For intensities close to 0, the standard deviation reaches a constant that corresponds to an additive noise at background level. For large intensities, the standard deviation is approximately proportional to intensity. Supplementary Table 8 gives the additive and multiplicative parameters for each hybridization. Note that the additive parameter is hybridization-specific while the multiplicative parameter is common to all hybridizations of the same type (cDNA and gDNA). The additive parameter is similar across all hybridizations. However, the multiplicative parameter is about 2 fold smaller for genomic DNA hybridizations. This implies that for similar intensities, the genomic DNA hybridizations have better precision and thus weight more in the fitting than the cDNA hybridizations.

Supplementary Table I Transcript expression levels, confidence intervals and differential expression FDR

See suppl_table_1.xls

Supplementary Table II List of sense-antisense pairs and their expression levels

See suppl_table_2.xls

Supplementary Table III Transcription factor target sets enrichment for differentially expressed genes between S and Y

See suppl_table_3.xls

Supplementary Table IV 32 PHO pathway genes (Ogawa *et al*, 2000; Wykoff *et al*, 2007) covered by our dataset (in alignment of Y and S genome and with at least 20 probes).

See suppl_table_4.xls

Supplementary Table V Growth phenotypes in Arsenate containing media for 184 genotyped Y/S segregants

See suppl_table_5.xls

Supplementary Table VI List of strains used in this study

Strain name	Genetic background	Parental strain	Genotype	Reference
S1003	S	S96	*MATa/α lys5/lys5*	(Steinmetz *et al*, 2002)
S1766	S	S1003	*MATa lys5*	This study
S1767	S	S1003	*MATα lys5*	This study
S1769	S	S1003	*MATα lys5*	This study
S1776	S	S1766	*MATa lys5 pho84Δ::NatMX4*	This study
S96	S	S288c	*MATa lys5*	(Steinmetz *et al*, 2002)
XHS768	Y/S	YHS959 x S1767	*MATa(Y)/α(S) hoΔ::loxP-KanMX4-loxP/ho LYS2/lys2 LYS5/lys5*	This study
XHS769	Y/S	YHS960 x S1766	*MATa(S)/α(Y) hoΔ::loxP-KanMX4-loxP/ho LYS2/lys2 LYS5/lys5*	This study

XHS770	Y/S	YHS961 x S1769	*MATa(Y)/α(S) hoΔ::loxP-KanMX4-loxP/ho LYS2/lys2 LYS5/lys5*	This study
XHS788*	Y/S	YHS960 x S1776	*PHO84-Y/pho84-SΔ::NatMX4*	This study
XHS789*	Y/S	YHS969 x S1767	*pho84-YΔ::HygMX4/PHO84-S*	This study
YHS957	Y	YJM155	*HO/hoΔ::loxP-Kan-loxP*	This study
YHS959	Y	YJM155	*MATa lys2*	This study
YHS959	Y	YHS957	*MATa lys2*	This study
YHS960	Y	YHS957	*MATα lys2*	This study
YHS961	Y	YHS957	*MATa lys2*	This study
YHS969	Y	YHS959	*MATa lys2 pho84Δ::NatMX4*	This study
YJM145	Y	Clinical isolate	*MATa/α*	(Steinmetz *et al*, 2002)
YJM155	Y	YJM145	*MATa/α lys2/lys2*	(Steinmetz *et al*, 2002)
YJM789	Y	YJM145	*MATα ho::hisG lys2*	(Wei *et al*, 2007)

* reciprocal hemizygote strains used to confirm *PHO84* as trans-acting factor for the PHO pathway regulation

Supplementary Table VII Allelic expression ratios from sequence traces.

See suppl_table_7.xls

Supplementary Table VIII Asymptotic standard deviation parameters per hybridization. There was a mistake in the naming of the original CEL files of the genomic DNA hybridizations. The true genotypes are those given in the "sample.type" column. Correct annotations of those samples are also provided in ArrayExpress.

See suppl_table_8.xls

Supplementary Figure S1 Cumulative distribution of allelic expression ratios for the transcripts allelic expression ratios for the 454 transcripts with significant ADE (FDR<0.05).

Supplementary Figure S2 Allelic differential expression and polymorphism density in promoter regions along the genome. For each nuclear chromosomes, the ADE coefficient (ranging from 0 for equal expression to 1 for mono-allelic expression) of each transcript is displayed in the upper lane (green gradient). The lower lane (purple gradient) shows the density of polymorphisms in the promoter region (500 bp upstream of the transcript start site). The inset displays a highly polymorphic region on chromosome I encoding several members of the *DUP240* gene family. Barplots show the fitted expression levels for the S allele (blue) and Y allele (red). Error bars represent 95% confidence intervals. All genes exhibit significant ADE (FDR < 0.05).

Supplementary Figure S3 Distribution of the ratio of cis-regulatory divergence to the total regulatory divergence (Methods) for the 455 transcripts with at least 1.5-fold expression difference between the S and the Y strains and confident ADE estimates. Upper panel, cumulative distribution. Plain circles indicate position for ratio of 1/3, median ratio (0.40) and 2/3. Lower panel, histogram of the same quantity.

Supplementary Figure S4 Arsenate resistance of *PHO84* reciprocal hemizygotes. Strains S1003 (S/S), YJM155 (Y/Y), XHS768 (Y/S), XHS788 (PHO84-Y/–) and XHS789 (–/PHO84-S) were grown for 48 hours on control YPD plates (left panel) and YPD + Arsenate 2mM (right panel).

Supplementary Figure S5 Relative ADE coefficients fitted using the 3 cDNA samples of the hybrid only versus the full dataset (i.e., 3 cDNA of the hybrid + 6 cDNA of the parental strains). Both analyses also use the genomic DNA hybridizations of the parental strains. For the expressed transcripts with 8 CSPs or more, the relative ADE coefficients $\frac{h_Y - h_S}{h_Y + h_S}$ inferred from each analysis are plotted against each other (top of panels, Pearson's correlation coefficient). The y=x line (grey) is provided as a guide.

Supplementary Figure S6 Relative ADE coefficient in individual hybrid samples. Sample-to-sample variation was assessed by fitting allelic levels using each of the three biological hybrid replicates independently (sample 1…3). For the expressed transcripts with 8 CSPs or more, the relative ADE coefficients $\frac{h_Y - h_S}{h_Y + h_S}$ inferred from each sample individually are plotted against each other (top of panels, Pearson's correlation coefficient). The y=x line (grey) is provided as a guide.

References

Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM (2008) High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454:** 479-485.

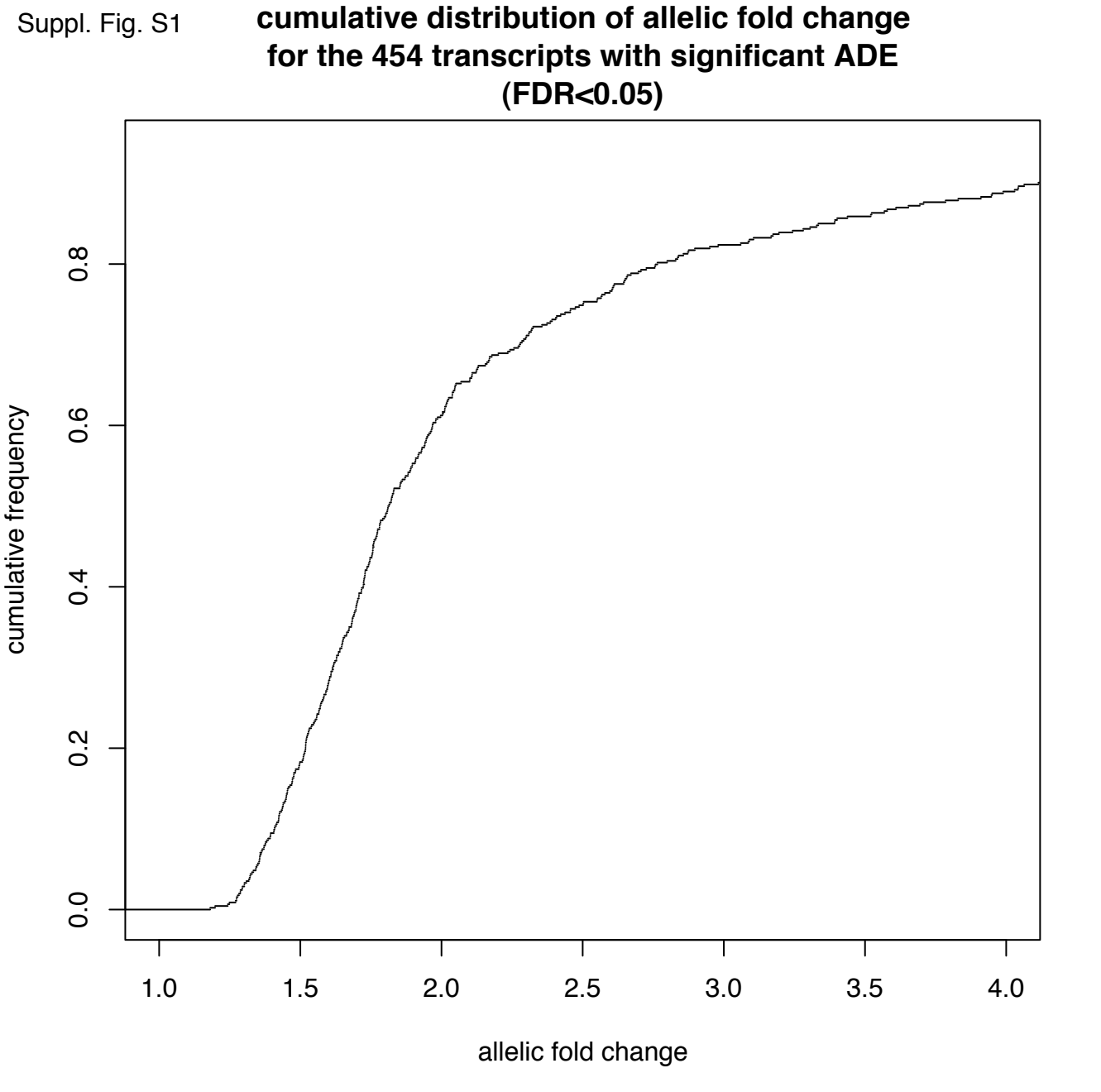
Ogawa N, DeRisi J, Brown PO (2000) New components of a system for phosphate accumulation and polyphosphate metabolism in Saccharomyces cerevisiae revealed by genomic expression analysis. *Mol Biol Cell* **11:** 4309-4321.

Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, Davis RW (2002) Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416:** 326-330.

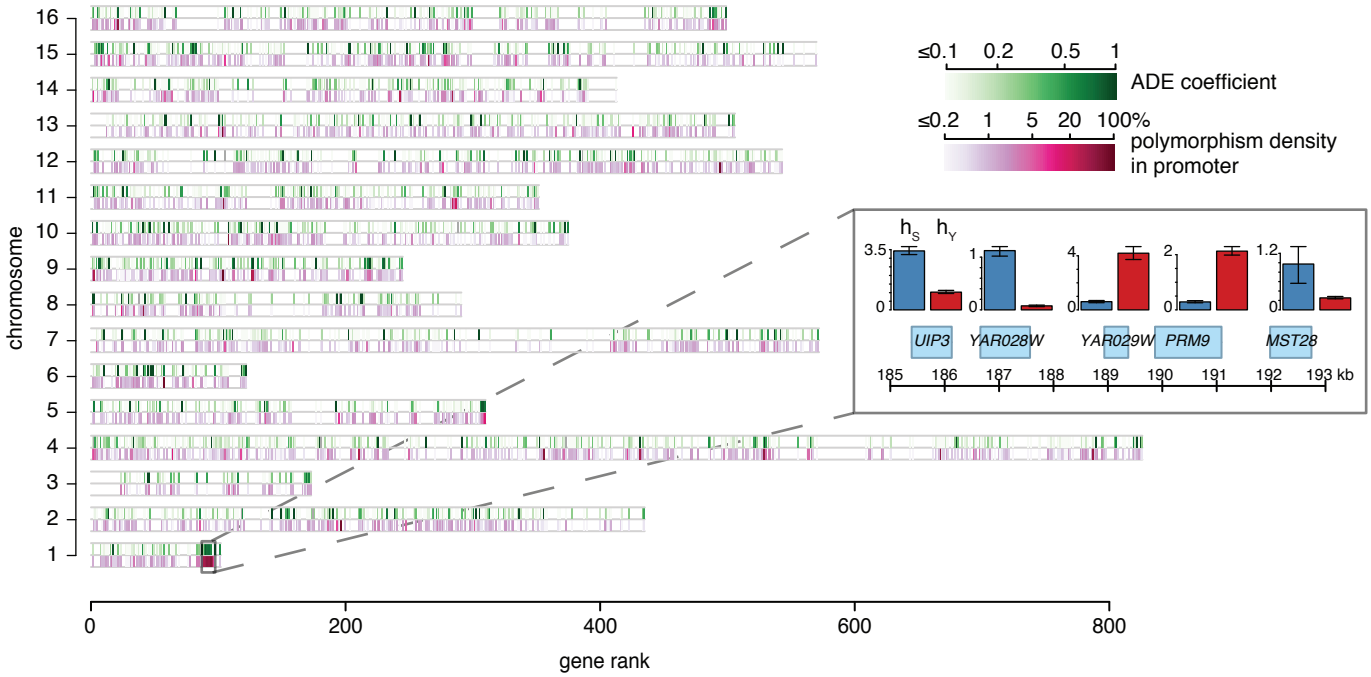
Wei W, McCusker JH, Hyman RW, Jones T, Ning Y, Cao Z, Gu Z, Bruno D, Miranda M, Nguyen M, Wilhelmy J, Komp C, Tamse R, Wang X, Jia P, Luedi P, Oefner PJ, David L, Dietrich FS, Li Y, et al. (2007) Genome sequencing and comparative analysis of Saccharomyces cerevisiae strain YJM789. *Proc Natl Acad Sci USA* **104:** 12825-12830.

Wykoff DD, Rizvi AH, Raser JM, Margolin B, O'Shea EK (2007) Positive feedback regulates switching of phosphate transporters in S. cerevisiae. *Mol Cell* **27:** 1005-1013.

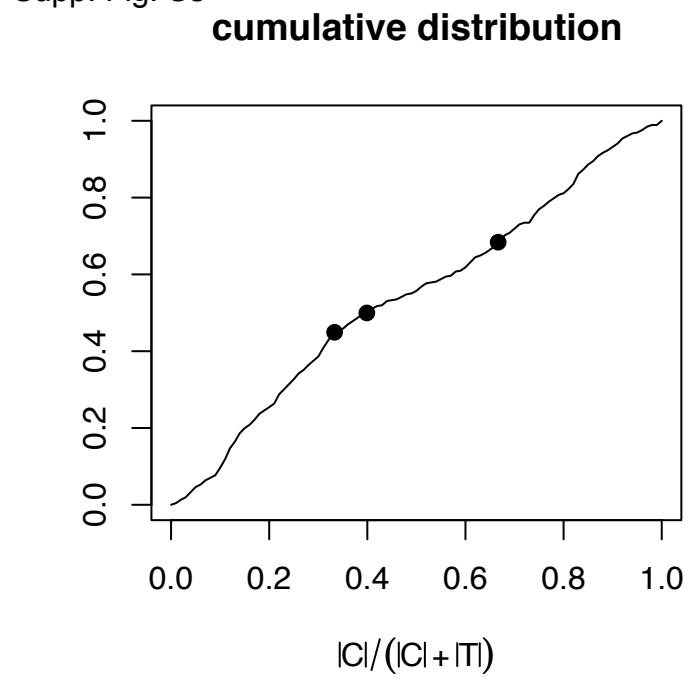
Suppl. Fig. S1

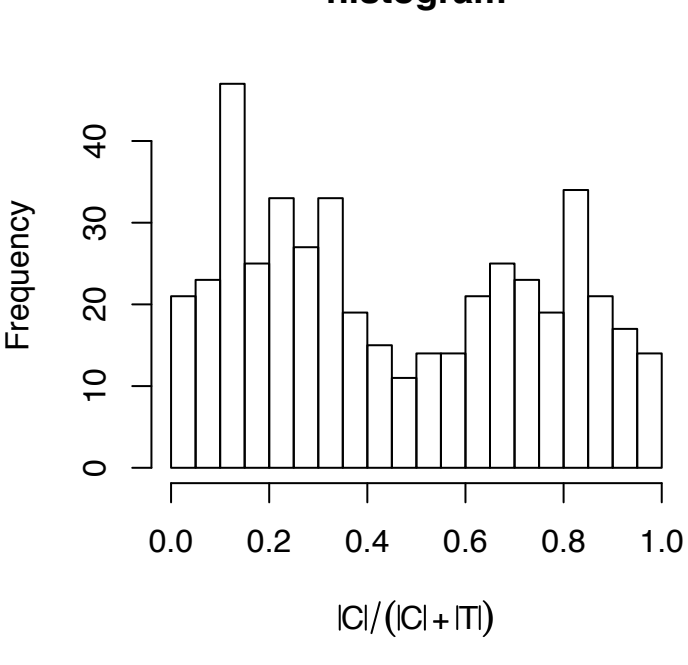

cumulative distribution of allelic fold change
for the 454 transcripts with significant ADE
(FDR<0.05)
cumulative frequency
0.0
0.2
0.4
0.6
0.8
1.0
1.5
2.0
2.5
3.0
3.5
4.0
allelic fold change

Supp. Fig. S2

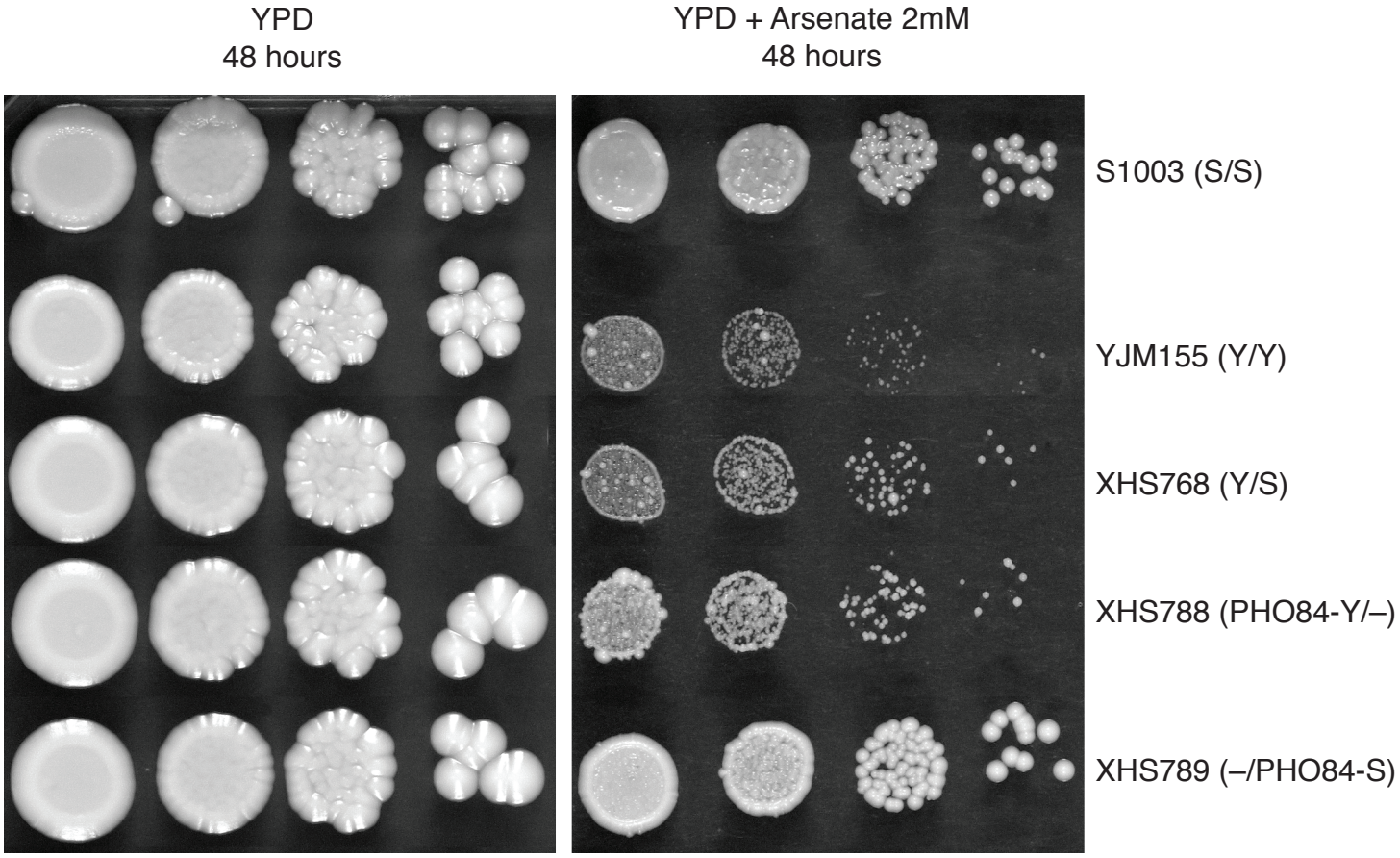

chromosome
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
200
400
600
800
gene rank
≤0.1 0.2 0.5 1
ADE coefficient
≤0.2 1 5 20 100%
polymorphism density
in promoter
h_S h_Y
3.5
0
1
0
4
0
2
0
1.2
0
UIP3
YAR028W
YAR029W
PRM9
MST28
185 186 187 188 189 190 191 192 193 kb

Supp. Fig. S3

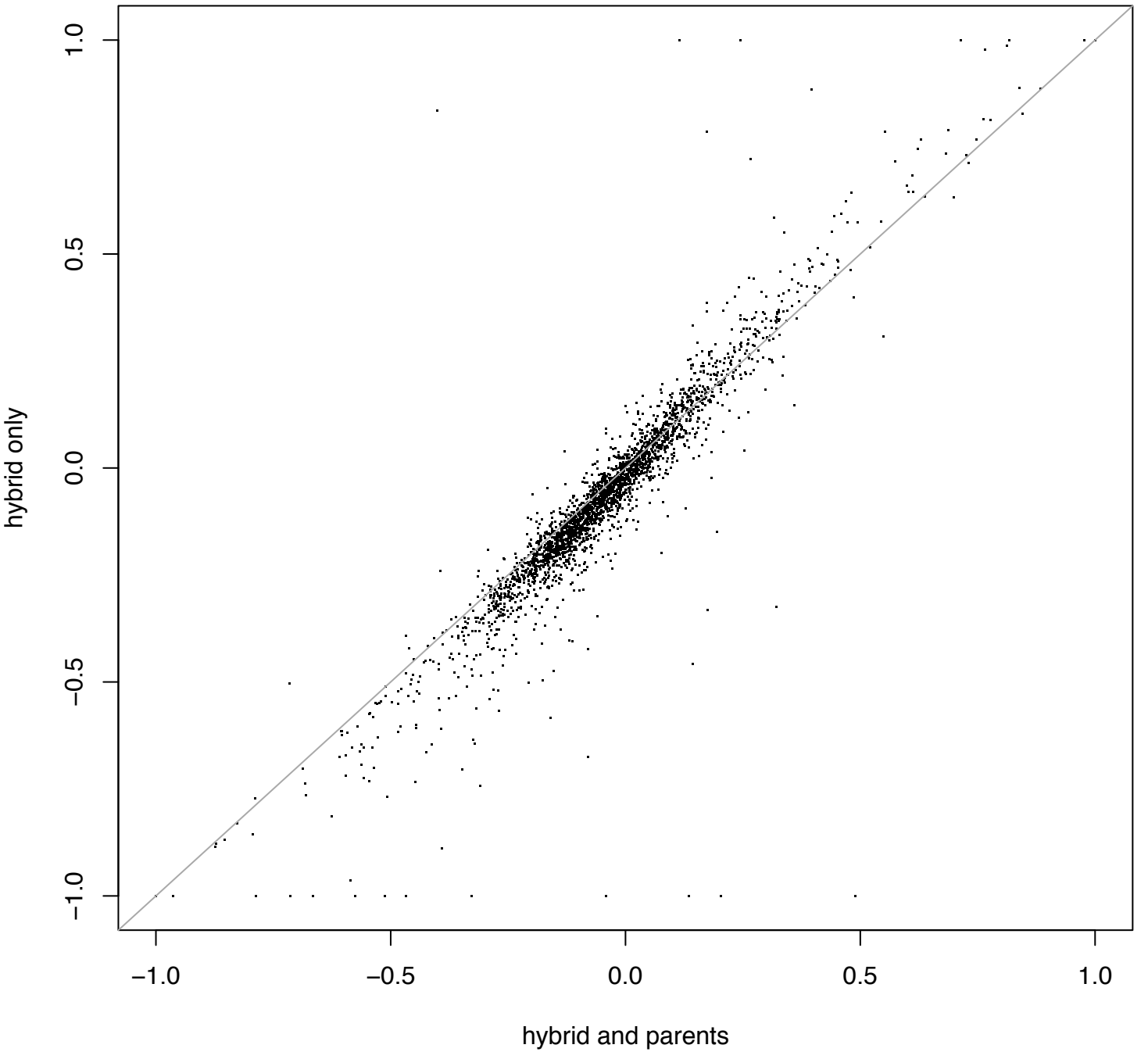

cumulative distribution
0.0
0.2
0.4
0.6
0.8
1.0
|C|/(|C|+|T|)


histogram
Frequency
0
10
20
30
40
0.0
0.2
0.4
0.6
0.8
1.0
|C|/(|C|+|T|)

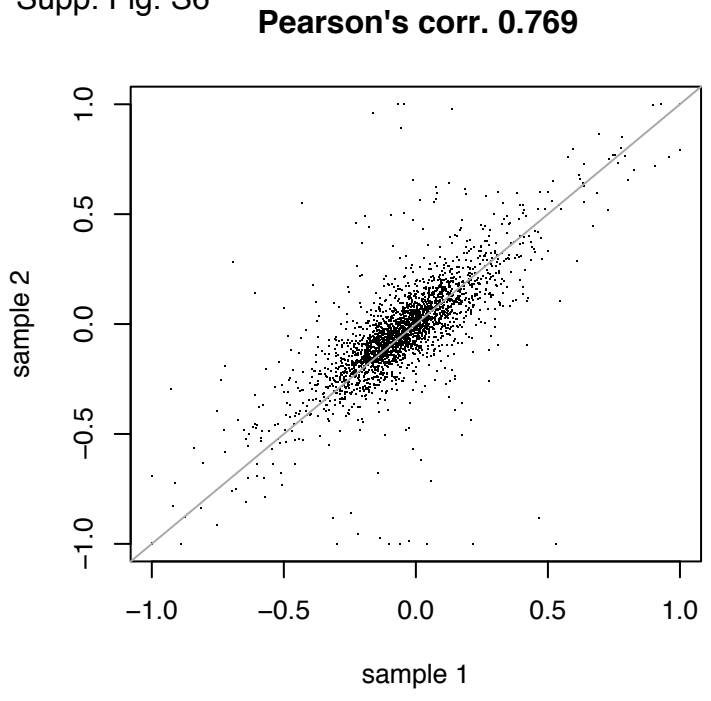
Suppl. Fig. S4

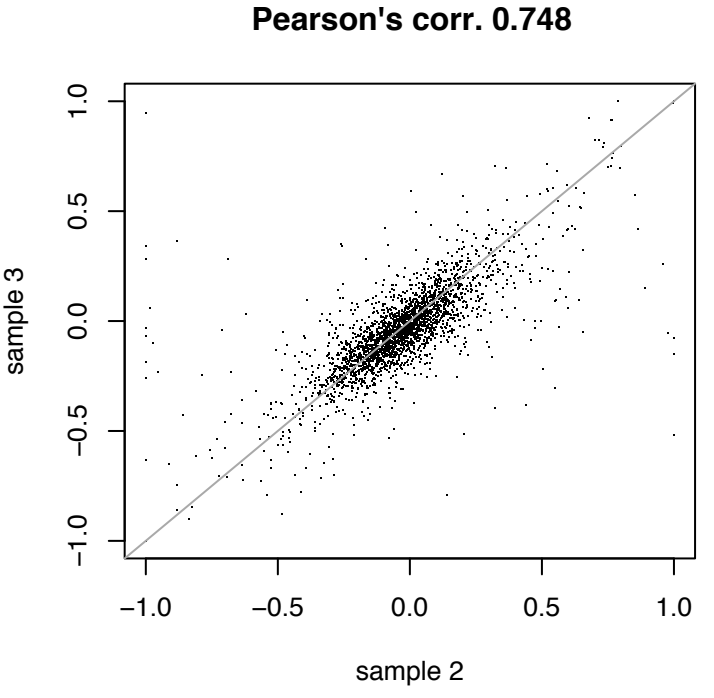

YPD
48 hours
YPD + Arsenate 2mM
48 hours
S1003 (S/S)
YJM155 (Y/Y)
XHS768 (Y/S)
XHS788 (PHO84-Y/–)
XHS789 (–/PHO84-S)

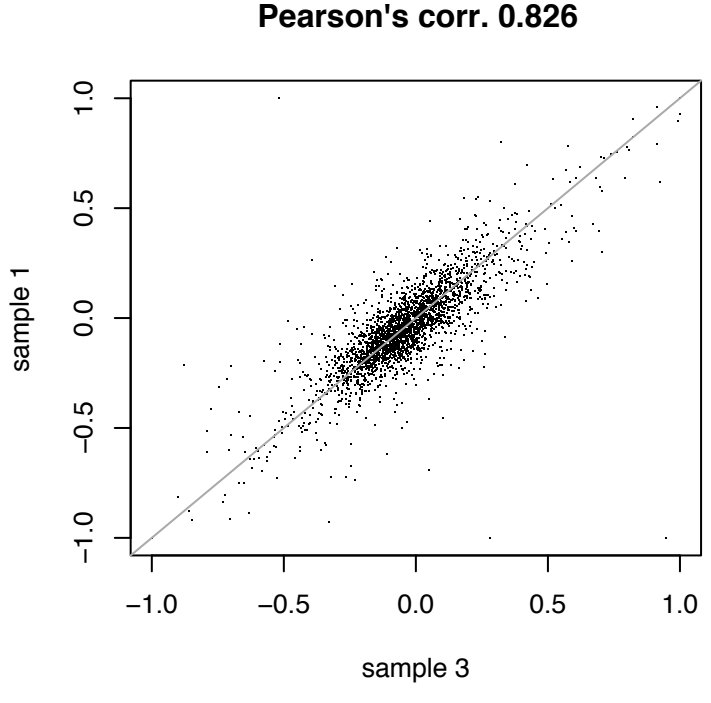
Supp. Fig. S5


Pearson's corr. 0.932
hybrid only
hybrid and parents
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0

Supp. Fig. S6


Pearson's corr. 0.769
sample 2
sample 1
−1.0
−0.5
0.0
0.5
1.0


Pearson's corr. 0.748
sample 3
sample 2
−1.0
−0.5
0.0
0.5
1.0


Pearson's corr. 0.826
sample 1
sample 3
−1.0
−0.5
0.0
0.5
1.0