molecular systems biology

Antisense expression increases gene expression variability and locus interdependency

Zhenyu Xu[1], Wu Wei[1], Julien Gagneur[1], Sandra Clauder-Münster, Miłosz Smolik, Wolfgang Huber and Lars M Steinmetz*

Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany
[1] These authors contributed equally to this work
* Corresponding author. Genome Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, Heidelberg 69117, Germany.
Tel.: + 49 6221 387 8389; Fax: + 49 6221 387 8518; E-mail: larsms@embl.de

Genome-wide transcription profiling has revealed extensive expression of non-coding RNAs antisense to genes, yet their functions, if any, remain to be understood. In this study, we perform a systematic analysis of sense–antisense expression in response to genetic and environmental changes in yeast. We find that antisense expression is associated with genes of larger expression variability. This is characterized by more 'switching off' at low levels of expression for genes with antisense compared to genes without, yet similar expression at maximal induction. By disrupting antisense transcription, we demonstrate that antisense expression confers an on-off switch on gene regulation for the *SUR7* gene. Consistent with this, genes that must respond in a switch-like manner, such as stress–response and environment-specific genes, are enriched for antisense expression. In addition, our data provide evidence that antisense expression initiated from bidirectional promoters enables the spreading of regulatory signals from one locus to neighbouring genes. These results indicate a general regulatory effect of antisense expression on sense genes and emphasize the importance of antisense-initiating regions downstream of genes in models of gene regulation.

Molecular Systems Biology **7**: 468; published online 15 February 2011; doi:10.1038/msb.2011.1
Subject Categories: functional genomics; RNA
Keywords: antisense regulation; gene expression; non-coding RNA; transcriptome

Introduction

Interleaved organization of transcription (Birney *et al*, 2007; Kapranov *et al*, 2007) is widespread in many genomes (David *et al*, 2006; He *et al*, 2008; Guell *et al*, 2009), raising the question of whether overlapping transcripts interact. Transcription antisense to coding genes represents ∼55% of the stable uncharacterized transcripts (SUTs) in yeast (Xu *et al*, 2009) and has been reported for a quarter of the protein coding genes in humans (He *et al*, 2008). For a handful of cases, regulatory roles of antisense expression on gene expression have been demonstrated. These involve a variety of mechanisms and effects—antisense can inhibit sense expression through transcriptional interference (Hongay *et al*, 2006) or histone modification (Camblong *et al*, 2007; Berretta *et al*, 2008; Houseley *et al*, 2008; Pinskaya *et al*, 2009). Such interactions can make gene activation faster (Uhler *et al*, 2007) or slower (Houseley *et al*, 2008). How widespread these regulatory effects are across the genome has so far, however, not been determined. We hypothesized that insight into the function of antisense expression could be gained by observing the behaviour of overlapping transcribed regions in response to short-term (environmental) and long-term (genetic) changes.

Results

We assessed genome-wide transcriptional response to genetic variation in *Saccharomyces cerevisiae* by profiling transcripts in 48 meiotic products (segregants) of an S288c/YJM789 hybrid strain (Figure 1A, Materials and methods and Supplementary Table S1). These segregants, among which genetic variation is shuffled by recombination, allow analysing transcriptome response to regulatory variation, keeping environment constant. We also analysed environmentally induced gene expression changes (keeping regulatory variation constant) across the main laboratory growth conditions of yeast (ethanol, galactose and glucose media, Figure 1B; Xu *et al*, 2009). Data were collected on high-resolution tiling arrays that measure strand-specific transcript levels genome-wide with 8-bp resolution (David *et al*,
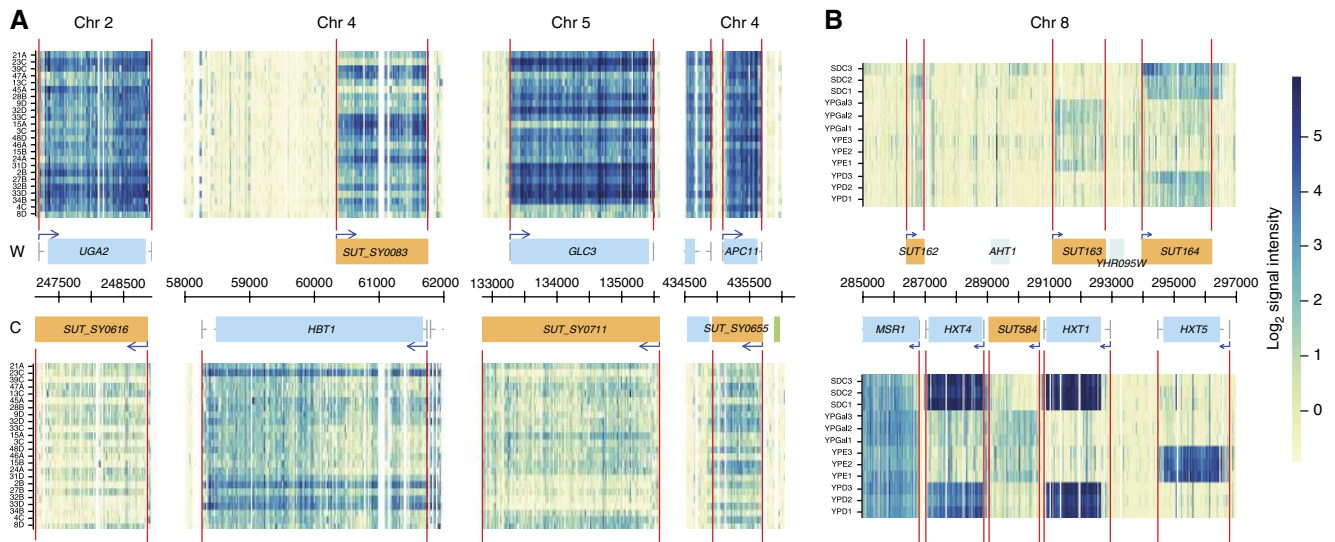
Figure 1 Genome-wide transcriptional response to genetic and environmental variations. (**A**) Four examples of sense–antisense transcript pairs (three anti-correlated and one positively correlated). Expression data are displayed along the chromosome (x axis) for the Watson (W, upper half) and the Crick (C, lower half) strands. Normalized signal intensities (higher in dark) are shown for 24 out of 48 segregants, randomly selected and ordered (y axis). Vertical red lines represent inferred transcript boundaries. Genome annotations are shown in the center: annotated ORFs (blue boxes), their mapped UTRs (dashed grey lines), SUTs (orange boxes) and transcript start sites (arrows). (**B**) Expression data along 15 kb of chromosome VIII across three replicates each for yeast grown in synthetic complete media with glucose (SDC), rich media with galactose (YPGal) and with ethanol (YPE); and three rows (summarizing nine replicates) for yeast grown in rich media with glucose (YPD; Xu et al, 2009).

2006). Observed transcripts (Materials and methods, Supplementary Tables S2 and S3) were classified as ORF-transcripts (ORF-Ts) when they mainly overlapped coding genes in the same orientation, and as SUTs when they mainly derived from unannotated genomic regions either antisense to genes or from intergenic regions (the term stable indicates that they are detected in wild-type cells as opposed to mutants of the exosome in accordance with our earlier definition; Xu et al (2009), Materials and methods). For legibility, we will use the terms ORF-T and gene interchangeably. In total, 613 (12%) of the ORF-Ts overlapped a SUT on the other strand (antisense transcript) in the segregant data set (Supplementary Table S4), and 474 (9%) in the environmental data set. The data set and expression plots for the whole genome are available in a searchable web database (http://steinmetzlab.embl.de/ASresponse).

As a control for our quantitation of sense and antisense transcript levels, we verified that the expression levels of transcripts in sense–antisense pairs were not significantly lower when estimated using the tiling array probes of the region of overlap than using the probes outside this region. These data show that any potential competition during hybridization between probes and antisense transcripts did not affect our level measurements (Supplementary Figure S1). Overall, ORF-Ts had much higher expression levels than antisense transcripts ($\sim$5.9-fold between medians, $P<2\times10^{-16}$, Wilcoxon rank-sum test). Furthermore, the larger number of genes with antisense transcripts found in the genetic data set is in agreement with our previous observation of more variation in SUT expression observed between the two parental strains than across changes in growth conditions (Xu et al, 2009).

Expression characteristics of genes with antisense transcripts

Notably, expression variation in response to our genetic and environmental changes was larger for genes with antisense transcripts than for genes without (Figure 2A and B, $P<2\times10^{-16}$ and $P=6\times10^{-12}$, respectively, Wilcoxon rank-sum test). Higher variability was also observed at evolutionary scales. Genes with antisense showed higher expression divergence across 5 yeast species (Tirosh et al, 2006; Figure 2C, $P=4\times10^{-12}$, one-tailed Wilcoxon rank-sum test here and in the following unless specified). Furthermore, larger variability between cells in a single population (i.e., cell-to-cell variability; Newman et al, 2006) was observed for protein abundance of genes with antisense (Figure 2D, $P=2\times10^{-4}$). All these observations on gene expression variability are reminiscent of properties of the TATA-box (Lopez-Maury et al, 2008), but remained significant when controlling for the presence of a TATA-box in gene promoters (Supplementary Figure S2, Materials and methods). These results indicate that, at different scales, antisense expression associates with a larger dynamic range of gene expression, and this association is independent of the increased expression variability known for TATA-containing genes (Lopez-Maury et al, 2008).

A larger dynamic range could be the result of lower minimal levels or higher maximal levels. Across the segregants, genes with antisense showed a notable depression at the lower end of their expression range, but almost no difference in the high range, compared with genes without antisense (Figure 2E). Similar observations on an independent strand-specific RNA-sequencing data set (Yassour et al, 2010) confirmed that these results are not an artefact due to saturation of the microarray
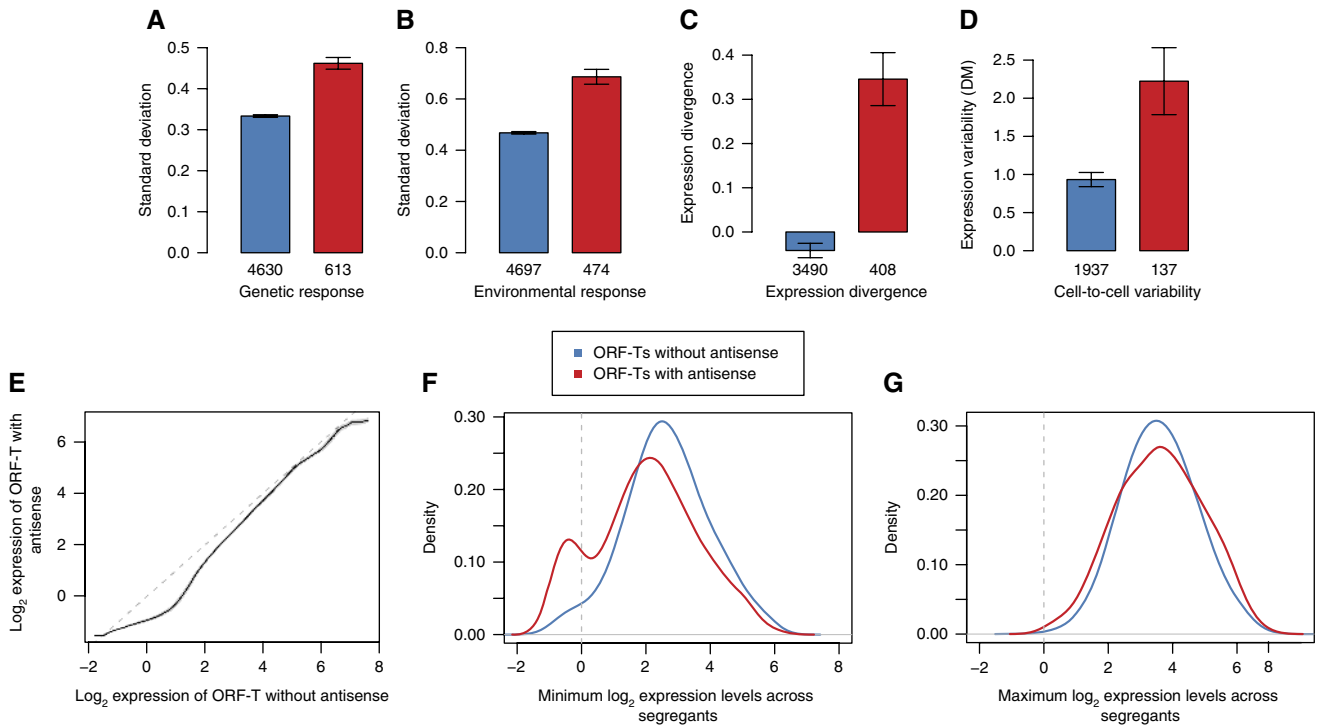
Figure 2 Antisense expression associates with larger gene expression variability. (**A–D**) Expression variability of ORF-Ts with or without antisense. (**A**) Mean of the $\log_2$ expression standard deviation across all segregants and (**B**) across environmental conditions, (**C**) mean of the expression divergence across five yeast species as provided by Tirosh *et al* (2006) and (**D**) mean of cell-to-cell protein expression variability (DM as provided by Newman *et al* (2006)) for ORF-Ts without antisense (blue) and ORF-Ts with antisense (red). Error bars indicate standard error of the mean. The number of ORF-Ts in each category is shown below the bar. (**E**) Q-Q plots of expression levels of ORF-Ts with and without antisense. The Q-Q plot compares two distributions by plotting for every quantile the two corresponding expression values against each other (expression level for ORF-Ts without antisense (*x* axis) against those for ORF-Ts with antisense (*y* axis)). Two data sets with the same distribution would align on the diagonal (grey line). ORF-Ts with antisense have lower expression levels (below the diagonal, bottom left) for the small quantiles, and similar expression levels for the large quantiles (top right). The shade around the curve (black line) represents bootstrap standard deviation (Materials and methods). (**F**, **G**) Smoothed histograms (distribution density estimates) of minimum (**F**) and maximum (**G**) expression values across the 48 segregants for ORF-Ts with antisense (red) and without (blue). Vertical line at *x*=0 indicates our threshold for calling a transcript expressed.

signals (Supplementary Figure S3 and Supplementary information). Specifically, genes with an antisense transcript had minimal levels significantly lower than genes without antisense (Figure 2F, $P < 2 \times 10^{-16}$). A large fraction of these per-gene minimum levels were consistent with no expression, that is, with microarray signal in the background range (18% for genes with antisense versus 5% for genes without, $P < 2 \times 10^{-16}$, one-sided Fisher test, see Materials and methods). In contrast, maximal expression levels were similar for both classes of genes (Figure 2G). Analogous behaviour was observed for the growth condition data (Supplementary information).

One interpretation of these observations is that antisense inhibits sense expression particularly at low levels of sense expression and that such inhibition is relaxed when sense expression is high. Another interpretation, although not in contradiction with the former, is that sense represses antisense expression and thus antisense is more easily expressed when sense expression is low—an interpretation that is perhaps in favour of a non-functional role of non-coding RNAs (Struhl, 2007). To find further support for a role (or lack thereof) of antisense expression in sense regulation, we examined the position of sense–antisense overlap.

The distributions of the $3'$ end positions of either sense or antisense transcripts peaked slightly beyond the trans-

cription start sites (TSS) of each other (98 ± 45 and 77 ± 19 bp, respectively, Figure 3A and Materials and methods). Thus, the typical arrangement of sense–antisense pairs involves an overlap of both promoter regions. In addition, variability of sense gene expression depended on the presence of this TSS overlap. Among genes with an antisense transcript, genes with an overlapped TSS showed larger expression variance across segregants and environmental conditions (Figure 3B, $P < 2 \times 10^{-16}$ and $P = 4 \times 10^{-5}$, respectively), larger expression divergence across species ($P = 4 \times 10^{-5}$) and larger cell-to-cell variability ($P = 0.09$; Supplementary Figure S4). Also, among the 282 genes of which the TSS was overlapped by an antisense transcript, 26% were switched off in at least one of the segregants, compared with only 11% of the 331 that were not overlapped at the TSS ($P = 1 \times 10^{-6}$, Fisher test). Hence, the effects on sense gene expression depended strongly on the overlap of the antisense transcript at the position of sense transcript initiation, favouring a model in which antisense expression affects sense expression.

Taken together, the genomic data support a model in which antisense expression induces a threshold-dependent or ultrasensitive (Koshland *et al*, 1982) on-off switch on sense gene regulation. This model proposes that in the absence of activation of the sense promoter, antisense expression switches off low, basal sense expression. In response to a
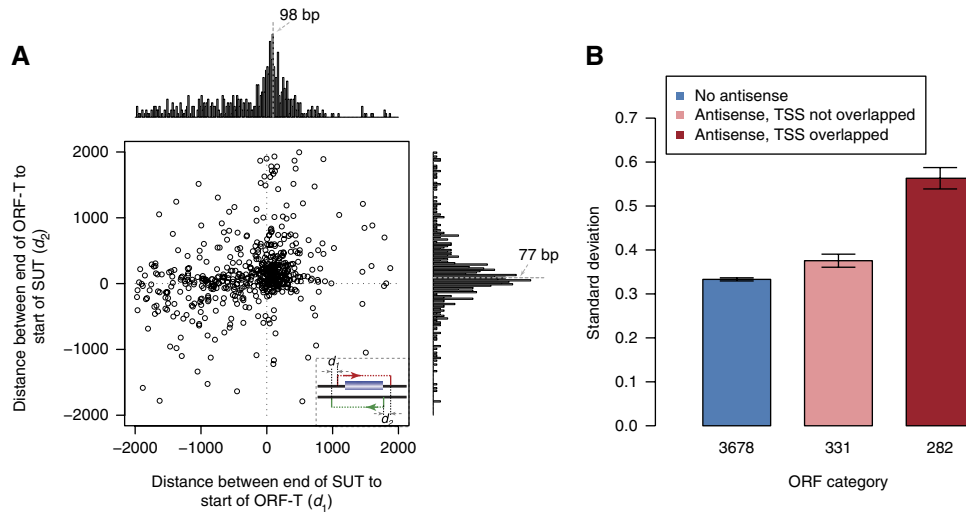
Figure 3 Sense expression variability depends on TSS overlap. (**A**) Overlap of sense–antisense pairs. d_1 (*x* axis) is the distance (base pairs) of the 3′ end of the antisense SUT to the TSS of the sense ORF-T and is positive if the 3′ end of the antisense extends beyond the sense TSS. d_2 (*y* axis) is the distance (base pairs) of the 3′ end of the sense ORF-T to the TSS of the antisense SUT. The panel shows the scatterplot of d_1 and d_2 and the histograms their marginal distributions, with dashed grey lines marking the modes of the distribution. (**B**) Variability across segregants for genes grouped by type of sense–antisense overlap. Mean of the expression standard deviation across the 48 segregants for ORF-Ts without antisense SUT (blue), with antisense SUT not overlapping the TSS (pink) and with antisense SUT overlapping the TSS (red). Error bars indicate standard error of the mean. The number of ORF-Ts in each category is shown below the bar.

sufficiently activating stimulus on the sense promoter, sense expression turns on and antisense inhibition is relaxed.

Validating the model

Elements of this model are supported by mechanistic studies. For example, experiments that block antisense expression have demonstrated an increase of sense expression for *PHO84* (Camblong *et al*, 2007), *IME4* (Hongay *et al*, 2006), *KCS1* (Nishizawa *et al*, 2008) and *GAL10* (Houseley *et al*, 2008; Pinskaya *et al*, 2009) showing that antisense expression represses sense expression. Analysis of data that we have published previously (Xu *et al*, 2009) reveals that in a mutant of *RRP6*, a component of the exosome machinery, in which the degradation of non-coding RNAs is impaired, 76 of 174 (44%) genes were repressed upon increased RNA levels of an antisense transcript that proceeded through their TSS (Materials and methods). This is significantly larger than the 25% of downregulated genes among those that lacked an antisense transcript (Fisher exact test, $P=3 \times 10^{-8}$), bolstering an argument for the inhibitory role of antisense in the regulation of multiple genes.

At high levels of gene expression, the effect of antisense appears reduced. The strength of a highly active gene promoter may override inhibitory effects exerted by antisense expression. In addition, reciprocal inhibition could explain the relaxation of inhibition at higher levels, where high sense expression inhibits antisense expression. Consistent with this, our sense–antisense overlap analysis showed an enrichment of sense transcripts overlapping the antisense promoter region. We also observed a significant enrichment for anti-correlation within sense–antisense pairs across conditions (Xu *et al*, 2009) and segregants, compared with random pairs of sense and antisense transcripts (Materials and methods, $P<2 \times 10^{-16}$, Supplementary Figure S5 and Figure 1A for

particular instances). Moreover, anti-correlation is stronger not only for pairs with overlap of the sense–TSS but also for those where only the antisense TSS is overlapped (compared with pairs with neither TSS overlapped, $P=2 \times 10^{-7}$ and 6×10^{-7}, respectively). Finally, an inhibitory function of sense on antisense expression has been demonstrated for *IME4*, where overexpression of the sense was shown to reduce antisense expression (Hongay *et al*, 2006). These data suggest that sense expression could display an inhibitory function on antisense expression.

So far, the threshold mediated on-off switch on gene regulation has not been directly tested. We tested this hypothesis on *SUR7*, a gene that has not been investigated for its antisense-mediated regulation before. *SUR7* exhibits both high and low levels of expression in two distinct conditions, and its antisense transcript (*SUT719*) can be disrupted without altering the sequence of the sense transcript.

In galactose media, *SUT719* is expressed antisense to *SUR7* and extends beyond the *SUR7* TSS (Figure 4A). *SUR7* is a gene of uncharacterized function and has been reported to be strongly downregulated in response to stimulation by α-factor pheromone (Roberts *et al*, 2000). We observed that *SUR7* is highly expressed in standard galactose media and is below detectable levels upon α-factor stimulation, whereas the antisense remains highly expressed in both conditions (Figure 4A). *SUT719* expression was disrupted without affecting the sequence of the *SUR7* RNA by deleting the Gal4 binding site of the *SUT719* promoter (Materials and methods). In agreement with our model, when disrupting antisense expression, expression of *SUR7* could be detected upon α-factor stimulation with a large increase compared with wild type (4.5-fold above background), whereas a moderate increase of expression was observed in the absence of α-factor (1.2-fold, Figure 4B). The possibility of a *GAL80*-mediated
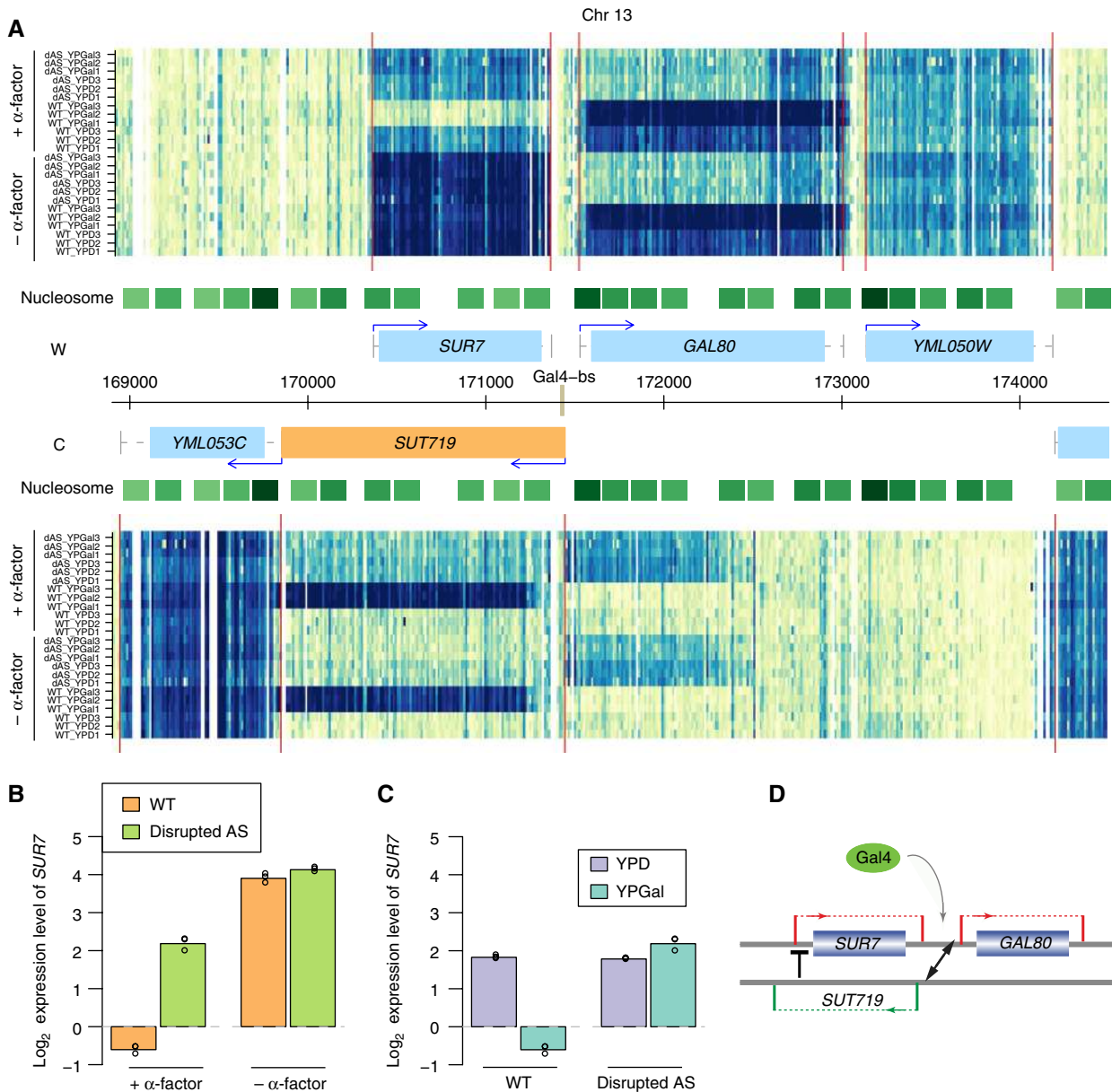
Figure 4 Antisense-mediated regulation of *SUR7*. (**A**) Expression maps around the *SUR7–GAL80* locus. Expression data is displayed along the chromosome (*x* axis) for the Watson (W, upper half) and the Crick (C, lower half) strands. Normalized signal intensities (higher in dark) are shown for the profiled samples (*y* axis, bottom-up): wild-type strain grown in YPD (WT_YPD1-3), wild-type strain grown in YPGal (WT_YPGal1-3), mutant strain with Gal4 binding site knocked out grown in YPD (dAS_YPD1-3), mutant strain with Gal4 binding site knocked out grown in YPGal (dAS_YPGal1-3), all for three biological replicates and without α-factor stimulation. The next 12 samples (bottom up) follow the same order, but with α-factor stimulation. Vertical red lines represent inferred transcript boundaries. Nucleosome positions (green tracks, darker for more significant scores; Mavrich *et al* (2008)) and genome annotations are shown in the center: annotated ORFs (blue boxes), their mapped UTRs (dashed grey lines), SUTs (orange boxes) and transcript start sites (arrows). (**B**) Mean of the log₂ expression levels of *SUR7* across three biological replicates for strains grown in YPGal, with (left) or without (right) α-factor stimulation for the wild-type strain (orange) or the antisense-disrupted strain (green). Values of individual samples are shown by open circles. (**C**) Mean of the log₂ expression levels of *SUR7* across three biological replicates for strains under α-factor stimulation grown in YPD (purple) or YPGal (turquoise), for the wild-type (left) or the antisense-disrupted strain (right). Values of individual samples are shown by open circles. (**D**) Model for spreading of regulatory signals on the *SUR7–GAL80* locus. *SUR7* and *GAL80* are two ORFs (blue boxes) in tandem configuration. The promoter of *GAL80* is bidirectional and initiates *SUT719*, a transcript antisense to *SUR7*. Regulatory signals in response to galactose media (Gal4 transcription factor, green oval) upregulate expression of *GAL80* as well as the antisense *SUT719*, thereby repressing *SUR7*.

feedback responsible for the upregulation of *SUR7* was ruled out by an experiment in which a drug-selectable cassette was inserted between the end of the *SUR7* transcript and the Gal4 binding site. Both experiments yielded the same conclusion on *SUR7* regulation, whereas the latter had no effect on *GAL80*

expression (Supplementary Figure 6, Materials and methods). These experiments demonstrate that antisense expression leads to threshold-dependent regulation on *SUR7* sense expression by specifically inhibiting sense expression when it is induced at low levels.

Spreading of regulatory signals

To obtain further support for antisense-mediated regulation, we examined neighbouring genes linked by non-coding RNAs. Specifically, we addressed the effect of bidirectional promoters on the regulation of tandem genes. Co-expression and functional correlation between adjacent genes have been observed (Cohen *et al*, 2000). Interleaved transcription (Kapranov *et al*, 2007) is a natural mechanism for building connections between adjacent genes. We have previously shown that antisense transcripts typically originate from bidirectional promoters shared with divergent genes (Xu *et al*, 2009). Combined with our current findings on antisense function, bidirectional transcription provides a possible mechanism for how the expression regulation of adjacent genes could be linked. In such an arrangement, exemplified by the *SUR7–GAL80* pair, a gene is under the control of its upstream promoter as well as a downstream promoter shared by an antisense and a downstream tandem gene (Figure 4A).

The antisense of *SUR7*, *SUT719*, initiates from the same nucleosome-depleted region as *GAL80*. *SUT719* responds to changes in sugar source, being expressed in galactose, but not in glucose, media. Its response is co-regulated with that of *GAL80*. In support of this, the deletion of the Gal4 binding site in the shared bidirectional promoter reduces the expression of both *GAL80* and *SUT719* (Figure 4A). In addition, we observed a complex pattern of expression of *SUR7* responding both to sugar source changes and to stimulation by α-factor pheromone. *SUR7* reaches high expression without α-factor in glucose and slightly lower levels in galactose. In the presence of α-factor, *SUR7* shows low levels of expression in glucose and is below array-detection levels in galactose (Figure 4A, wild type). Strikingly, when *SUT719* expression is disrupted, *SUR7* is no longer repressed after shifting from glucose to galactose media (Figure 4C), showing that the response of *SUR7* to galactose is mainly mediated by *SUT719* expression. Together, these results indicate that regulatory signals impinging on the *GAL80* promoter also affect the expression of the upstream gene, *SUR7*, by the regulated expression of an antisense transcript from the bidirectional promoter (Figure 4D).

The possibility that regulatory signals can spread across neighbouring loci by ncRNA expression, as shown here, stresses the importance of gene order and genomic organization (Kapranov *et al*, 2007). Because antisense expression can actually repress expression of sense genes, the relation is likely to be more complex than simple positive co-expression patterns within chromosomal domains as previously reported (Cohen *et al*, 2000; Ebisuya *et al*, 2008). Consistent with this, correlations between tandem gene pairs in the segregant data set are significantly smaller if the promoter of the downstream gene initiates a transcript antisense to the upstream gene, as in the *SUR7–GAL80* configuration (Supplementary Figure S7, median correlation 0.17 and 0.22, respectively, $P=6 \times 10^{-5}$, Wilcoxon rank test). These data support the hypothesis of antisense-mediated gene regulation between neighbouring loci.

Discussion

We have shown that antisense expression can induce threshold dependent gene regulation, by repressing sense expression

Figure 5 Model of antisense-mediated regulation. The sense gene (red dashed line, coding sequence as blue box) and the antisense SUT (green dashed line) typically extend beyond the TSS of each other. In the absence of sufficiently strong activating signals on the sense promoter, the antisense is expressed and inhibits sense expression (left T-shaped arrow). Upon activation of the sense promoter, the inhibition from the antisense SUT is relaxed, possibly through reciprocal inhibition of sense on antisense expression (right T-shaped arrow). This leads to threshold behaviour of gene regulation, where the gene is switched-off unless activation reaches a certain threshold.

particularly in the low range, whereas this inhibition is relaxed when sense expression is high. This enables an on-off switch on gene expression for antisense-containing genes, which leads to greater expression variability for antisense-containing genes. One simple possible mechanism for reduced inhibition at high levels is that reciprocal inhibition of sense on antisense relaxes the inhibition of antisense on sense expression (Figure 5). We have also shown that antisense expression initiated from bidirectional promoters can spread regulatory signals between neighbouring genes.

Our results underline the regulatory potential of the downstream region of a gene as a possible promoter of an antisense transcript. Hence, cloning the canonical region of a gene, defined by the promoter, the ORF and its UTRs, might not capture the whole local regulation if the cloned region does not include the possible antisense and its promoter. Similarly, computational predictions of *cis*-regulatory elements should include the 3′ region of genes.

Although sense–antisense pairs were enriched in anti-correlated expression patterns, we also observed a large proportion of positively or non-correlated expression pairs. Interestingly, all groups showed evidence of threshold-dependent ultrasensitive regulation (Supplementary Figure S8 and Material and methods). For example, for the 61 antisense transcripts with (approximately) constant levels of expression, the levels of their sense partners were reduced throughout the whole range (Supplementary Figure S8, green curve), which agrees with antisense-mediated inhibition, but with a weaker effect on the high range of sense expression. Consistent with this, higher variability was observed for all classes, but is more pronounced for the anti-correlated pairs. Overall, these observations support that the ultrasensitive regulation of gene expression induced by antisense is strengthened in the presence of, although not dependent on, anti-correlated sense–antisense expression behaviour. Furthermore, we note that the correlation coefficients are usually small; that is, a change of antisense expression is not always accompanied by a change of sense expression or vice versa suggesting that the main driving force of sense expression change is not antisense expression. Instead, the effect of antisense is more likely to be fine-tuning with a stronger effect on the low range than the high range of gene expression.

Across the segregants, 110 antisense-containing genes appeared switched off in at least one of the segregants ($\sim 2\%$ of all genes). Assuming that antisense transcripts overlapping sense TSS exert a regulatory role, the total number of genes that could be affected by antisense expression is 282 ($\sim 5.5\%$ of all genes). This covers about half of all antisense transcripts we detected. Nevertheless, due to the limited number of segregants and conditions that we profiled, the number of genes that are regulated by antisense could be larger.

It is not clear from the genomic data alone how, mechanistically, antisense expression exerts its role on sense expression. Our data cannot discriminate between a role of the antisense transcript or of the act of antisense transcription itself. Our analysis of sense–antisense overlap configuration supports an effect at the promoter level, but this could involve a variety of mechanisms. Silencing of the sense promoter through histone modifications induced by antisense transcript elongation has previously been suggested in the case of the *GAL10* gene (Houseley *et al*, 2008; Pinskaya *et al*, 2009). Also, pausing of RNA polymerases (transcribing a sense gene) on the promoter of an antisense transcript has been shown in *Escherichia coli* and suggested as a mediator of sense–antisense inhibition (Palmer *et al*, 2009).

Which gene categories benefit from antisense-mediated regulation? Condition-specific genes are more subject to transcriptional variability than housekeeping genes, as cells tune gene expression to activate cellular processes that respond to genetic and environmental changes. In our data, genes with antisense are depleted in essential genes (Materials and methods, $P=1 \times 10^{-11}$) yet enriched for environmental stress–response (Gasch *et al*, 2000; Materials and methods, $P=1 \times 10^{-6}$) and plasma membrane genes (enrichment screen for Gene Ontology categories, Materials and methods, $P=8 \times 10^{-5}$), which function in sensing and responding to external environmental signals. In addition, we showed increased expression variability between cells in a clonal population for genes with antisense expression. This variability could be advantageous within a population where cell-specific expression patterns enable some cells to be in an 'anticipatory' state for a sudden environmental change (Wykoff *et al*, 2007). Also along evolutionary time, a species may benefit from amplifying the regulatory impact of mutations for condition-specific genes, as opposed to growth-related genes. This would allow exploring transcriptional states beneficial to unforeseen changes (Lopez-Maury *et al*, 2008). Thus, antisense-mediated threshold regulation could provide a simple mechanism for short-term and long-term adaptation.

Notably, genes with antisense were more frequently switched off. Guaranteeing a gene to be off might be most important for genes whose qualitative presence (as opposed to quantitative abundance) can commit a cell into cell fate-altering transcriptional programmes. This is the case for *IME4*, whose expression has been shown to be determined by an antisense transcript that controls the entry into meiosis by repressing *IME4* in haploid cells (Hongay *et al*, 2006). Notably, transcription factors were enriched for genes with antisense expression (19% compared with 13% for other genes, $P=0.02$, Materials and methods). This supports the hypothesis that antisense-mediated switching off is important for controlling cell fate decisions.

If the observed enrichment of antisense expression among condition-specific genes is the result of natural selection, one can ask whether this is exerted by a positive selection for the presence of antisense in condition-specific genes, or by stronger negative selection against antisense expression for essential genes. A recent deep sequencing study (Yassour *et al*, 2010) identified about 1100 genes with antisense expression in rich media. In agreement with our observations, these are enriched for stress–response and condition-specific genes and tend to show opposite patterns of regulation than their antisense counterparts when profiled in the relevant conditions. Following a handful of cases across 5 yeast species, this study showed conservation of antisense expression and of anti-correlation. This provides further support to a conserved functional role of antisense expression and regulation.

As antisense expression is a universal feature of eukaryotic genomes (Kapranov *et al*, 2007), our results in yeast may generalize to higher eukaryotes. The non-coding RNA transcriptome is more complex in humans, nevertheless we observed that genes with an antisense show larger variance across five human cell lines (He *et al*, 2008; Materials and methods, Supplementary Figure S9, $P=8 \times 10^{-6}$). Thus, antisense-mediated threshold regulation of genes could be an ancient mechanism to enhance gene expression response to genetic and environmental variation.

Materials and methods

Data availability

Raw array data are available from ArrayExpress (http://www.ebi.ac.uk/arrayexpress) under accession numbers E-TABM-845 and E-TABM-1096. The data set and expression plots for the whole genome are available in a searchable web database (http://steinmetzlab.embl.de/ASresponse).

Strains and media

The segregant data set consists of 48 of the 184 segregants from Mancera *et al* (2008), derived from a cross of *S. cerevisiae* strains S96 (MATa ho:: lys5 gal2) and YJM789 (MATα ho::hisG lys2 gal2) (Supplementary Table S1). The segregants were grown to mid-exponential phase (OD$_{600}$$\sim 1.0$) in YPD (2% peptone, 1% yeast extract, 2% dextrose). Strains for the sense–antisense experiments of *SUR7* were constructed in a S288c bar1Δ background. The antisense transcript, *SUT719*, was disrupted by Gal4 binding site deletion or by KanMX cassette integration. For the binding site deletion, the Gal4 binding site between *SUR7* and *GAL80* (chromosome 13, 171422–171438) was excised applying the Cre/*loxP* recombination system, leaving two adjacent loxP sites with 72 bp instead of the binding site. For the cassette integration, a KanMx cassette was integrated at position 171376 on chromosome 13, encoded on the crick strand, which lies between the Gal4 binding site and the end of the *SUR7* transcript. To collect RNA for expression profiling, these strains were cultured either in YPD or YPGal (2% peptone, 1% yeast extract, 2% galactose) to mid-exponential phase (OD$_{600}$$\sim 0.5$–0.7) and split into two halves. To one half, α-factor (Zymo Research, cat. Y1001) was added to a final concentration of 1.5 nM, the other half served as a control. Then, strains were grown for two additional hours until they reached mid-exponential phase (OD$_{600}$$\sim 1.0$).

Sample preparation

All strains were collected from 100 ml of complex media at mid-exponential phase (OD$_{600}$$\sim 1.0$). Total RNA was isolated by a standard

hot phenol method. Poly(A) RNA was isolated from 2.5 mg of total RNA by using the Oligotex mRNA Maxi kit (Qiagen). Each sample of poly(A) RNA was treated with RNase-free DNaseI using Turbo DNA-free kit (Ambion). For first-strand cDNA synthesis, 9 μg of poly(A) RNA was mixed with 4.5 μg of random hexamers, 0.09 μg of oligo(dT) primer and incubated at 70°C for 10 min, then transferred on ice. The synthesis included 2000 units of SuperScript II Reverse Transcriptase, 50 mM Tris–HCl, 75 mM KCl, 3 mM MgCl$_2$, 0.01 M DTT, dNTP + dUTP mix (0.25 mM for dCTP, dATP and dGTP; 0.2 mM for dTTP and 0.05 mM for dUTP, Invitrogen), 6.25 μg/ml actinomycin-D in a total volume of 200 μl at 42°C for 1 h. Samples were then subjected to RNase treatment of 20 min at 37°C (30 units RNase H, Epicentre, 60 units of RNase Cocktail, Ambion). First-strand cDNA was purified using the MinElute PCR purification kit (Qiagen), and 4.5 μg were fragmented and labelled using the GeneChip WT Terminal labelling kit (Affymetrix) according to manufacturer's protocol. The labelled cDNA samples were denatured in a volume of 300 μl containing 50 pM control oligonucleotide B2 (Affymetrix) and Hybridization mix (GeneChip Hybridization, Wash and Stain kit, Affymetrix), of which 220 μl were hybridized per array (*S. cerevisiae* yeast tiling array, Affymetrix, PN 520055). Hybridizations were carried out at 45°C for 16 h with 60 r.p.m. rotation. The staining was carried out using the GeneChip Hybridization, Wash and Stain kit with fluidics protocol FS450_0001 in an Affymetrix fluidics station.

Genome sequence and annotation

The genome annotation (.gff file) for S288c was obtained from the Saccharomyces Genome Database on 19th August 2009. The sequence for YJM789 was obtained from Wei *et al* (2007) and aligned to the S288c genome using the procedure described therein (Wei *et al*, 2007).

Array data analysis

Arrays profiled for segregant strains were normalized with S288c genomic DNA as reference (Huber *et al*, 2006; Supplementary Table S1). Only the probes matching exactly and uniquely to both S288c and YJM789 genome and at the same alignment position were considered. The normalized data were jointly segmented based on the alignment between S288c and YJM789 using a segmentation algorithm (Huber *et al*, 2006), and the automatically identified segments were curated using a custom web interface (Xu *et al*, 2009). This defined the set of manually curated transcripts for the segregant data set (Supplementary Table S2 and Supplementary Information for transcript boundary accuracy assessment). For each transcript and each segregant, expression level was estimated by the midpoint of the shorth (shortest interval that covers half the values) of the normalized probe intensities lying within the transcript (Supplementary Table S3). The expression level cut-off for calling a transcript expressed was obtained using the same procedure as previously described (David *et al*, 2006). Briefly, the distribution of background microarray signal intensities was estimated from the intensities of the probes outside transcript boundaries. The cut-off for an intensity to be significantly above background was then set at an estimated FDR of 0.05. For the growth condition data set, transcript boundaries and levels were taken from Xu *et al* (2009), restricting to SUTs and ORF-Ts expressed in the media YPD, YPGal, YPE and SDC. Every reported transcript was expressed in at least one condition of the environmental data set or in one segregant of the segregant data set.

Transcript categorization

The manually curated transcripts were overlapped with genome annotation features and classified as (1) SUT, if they did not overlap with existing annotation; (2) ORF-T, if they overlapped with a verified or uncharacterized ORF; (3) other, otherwise. Antisense SUTs were defined as SUTs that overlapped with an ORF-T. Sense ORF-Ts were defined as ORF-Ts with at least one overlapping antisense SUT (Supplementary Tables S2 and S4). For transcripts on the condition data, we used the categorization of Xu *et al* (2009).

Expression divergence across yeast species

We used expression divergence of yeast ORFs as provided by Tirosh *et al* (2006). All analyses on expression divergence were done using the transcript annotation defined on the segregant data set.

Cell-to-cell variability

We used cell-to-cell protein expression variability as measured by the DM coefficient provided by Newman *et al* (2006) for the YPD condition. All analyses of cell-to-cell protein expression variability were done using the transcript annotation defined on the condition data set.

Effects of antisense controlling for the presence of a TATA-box

The list of genes with a TATA-box was obtained from Basehoar *et al* (2004). We modelled each measure (gene expression variance across conditions and segregants, expression divergence across yeast species and cell-to-cell variability) with an ANOVA model as the sum of an effect contributed by antisense if present and an effect contributed by the TATA-box if present (linear model with no interaction). Statistical significance of the effect of the antisense being not 0 was tested by the *t*-test.

Overlap configuration

All analyses on overlap configurations were performed using the transcript annotation defined on the segregant data set. For each sense–antisense pair, the distance of the 3′ end of the sense ORF transcript relative to the TSS of the antisense-SUT (positive if the 3′ UTR extends beyond the TSS, negative otherwise) was computed, and similarly for the 3′ end of the antisense-SUT. The peak of the distribution of each of these two values (Figure 3A, upper and rightmost panels) was estimated by the midpoint of the shorth (shortest interval containing half of the values). Standard deviations for the peak position were computed by bootstrapping the cases a 1000 times using the R package 'boot'.

Enrichment for anti-correlation

Random sense–antisense pairings were generated by reshuffling the antisense transcripts, keeping the sense transcripts fixed. Enrichment or depletion for anti-correlation (negative Pearson's correlation coefficients) in the actual data set compared with the random was tested using the two-sided Fisher test.

Differential expression between rrp6 mutant and wild type

Statistical significance for differential expression between rrp6 mutant and wild type was tested using limma (Smyth *et al*, 2003) and followed by Storey's q-value correction (Storey and Tibshirani, 2003) with a false discovery rate of 0.05. The sense–antisense pairs here are defined using both SUT and CUT annotation as described before (Xu *et al*, 2009).

Tandem gene pairs

Genes in a tandem pairs were defined as consecutive ORF-Ts on the same strand separated by < 3 kb.

Ultrasensitivity behaviour as a function of expression correlation between sense and antisense

Sense–antisense pairs were tested for significant positive or negative correlation (Pearson's correlation coefficient, false discovery rate

<0.05 using Storey's method; Storey and Tibshirani, 2003). We defined as 'variable' the transcripts with a standard deviation in the top 50%. ORF-Ts with antisense were split into four groups: the anti-correlated ones (165 cases, 27%), the positively correlated ones (108, 18%, e.g., SUT_SY0655), those with no significant correlation but a variable antisense (279, 45%, e.g., SUT_SY0338) and those with no significant correlation and a not variable antisense (61, 10%, e.g., SUT_SY0117).

Gene ontology analysis

The *P*-value of the Fisher test for every cellular component term of the Gene Ontology (Ashburner *et al*, 2000), followed by a Holm correction for multiple testing (family-wise error rate), was computed with the software Ontologizer (Bauer *et al*, 2008) using the combined set of ORFs with an antisense in either the segregant data set or the condition data set. The same was carried out for the biological process terms and the molecular function terms. We used Gene Ontology annotations obtained from the Saccharomyces Genome Database on 22 October 2009.

Other gene categories

We used environmental stress-induced ORFs as provided by Gasch *et al* (2000). ORFs were classified as transcription factors if in the set of transcription factors defined by Harbison *et al* (2004). Essential ORFs were obtained from the Stanford yeast deletion project (http://www-sequence.stanford.edu/group/yeast_deletion_project/). Enrichment analysis was performed using the combined set of ORFs with an antisense in either the segregant data set or the condition data set.

Variability across human cell lines

For strand-specific RNA-seq in human cell lines, we used the counts of distinct reads per gene locus provided by He *et al* (2008) and Ensembl transcript annotation as of March 2008. For each experiment and for each gene, an approximately variance-stabilized measure of expression level was obtained as the square root of the normalized read number, itself defined as the number of reads divided by the gene length and by the library size (total number of unique reads in the experiment). Expression for each cell line was obtained as the median across technical replicates. The antisense expression value was similarly computed for the same gene boundaries using the reads mapping to the opposite strand. We called an antisense detected if its normalized read number was greater than 2 divided by the median gene length and the median library size. (Two antisense reads is the cut-off used in the original study for reporting antisense expression). To control for remaining expression level effect on the variance (not removed by the approximate variance stabilization), genes with at least one mapped read were grouped by expression level into 10 bins of equal size (Supplementary Figure S9). We then modelled expression variance as the sum of an effect contributed by antisense if present and an effect contributed by the expression bin (linear model with no interaction). Statistical significance of the effect of the antisense being not 0 was tested by the *t*-test.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

Author contributions: L.M.S., Z.X., W.W., J.G. designed the research; Z.X., W.W., J.G. analysed the data with the help of L.M.S. and W.H.; M.S., Z.X. and S.C. performed the array hybridizations; Z.X. and J.G. annotated the transcripts; L.M.S., J.G. and W.H. supervised the research; J.G., Z.X., L.M.S., W.W. and W.H. wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25:** 25–29

Basehoar AD, Zanton SJ, Pugh BF (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116:** 699–709

Bauer S, Grossmann S, Vingron M, Robinson PN (2008) Ontologizer 2.0–a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* **24:** 1650–1651

Berretta J, Pinskaya M, Morillon A (2008) A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev* **22:** 615–626

Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P *et al* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799–816

Camblong J, Iglesias N, Fickentscher C, Dieppois G, Stutz F (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* **131:** 706–717

Cohen BA, Mitra RD, Hughes JD, Church GM (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet* **26:** 183–186

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM (2006) A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci USA* **103:** 5320–5325

Ebisuya M, Yamamoto T, Nakajima M, Nishida E (2008) Ripples from neighbouring transcription. *Nat Cell Biol* **10:** 1106–1113

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11:** 4241–4257

Guell M, van Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kuhner S, Rode M, Suyama M, Schmidt S, Gavin AC, Bork P, Serrano L (2009) Transcriptome complexity in a genome-reduced bacterium. *Science* **326:** 1268–1271

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* **431:** 99–104

He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW (2008) The antisense transcriptomes of human cells. *Science* **322:** 1855–1857

Hongay CF, Grisafi PL, Galitski T, Fink GR (2006) Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell* **127:** 735–745

Houseley J, Rubbi L, Grunstein M, Tollervey D, Vogelauer M (2008) A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Mol Cell* **32:** 685–695

Huber W, Toedling J, Steinmetz LM (2006) Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics* **22:** 1963–1970

Kapranov P, Willingham AT, Gingeras TR (2007) Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet* **8:** 413–423

Koshland Jr DE., Goldbeter A, Stock JB (1982) Amplification and adaptation in regulatory and sensory systems. *Science* **217:** 220–225

Lopez-Maury L, Marguerat S, Bahler J (2008) Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* **9:** 583–593

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

Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I, Pugh BF (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* **18:** 1073–1083

Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441:** 840–846

Nishizawa M, Komai T, Katou Y, Shirahige K, Ito T, Toh EA (2008) Nutrient-regulated antisense and intragenic RNAs modulate a signal transduction pathway in yeast. *PLoS Biol* **6:** 2817–2830

Palmer AC, Ahlgren-Berg A, Egan JB, Dodd IB, Shearwin KE (2009) Potent transcriptional interference by pausing of RNA polymerases over a downstream promoter. *Mol Cell* **34:** 545–555

Pinskaya M, Gourvennec S, Morillon A (2009) H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. *EMBO J* **28:** 1697–1707

Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, Bennett HA, He YD, Dai H, Walker WL, Hughes TR, Tyers M, Boone C, Friend SH (2000) Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* **287:** 873–880

Smyth GK, Yang YH, Speed T (2003) Statistical issues in cDNA microarray data analysis. *Methods Mol Biol* **224:** 111–136

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* **100:** 9440–9445

Struhl K (2007) Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* **14:** 103–105

Tirosh I, Weinberger A, Carmi M, Barkai N (2006) A genetic signature of interspecies variations in gene expression. *Nat Genet* **38:** 830–834

Uhler JP, Hertel C, Svejstrup JQ (2007) A role for noncoding transcription in activation of the yeast PHO5 gene. *Proc Natl Acad Sci USA* **104:** 8011–8016

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

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457:** 1033–1037

Yassour M, Pfiffner J, Levin JZ, Adiconis X, Gnirke A, Nusbaum C, Thompson DA, Friedman N, Regev A (2010) Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome Biol* **11:** R87