

Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D

Fabiana Perocchi, Zhenyu Xu, Sandra Clauder-Münster and Lars M. Steinmetz*

European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Received June 6, 2007; Revised and Accepted August 20, 2007

ABSTRACT

Recent transcription profiling studies have revealed an unanticipatedly large proportion of antisense transcription across eukaryotic and bacterial genomes. However, the extent and significance of antisense transcripts is controversial partly because experimental artifacts are suspected. Here, we present a method to generate clean genome-wide transcriptome profiles, using actinomycin D (ActD) during reverse transcription. We show that antisense artifacts appear to be triggered by spurious synthesis of second-strand cDNA during reverse transcription reactions. Strand-specific hybridization signals obtained from *Saccharomyces cerevisiae* tiling arrays were compared between samples prepared with and without ActD. Use of ActD removed about half of the detectable antisense transcripts, consistent with their being artifacts, while sense expression levels and about 200 antisense transcripts were not affected. Our findings thus facilitate a more accurate assessment of the extent and position of antisense transcription, towards a better understanding of its role in cells.

INTRODUCTION

Non-protein-coding RNAs are postulated to represent an important development in the genetic operating system of higher eukaryotes, where the ratio of non-coding to protein-coding DNA rises as a function of phenotypic complexity (1). Among them, antisense transcripts have been implicated in transcription, processing, stability, transport and translation of their corresponding complementary RNAs (2,3). Their role appears to be regulatory (4) and has been implicated in human genetic diseases (5).

Ongoing large-scale studies using cDNA sequencing and genome tiling arrays have reported a widespread

occurrence of antisense transcripts in the genomes of simple organisms (6–9), as well as higher eukaryotes (10–12). The unexpected extent of antisense transcription over the genomes, as well as the observation of mirrored transcription for several sense–antisense pairs have raised the concern that some of these signals can be due to experimental artifacts (13). However, these concerns have remained speculative and no comprehensive independent experimental validation has been performed yet.

One possibility is that artifacts arise during target preparation by reverse transcription, confounding the interpretation of array data when both genomic strands are interrogated (Figure 1). Indeed, reverse transcriptase has a tendency to generate spurious second-strand cDNA based on its DNA-dependent DNA polymerase activity (14). Among several possibilities, priming of first-strand cDNA could occur by either a hairpin loop at its 3' end or by re-priming, from either RNA fragments formed by degradation of RNA templates or from primers used for the first-strand synthesis (15,16).

ActD is a known inhibitor of transcription, as well as DNA replication. Less well known is the fact that ActD can also selectively prevent second-strand cDNA synthesis during reverse transcription (17) due to the specific inhibition of DNA-dependent, but not RNA-dependent, DNA-synthesis (18) (Figure 1). This inhibitory effect of ActD appears related to its ability to bind deoxyguanosine residues on single- and double-stranded DNA, possibly preventing either the annealing of DNA during priming or the elongation of DNA-dependent DNA-synthesis (19).

Here we show that a standard reverse transcription reaction, used for the synthesis of DNA complementary to RNA templates, is a major source of artifactual antisense transcripts. We demonstrate this using an array that contains 6.5 million probes tiling both strands of the full genomic sequence of *S. cerevisiae* at an average of 8 nt intervals on each strand and a 4 nt offset of the tile between strands (7). We propose that ActD can be used in cDNA preparations during reverse transcription reactions for hybridization to microarrays, to generate clean transcriptome profiles. We show that antisense artifacts

*To whom correspondence should be addressed. Tel: +49 6221 387-389; Fax: +49 6221 387-518; Email: larsms@embl.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

are eliminated in samples prepared with ActD. These findings have implications for other genome-wide approaches involving reverse transcription.

MATERIALS AND METHODS

Sample preparation

S1003 (MATa/ α gal2/gal2 lys5/lys5), an S288c background strain, was grown in 100 mL of rich medium (2% Difco peptone, 1% yeast extract, 2% dextrose) to mid-exponential phase (OD₆₀₀ ~ 1.0). Total RNA was isolated by a standard hot phenol method. Poly(A) RNA was isolated from 1 mg of total RNA by using the Oligotex mRNA Midi 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 μ g of poly(A) RNA was mixed with 4.5 μ g of random hexamers and incubated at 70°C for 10 min, then transferred on ice. The synthesis included 2000 U of SuperScript II Reverse Transcriptase, 50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂, 0.01 M DTT, 0.25 mM dNTPs mix (Invitrogen) in a total volume of 200 μ l at 42°C for 1 h. Samples prepared with ActD were treated the same, except that the drug was added after the denaturing step at 70°C to a final concentration of 6 μ g/ml. Samples were then subjected to RNase treatment of 20 min at 37°C (30 U RNase H, Epicentre, 60 U of RNase Cocktail, Ambion). First-strand cDNA was purified by standard phenol extraction (1.5 ml Phase lock gel, Eppendorf) and ethanol precipitation. The sample was dissolved in DEPC water and digested by using 0.1 U DNase I (Invitrogen) in 1 \times One-Phor-All buffer (Amersham Pharmacia) and 1.5 mM CoCl₂ (Roche) solution at 37°C, to yield fragments of 50–100 bp in size. Each sample was 3' end-labeled with 0.07 mM Biotin-N6-ddATP (Enzo Life Sciences) using 400 U of Terminal Transferase (Roche) for 2 h at 37°C.

For genomic DNA hybridizations, the diploid yeast strain (S1003) was grown in rich media to saturation and genomic DNA was purified by using the Genomic tip protocol (Genomic DNA kit, Qiagen) including RNase A treatment. 10 μ g of DNA was fragmented with 0.2 U of DNaseI (Invitrogen) for 5 min at 37°C and labeled as described for cDNA.

Hybridization to arrays

For genomic DNA and cDNA from poly(A)-RNA, labeled samples of 10 and 4.5 μ g respectively were denatured in a solution containing 100 mM Mes, 1 M [Na⁺], 20 mM EDTA, 0.01% Tween-20, 50 pM control oligonucleotide B2 (Affymetrix), 0.1 mg/ml herring sperm DNA, and 0.5 mg/ml BSA in a total volume of 300 μ l, from which 220 μ l were hybridized per array. Hybridizations were carried out at 45°C for 16 h with 60 r.p.m. rotation. Our analysis is based on five replicate sample hybridizations of poly(A) RNA without ActD, three replicate sample hybridizations of poly(A) RNA with ActD and three of genomic DNA.

Data preprocessing

Arrays were normalized using the DNA reference normalization (20) as implemented in the tiling array package of Bioconductor (21). Only the probes that are unique and exactly match to the S288c genome were further considered. All arrays were then segmented together using the segmentation algorithm (20) as implemented in the tiling array package of Bioconductor.

Calculation of background threshold

To estimate a common background threshold for both ORFs and segment comparisons, we used the segments that do not overlap with any annotated, transcribed features. Only data from the hybridizations without ActD were used to determine the background threshold. The values of the segments were first sorted and the midpoint of the shorth (the shortest interval that covers half of the values) of the first 99.9% of the data was used to fit a normal distribution to determine the threshold at which the false discovery rate is 0.1%, as performed in David *et al.* (7). This estimated background threshold was used to rescale the probe intensity data, setting the background threshold to zero for all probes.

Calculation of sense and antisense strand signals for ORFs

To compare the signal for both sense and antisense strands of ORFs between hybridizations performed with and without ActD, the probes were mapped to the annotated features (davidTiling Package in Bioconductor, gff data) and the same boundaries opposite to the annotated features. The signals of the probes were then averaged among the replicates and the median value of all the probes that mapped to the feature was used to represent the signal of that feature. An ORF or its antisense region was considered as transcribed if the expression signal was above 0 (Supplementary Table 1).

Categorization of segments

After segmentation, the probe signals within the boundaries of the segments were averaged among replicates and the median value of the averages was used to represent the expression signal of the segment. Segments were then assigned to different categories depending on how they overlapped with annotated features, as described in David *et al.* (7). The segments in reaction with and without ActD were categorized separately and classified as annotated ORF, annotated ncRNA, novel isolated, novel antisense, excluded, dubious gene. The excluded segments were the segments that have more than 50% non-unique probes. The novel isolated segments were the segments that do not overlap with any transcribed features on either strand. The novel antisense filtered segments are the segments that do not overlap with any transcribed features but overlap with a transcribed feature on the opposite strand. Antisense and isolated segments had to further fulfill the following two criteria: (i) have a length longer than 48 bp and (ii) be flanked by segments with reduced signal on both sides. If not fulfilling both criteria, they were

classified as unassigned antisense or unassigned isolated, respectively (Supplementary Table 2).

Comparison to cDNA sequencing

We compared the stringent set of antisense detected by cDNA sequencing of strain S288c (8) to our data. Comparisons were based on gene names of the sense transcripts. Of 7562 features (e.g. genes) (davidTiling Package in Bioconductor, gff data), 218 features show antisense sequences by cDNA sequencing, 229 features have antisense segments in data generated with the addition of ActD, and 589 features have antisense segments in data generated in the absence of ActD. We found 26 features in common between the cDNA sequencing dataset and the ActD+ array dataset, and 36 features in common with the ActD- array dataset. A Fisher's exact test was performed on the numbers above.

Comparison to computational filtering

The antisense segments were flagged by the filter if any median signal of a 100 bp sliding region on the opposite strand of the antisense segment was larger than the signal of the antisense segment itself, as applied previously (7). Two comparisons were performed between the two datasets (with and without ActD). In a first comparison the filter was applied to both datasets generated with and without ActD. In a second comparison the dataset on antisense segments detected in the presence of ActD was compared with the dataset on antisense segments detected in the absence of ActD but after filtering. The differences were manually checked. A list of these segments is available in Supplementary Table 4 and the expression profiles are available online on our website (www.ebi.ac.uk/huber-srv/actinomycinD).

Strand-specific RT-PCR

Total RNA was isolated from strain S1003 (MATa/ α gal2/gal2 lys5/lys5) as described above and was treated with RNase-free DNase I (TURBO DNase, Ambion) for 25 min at 37°C. DNA-free RNA samples were split into +RT and -RT reactions. In +RT reactions, three replicates for each antisense RNA were performed in the presence of reverse transcriptase. In each reaction 2 pmol of primer, designed to hybridize to the antisense transcript (Supplementary Table 5), was added and annealed to 2 μ g of total RNA by 5 min incubation at 65°C, followed by cooling on ice. A strand-specific primer targeting *ACT1* mRNA was always included as an internal control. First-strand cDNA was synthesized by using 200 U of SuperScript II RT (Invitrogen) for 50 min at 42°C. The enzyme was heat inactivated at 70°C for 15 min. RNA complementary to the cDNA was removed by *Escherichia coli* RNase H (10 U, Epicentre) and remaining RNAs were digested with 20 U of RNase Cocktail (Ambion). The -RT reactions were treated identically to the +RT reactions, with the exception that reverse transcriptase was omitted.

PCR was performed for the antisense transcripts of interest and for *ACT1* individually by using 1 μ l of the RT reaction as a template, two gene specific primers (250 nM

each) (Supplementary Table 5), 200 μ M dNTP and 1 U of Ampli Taq Gold (Applied Biosystems). As a positive control PCR reactions were performed with genomic DNA as a template. The following thermal profile was used for amplification: incubation at 95°C for 10 min followed by 25 cycles of three-step amplification at 95°C for 30 s, 60°C or 52°C (dependent on primer) for 30 s and 72°C for 30 s.

RESULTS AND DISCUSSION

To assess the extent of artifacts generated by reverse transcription, we performed five replicates (biological) of standard reverse transcription reactions lacking ActD (ActD-) on RNA samples from yeast grown in rich media and analyzed cDNA targets by hybridization to tiling arrays. For sense transcripts, hybridization signals were concordant between all replicates. However, two types of signals were registered on the antisense strands opposite to expressed genomic regions. In one class, the signal intensities corresponded proportionally to the intensities of the sense counterparts (Figure 2A, upper panel); furthermore, substantial variability existed across replicates (Figure 2B and C, upper panel). In the other class, antisense signals were highly reproducible across all replicates and did not correlate with sense strand expression levels (Figure 2D, upper panel). These differences suggested that the first class of antisense signals might be artifacts, potentially triggered by spurious second-strand synthesis during reverse transcription as proposed by the models (Figure 1). Besides growth in rich media, we have seen the same pattern for several other conditions (data not shown). We postulated that the putative artifacts can be resolved using ActD. Indeed, three replicates of reverse transcription reactions performed with ActD (ActD+) resulted in expression signals below background for the first class, but not for the second class of antisense regions (Figure 2A-D, lower panels). In addition to the examples shown in Figure 2, profiles for all genomic regions are available online (<http://www.ebi.ac.uk/huber-srv/actinomycinD>).

A genome-wide comparison of expression signals with or without ActD over all coding genes [as defined by ORFs in the *Saccharomyces* Genome Database (SGD, <http://www.yeastgenome.org>)] and their opposite regions supported the interpretation that ActD reduces artifactual antisense signals. The number of antisense regions detected as expressed above background declined from 1046 in ActD- to 325 in ActD+. Moreover, only 25% (260/1046) of the cases observed upon standard first-strand cDNA synthesis are still detectable in ActD+ (Supplementary Table 1). Furthermore, consistent with its specific role in second-strand synthesis, ActD did not affect the number of sense transcripts detected above background. Out of 5703 coding-genes (ORFs), the majority is expressed above background in both conditions: 5214 ORFs in ActD+ and 5186 ORFs in ActD-; in addition, there is a nearly complete overlap between the two gene lists (5172 ORFs) (Supplementary Table 1).

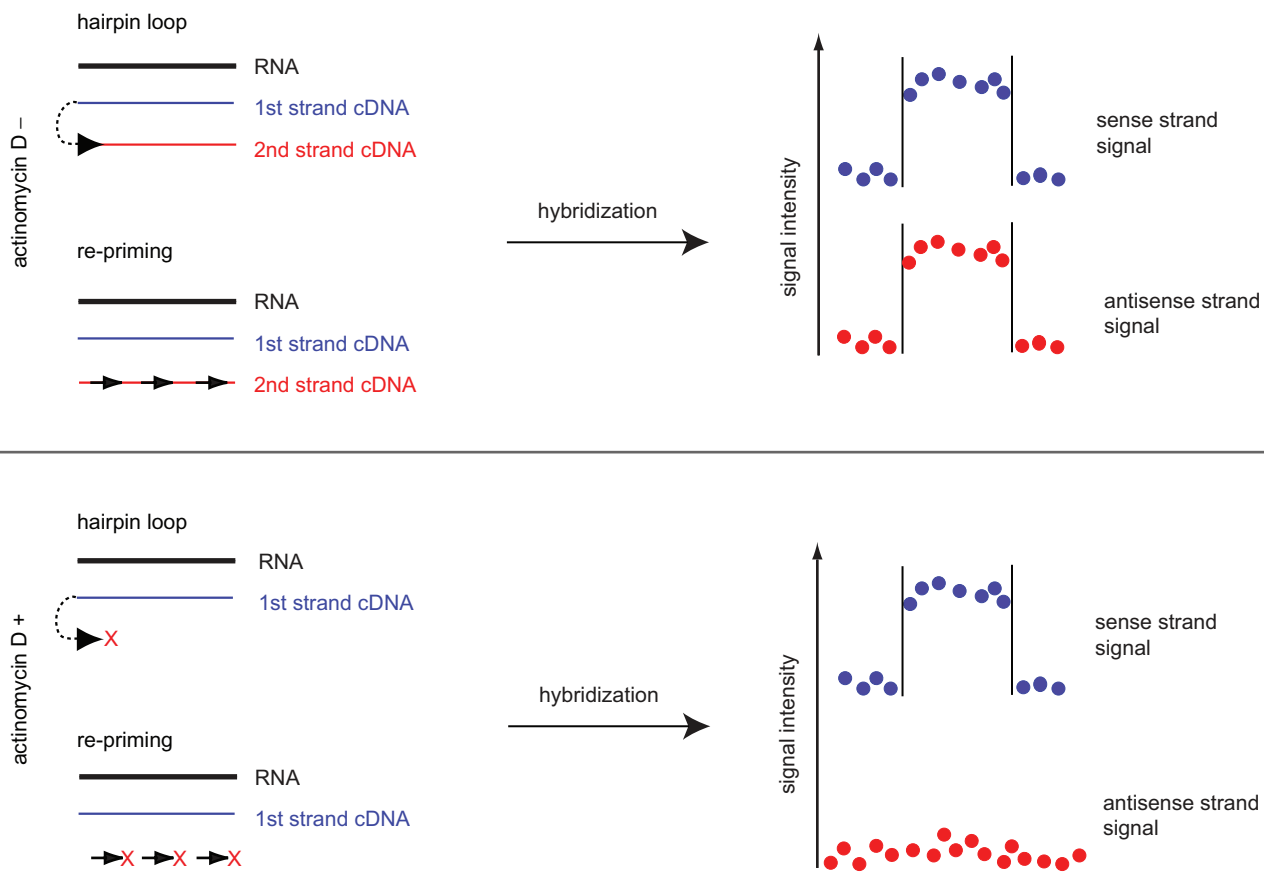


Figure 1. Hypothetical models for spurious second-strand synthesis by reverse transcription (upper panel) and the mode of inhibition by ActD (lower panel). During sample processing first-strand cDNA is generated from RNA molecules by reverse transcription. Spurious second-strand cDNA synthesis could occur using the first-strand cDNA as a template. Two possible mechanisms are: self-priming through hairpin loop extension of the first-strand cDNA, or re-priming either from RNA fragments formed during degradation of the RNA templates or from primers present in the reaction. In both cases, hybridization to strand-specific arrays would lead to artifactual antisense signals. ActD inhibits second-strand synthesis (lower panel, red cross), putatively by binding to bases of single-stranded DNA molecules, and therefore artifactual antisense-strand signals would not be detected.

Tiling arrays can be used for *de novo* identification of transcripts (7). Therefore, the hybridization signals of probes were examined along their chromosomal positions irrespective of previous annotation and separately for each strand. The profiles were partitioned into segments of constant hybridization intensity using a segmentation algorithm (7,20). We analyzed the effect of ActD for sense and antisense transcripts defined by segmentation (Supplementary Table 2). Antisense segments were defined based on three criteria: (i) the segments are expressed and overlap annotated genes located on the opposite strand, but not on the same strand, (ii) the segment lengths are longer than 48 bp and (iii) they are flanked by segments with reduced hybridization signal on both sides. A comparison shows that ActD has no quantitative effect on first-strand cDNA synthesis, since there is concordance between the expression levels of sense segments measured in ActD+ and ActD- reactions (Figure 3). However, the addition of ActD reduces the expression level below background for more than half of the antisense segments: among a total of 553 antisense segments, 347 give signal above background only in ActD- and 14 only in ActD+, while 192 antisense segments are detected in both

conditions (Supplementary Table 3). Therefore, strikingly, the number of antisense segments observed by using a standard reverse transcription protocol is decreased by 64% (347/539) with the inclusion of ActD.

Three independent lines of evidence are in agreement with the array hybridization results generated in the presence of ActD. First, we compared our array hybridization data to the stringent set of antisense transcripts obtained by full-length cDNAs sequencing (8). Despite complementarities between these two methods, as well as the comparison being between different experimental growth conditions (rich media versus minimal media), a better overlap was achieved for the ActD+ than the ActD- dataset (P value of $1.6e-9$ versus $2.5e-6$, Fisher's exact test). Second, we evaluated the effect of a stringent computational filter on removing putative antisense artifacts from hybridization results. This filter was previously developed to computationally remove putative antisense artifacts by requiring segments to have higher expression signal than seen on the opposite strand for at least part of their length (7). On ActD- data, the computational filter reduced the number of antisense segments by 63% (337/539). In contrast, the filter had only

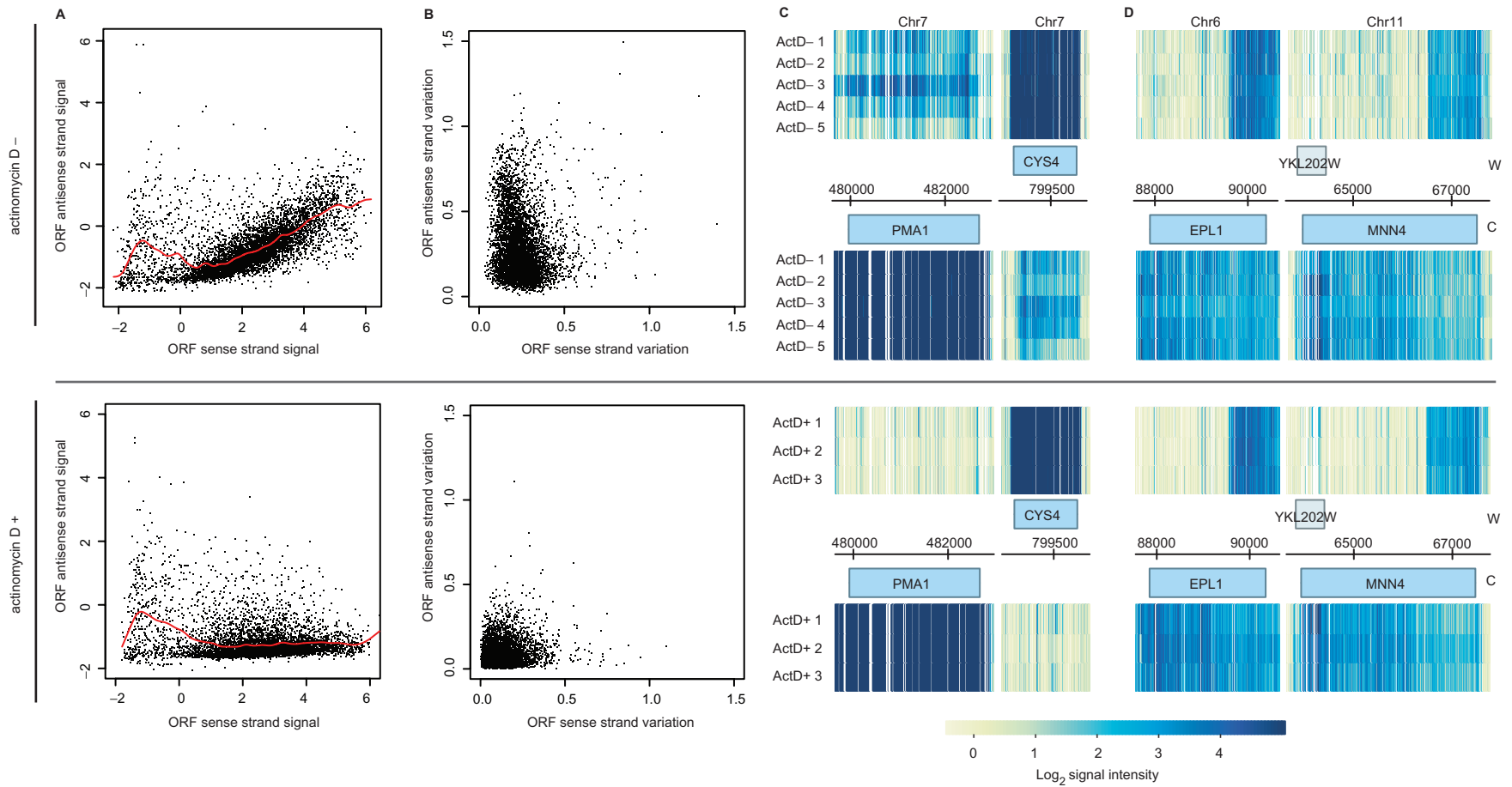


Figure 2. Genome-wide analysis of sense and antisense signals from hybridizations performed in either the absence (ActD⁻) or presence (ActD⁺) of ActD during reverse transcription. (A) Dependence of antisense signal intensity on the sense expression level. Red curves represent the fitted hybridization signals using smoothing splines. (B) Variation in signal intensities among biological replicates for the sense and antisense strands (SD). (C) Two examples of spurious antisense transcripts resolved by ActD. (D) Two examples of consistent antisense transcripts unaffected by ActD. In both panels C, and D, normalized signal intensities are shown for probes along their chromosomal position using a color gradient (*x*-axis; W, Watson strand; C, Crick strand). Upper panel, five ActD⁻ replicates (*y*-axis, ActD⁻, 1–5); lower panel, three ActD⁺ replicates (*y*-axis, ActD⁺ 1–3). Profiles for all genomic regions are available online. (www.ebi.ac.uk/huber-srv/actinomycinD).

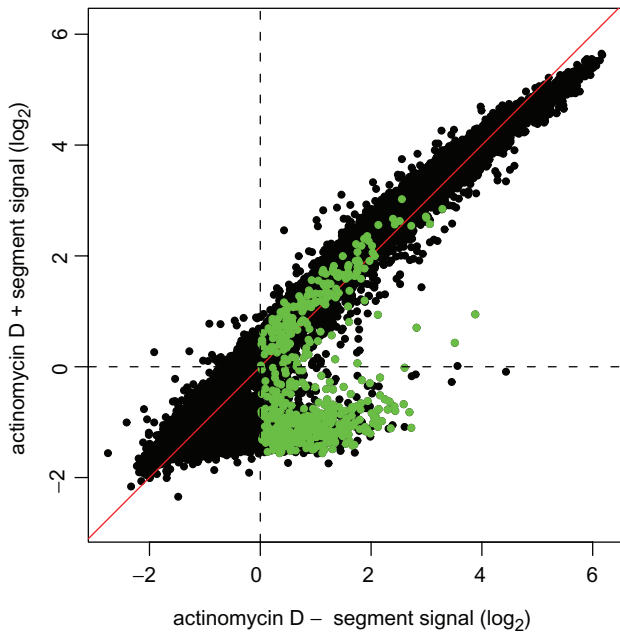


Figure 3. Scatter plot of segment expression levels between arrays generated with and without ActD. Green dots, antisense segments detected above background in ActD⁻ samples; Red line, identical expression in both conditions; dashed lines, background thresholds.

a mild effect on the number of antisense segments detected in ActD⁺: of 206 antisense segments, only 39 were filtered out. In addition, a comparison of antisense segments detected in ActD⁻ after filtering, and antisense segments detected in ActD⁺, shows good (but imperfect) concordance (158/250) (Supplementary Table 4). Since computational filters always face the dilemma of compromising between false positives and false negatives, the experimental improvement using ActD is more advantageous. This is clearly the case for antisense detection as the majority of antisense transcripts are weakly expressed (Supplementary Table 3) and thus hard to distinguish from noise. Notably, several antisense segments filtered out in ActD⁻ (48 segments, Figure 4A orange dots) are detectable in ActD⁺, which suggests that these cases were erroneously filtered (manual inspection confirms this). In addition, ActD⁺ also resolves artifactual antisense segments that erroneously passed the computational filter (44 segments, Figure 4A green dots). Third, we conducted semi-quantitative strand-specific RT-PCR analysis for 10 antisense segments (Figure 4B): of the three antisense transcripts detected both in ActD⁺ and ActD⁻ (*MBR1*, *EPL1*, *MRK1*), all three yielded a signal by strand-specific RT-PCR. In contrast, of seven antisense segments detected exclusively in ActD⁻ (*CYS4*, *EMP24*, *PNC1*, *MDH3*, *HAC1*, *TRR1*, *PFK1*), all seven yielded negative results by strand-specific RT-PCR.

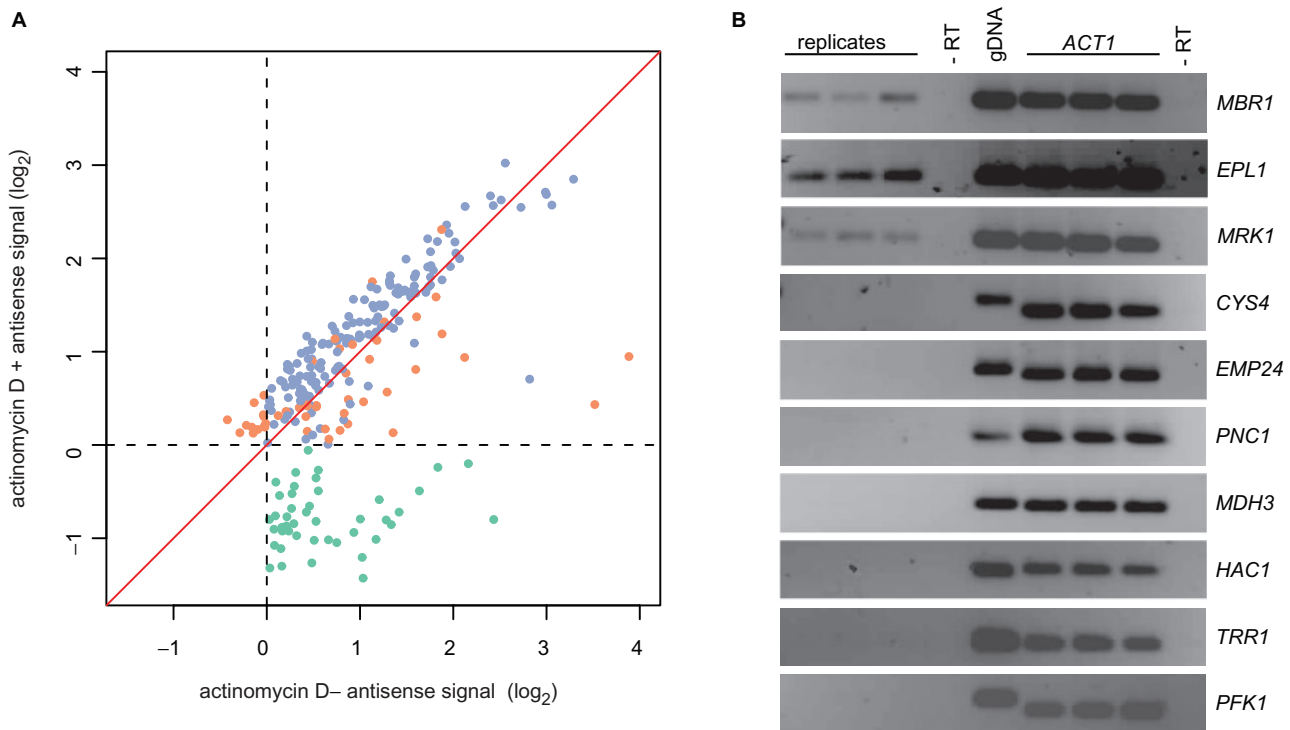


Figure 4. Validation of antisense transcripts. (A) Comparison of ActD and computational filtering. Expression level of antisense segments in reactions with (ActD⁺) and without ActD (ActD⁻). For ActD⁻ data, only antisense segments which passed the computational filter are shown. The dotted lines represent the background thresholds. Blue dots, segments common between ActD⁺ and filtered ActD⁻; orange dots, unique to ActD⁺; green dots, unique to filtered ActD⁻. A list of these segments is available (Supplementary Table 4). (B) Strand-specific RT-PCR of antisense transcripts. Three antisense RNAs opposite to *MRB1*, *EPL1* and *MRK1* yielded a RT-PCR product. Seven regions for which putative antisense artifacts were detected by array analysis opposite of *CYS4*, *EMP24*, *PNC1*, *MDH3*, *HAC1*, *TRR1* and *PFK1* did not yield a RT-PCR product. *ACT1* sense RT-PCR reactions served as internal controls for each individual RT-PCR; -RT designate RT reactions in which the reverse transcriptase was omitted; gDNA indicates PCR reactions with genomic DNA as a template; reactions were performed in triplicate.

CONCLUSIONS

Our results indicate that about half of all antisense signals observed in yeast by conventional protocols hybridizing first-strand cDNA to high-density tiling arrays are experimental artifacts. These artifacts appear to be largely triggered by spurious synthesis of second-strand cDNAs during reverse transcription reactions. Although the inhibitory effect of ActD on spurious second-strand synthesis during reverse transcription has been known for over 30 years (17), its application to microarray cDNA hybridization has been largely overlooked. Recent technological advancements in microarray design have allowed profiling transcription at the level of strand-specificity, but when conventional protocols for cDNA generation were used, the data analysis required stringent computational filtering to reduce the artifacts (7). The development of more sensitive and high resolution technologies needs to be complemented by the development of more accurate protocol designs.

We find that ActD enables a substantial experimental improvement to resolve the artifacts early during generation of the raw data. Although the precise molecular mechanism of this inhibition still awaits further elucidation, ActD likely acts through exerting an inhibitory effect on double-stranded DNA synthesis. Its inclusion in reverse transcription reactions increases both the sensitivity and the accuracy of array-detected antisense signals. Thus, we encourage the use of ActD when strand-specific transcription is interrogated by microarrays.

As reverse transcription is used in technologies beyond array hybridization, the use of ActD has broader applications. While widespread antisense transcription remains detectable in the yeast genome, the extent and identity of this transcription over the genomes of several other organisms might be reconsidered in light of these findings.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

National Institutes of Health (GM068717, HG000205 to L.M.S.) and the Deutsche Forschungsgemeinschaft (STE 1422/2-1 to L.M.S.). We thank Ye Ning and Marina Granovskaia for advice on experimental methods, Eran Segal, Wolfgang Huber, Himanshu Sinha, Lior David and Julien Gagneur for helpful comments on the manuscript, Joern Toedling for website template, Wolfgang Huber for providing the computing resources, and the contributors to the BIOCONDUCTOR (www.bioconductor.org) and R (<http://www.R-project.org>) projects for their software. Funding to pay the Open Access publication charges for this article was provided by the European Molecular Biology Laboratory.

Conflict of interest statement. None declared.

REFERENCES

- Mattick,J.S. (2004) RNA regulation: a new genetics? *Nat. Rev. Genet.*, **5**, 316–323.
- Vanhee-Brossollet,C. and Vaquero,C. (1998) Do natural antisense transcripts make sense in eukaryotes? *Gene*, **211**, 1–9.
- Munroe,S.H. and Zhu,J. (2006) Overlapping transcripts, double-stranded RNA and antisense regulation: a genomic perspective. *Cell. Mol. Life Sci.*, **63**, 2102–2118.
- Lapidot,M. and Pilpel,Y. (2006) Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep.*, **7**, 1216–1222.
- Tufarelli,C., Stanley,J.A., Garrick,D., Sharpe,J.A., Ayyub,H., Wood,W.G. and Higgs,D.R. (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.*, **34**, 157–165.
- Samanta,M.P., Tongprasit,W., Sethi,H., Chin,C.S. and Stolc,V. (2006) Global identification of noncoding RNAs in *Saccharomyces cerevisiae* by modulating an essential RNA processing pathway. *Proc. Natl Acad. Sci. USA*, **103**, 4192–4197.
- David,L., Huber,W., Granovskaia,M., Toedling,J., Palm,C.J., Bofkin,L., Jones,T., Davis,R.W. and Steinmetz,L.M. (2006) A high-resolution map of transcription in the yeast genome. *Proc. Natl Acad. Sci. USA*, **103**, 5320–5325.
- Miura,F., Kawaguchi,N., Sese,J., Toyoda,A., Hattori,M., Morishita,S. and Ito,T. (2006) A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc. Natl Acad. Sci. USA*, **103**, 17846–17851.
- Selinger,D.W., Cheung,K.J., Mei,R., Johansson,E.M., Richmond,C.S., Blattner,F.R., Lockhart,D.J. and Church,G.M. (2000) RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat. Biotechnol.*, **18**, 1262–1268.
- Yamada,K., Lim,J., Dale,J.M., Chen,H., Shinn,P., Palm,C.J., Southwick,A.M., Wu,H.C., Kim,C. *et al.* (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science*, **302**, 842–846.
- Bertone,P., Stolc,V., Royce,T.E., Rozowsky,J.S., Urban,A.E., Zhu,X., Rinn,J.L., Tongprasit,W., Samanta,M. *et al.* (2004) Global identification of human transcribed sequences with genome tiling arrays. *Science*, **306**, 2242–2246.
- Carninci,P., Kasukawa,T., Katayama,S., Gough,J., Frith,M.C., Maeda,N., Oyama,R., Ravasi,T., Lenhard,B. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.
- Johnson,J.M., Edwards,S., Shoemaker,D. and Schadt,E.E. (2005) Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet.*, **21**, 93–102.
- Spiegelman,S., Burny,A., Das,M.R., Keydar,J., Schlom,J., Travnicek,M. and Watson,K. (1970) DNA-directed DNA polymerase activity in oncogenic RNA viruses. *Nature*, **227**, 1029–1031.
- Gubler,U. (1987) Second-strand cDNA synthesis: classical method. *Methods Enzymol.*, **152**, 325–329.
- Gubler,U. (1987) Second-strand cDNA synthesis: mRNA fragments as primers. *Methods Enzymol.*, **152**, 330–335.
- Muller,W.E., Zahn,R.K. and Seidel,H.J. (1971) Inhibitors acting on nucleic acid synthesis in an oncogenic RNA virus. *Nat. New Biol.*, **232**, 143–145.
- Ruprecht,R.M., Goodman,N.C. and Spiegelman,S. (1973) Conditions for the selective synthesis of DNA complementary to template RNA. *Biochim. Biophys. Acta*, **294**, 192–203.
- Goldberg,I.H., Rabinowitz,M. and Reich,E. (1962) Basis of actinomycin action. I. DNA binding and inhibition of RNA-polymerase synthetic reactions by actinomycin. *Proc. Natl Acad. Sci. USA*, **48**, 2094–2101.
- Huber,W., Toedling,J. and Steinmetz,L.M. (2006) Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics*, **22**, 1963–1970.
- Gentleman,R.C., Carey,V.J., Bates,D.M., Bolstad,B., Dettling,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.