Published as: *Nat Genet.* 2008 August ; 40(8): 977–986.

Dynamic transcriptome of *Schizosaccharomyces pombe* shown by RNA-DNA hybrid mapping

Natalie Dutrow^{1,2}, David A Nix^{3,4}, Derick Holt^{1,2}, Brett Milash⁴, Brian Dalley⁵, Erick Westbroek^{1,2}, Timothy J Parnell^{1,2}, and Bradley R Cairns^{1,2}

¹Department of Oncological Sciences, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA.

²Howard Hughes Medical Institute and Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA.

³Research Informatics, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City 84112, USA.

⁴Bioinformatics Core Facility, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City 84112, USA.

⁵Microarray Core Facility, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City 84112, USA.

Abstract

We have determined the high-resolution strand-specific transcriptome of the fission yeast *S. pombe* under multiple growth conditions using a novel RNA-DNA hybridization mapping (HybMap) technique. HybMap uses an antibody against an RNA-DNA hybrid to detect RNA molecules hybridized to a high-density DNA oligonucleotide tiling microarray. HybMap showed exceptional dynamic range and reproducibility, and allowed us to identify strand-specific coding, noncoding and structural RNAs, as well as previously unknown RNAs conserved in distant yeast species. Notably, we found that virtually the entire euchromatic genome (including intergenics) is transcribed, with heterochromatin dampening intergenic transcription. We identified features including large numbers of condition-specific noncoding RNAs, extensive antisense transcription, new properties of antisense transcripts and induced divergent transcription. Furthermore, our HybMap data informed the efficiency and locations of RNA splicing genome-wide. Finally, we observed strand-specific transcription islands around tRNAs at heterochromatin boundaries inside centromeres. Here, we discuss these new features in terms of organism fitness and transcriptome evolution.

Recently, genome-wide approaches to monitor transcription have revealed several surprising features of the eukaryotic transcriptome, including extensive intergenic transcription, antisense transcription and a notable number of noncoding RNAs (ncRNA)¹⁻⁶. Furthermore, it has been found that, in some cases, it is the act of transcription (or the attendant chromatin modifications) that is important, rather than the ncRNA produced⁷.

© 2008 Nature Publishing Group <http://www.nature.com/naturegenetics>

Correspondence should be addressed to B.R.C. (brad.cairns@hci.utah.edu).

AUTHOR CONTRIBUTIONS N.D., D.A.N., and B.R.C.; system design and experimental approaches. B.D.; array method optimization. D.A.N., N.D., B.M., T.J.P., D.H., and B.R.C.; array design, data analysis methods and data analysis. E.W.; feature computation. N.D., D.A.N., and D.H.; figures. B.R.C., N.D. and D.A.N. wrote the manuscript.

Accession codes. NCBI GEO: *S. pombe* transcriptome microarray data have been deposited with accession number GSE11619.

Note: Supplementary information is available on the Nature Genetics website.

Together, these studies have initiated a new era in defining genetic elements and generated high interest in understanding the purpose and function of ncRNAs in the genome. Genome-wide transcriptome mapping has emerged as an important approach to help understand transcriptional dynamics and transcription-chromatin relationships, and to identify candidate functional ncRNAs.

Traditional transcript profiling approaches are often fraught with bias because of difficulties in labeling the isolated RNAs for detection on microarray formats⁸. To circumvent these limitations, we made use of a previously characterized RNA-DNA hybrid antibody that has been used to quantify RNA levels of structured and small RNAs in an array format⁹. Here, we adapted and extended this approach, creating a technique for acquiring a strand-specific transcriptome of an entire genome, termed HybMap, that is not susceptible to many of the common technical limitations of traditional methods. To apply HybMap, we chose the fission yeast *Schizosaccharomyces pombe*, an important model organism for understanding transcription and chromosomal biology with a highly compact genome. Basic elements of the *S. pombe* genome include a total of 13.8 Mb of DNA on three chromosomes, 20 kb of mitochondrial DNA, and fewer than 5,000 protein-coding genes (1 per ~2.5 kb). Detailed features of the genome will be discussed in the context of the transcriptome.

Our work had four goals: (i) to develop a technique for deriving an accurate strand-specific transcriptome of a genome at high resolution and low cost, (ii) to apply this technique to derive the *S. pombe* transcriptome, (iii) to uncover new candidate ORFs and ncRNAs for future studies and (iv) to reveal new features of transcription that inform our understanding of the relationship between transcription and chromatin.

RESULTS

HybMap RNA-DNA hybrid mapping

To create a high-resolution and strand-specific array, we tiled both strands of the entire *S. pombe* genome with 458,566 60-mer DNA oligonucleotide ‘probes’ (55-base resolution, 5-base overlap). All the 60-mer probes on the top (forward) strand had a precise 60-mer reverse complement probe on the bottom (reverse) strand, allowing strand-specific transcription to be monitored. We also included 500 60-mer control probes from the zebrafish (*Danio rerio*) genome not present in the *S. pombe* genome for array normalization and to provide a nonspecific hybridization background level of signal useful for comparison to true *S. pombe* probes; the median level of the zebrafish control probes will hereafter be termed ‘background’ in the text and will be depicted as a dotted line in all figures.

The HybMap technique can utilize total RNA isolated without poly(A) selection, reverse transcription, RNA (or DNA) amplification or nucleic acid labeling; thus omitting steps that might strongly bias the RNA pool and/or disfavor the isolation of small noncoding or structured RNAs. Purified total RNA was hybridized to the tiling array, and then incubated with a monoclonal antibody (S9.6) that recognizes an RNA-DNA hybrid. We then incubated arrays with a secondary antibody labeled with the fluorophore Cy3, allowing quantitation (Supplementary Methods online). The S9.6 antibody to RNA-DNA hybrids has been used previously for transcript detection in an array format^{9,10}. S9.6 has negligible sequence specificity¹⁰ and does not show a significant bias for GC content (Supplementary Fig. 1 online). S9.6 recognizes a hybrid of ~15 bp⁹, a result we have independently verified by analyzing signal from tRNAs with different registers of overlap with array probes (Supplementary Fig. 2 online). S9.6 is highly sensitive to mismatch: one mismatched base pair in a stretch of 15 bp reduces signal by 80-fold, and a second mismatch reduces ~20,000 fold (Supplementary Fig. 3 and Supplementary Table 1 online).

Evidence for extensive transcription in *S. pombe*

We purified total RNA from two separate cultures (biological replicates) of a prototrophic strain grown in rich medium (poly(A) RNA presented later) and carried out HybMap (see Methods). HybMap provided ~7,600-fold range of signal intensity (12.9 on a log₂ scale) and virtually all the probes on the array showed mean signal intensity well above the background control probes (Supplementary Fig. 4 online). Notably, 99.6% of all smoothed 2-probe windows had a signal above background, providing initial evidence for widespread transcription of the euchromatic *S. pombe* genome. Widespread transcription in other genomes has been reported previously^{11,12}; however, our use of internal control probes and a strand-specific approach allows for a direct and quantitative comparison genome-wide. We note that HybMap, like all array formats, does not directly measure transcription rate, but rather provides a steady-state RNA measurement.

Transcription in relation to a defined baseline

Determining whether a gene is activated or silenced requires a baseline standard of reference. The vast majority of intergenic regions (heterochromatin regions excepted) seemed to be transcribed at a very similar level, ~12-fold above background (Fig. 1a). To accurately quantify this basal level, we determined the median value of transcription of all intergenic regions, following their trimming on either end by 240 bp to limit the impact of numerous 5' and 3' UTR extensions (discussed below). This trimmed intergenic value is hereafter referred to as the 'baseline' and is depicted by a solid line in the figures; the baseline is 12-fold above the background. Notably, signal intensity generally deflects from baseline along the physical map in accordance with annotated genes and their strand specificity.

In accordance with extensive transcription, 87% of 5,049 genes, 79% of interrogated base pairs (25,180,884) and 75% of oligonucleotide probes were expressed at or above the baseline level (Fig. 1b). On the basis of signal intensity, we calculated that the average annotated gene is transcribed ~12-fold above baseline and 149-fold above background. Median signals for different gene classes are shown in Figure 1a. For Pol II genes, the highest signals, 630-fold above baseline and ~7,600-fold above background, mapped to ribosomal protein genes.

Transcriptional features of euchromatin

To illustrate the high-resolution features of euchromatin, we show the HybMap signal intensity of two genomic regions alongside current annotation (Fig. 2a,b). One region has a relatively simple transcriptional architecture; most genes reside in tandem orientation on one strand (bottom/reverse), and the data generally conforms to current annotation, with UTR designations excepted (Fig. 2a). The HybMap results of the second region deviate considerably from annotation in several locations, with, for example, attribution of *sno52* to the incorrect strand, longer untranslated RNA (UTR) extensions, and the appearance of antisense transcription (Fig. 2b). We will analyze elements from both of these regions in detail throughout the manuscript to illustrate concepts.

Features of UTRs and RNA processing

As in other organisms^{1,3}, we commonly observed the extension of 5' - and especially 3' - UTRs for Pol II genes beyond current annotation: 65% have extensions beyond their annotated 5' end, and 87% have extensions beyond their annotated 3' end; the median 5' UTR is 81 bp and the median 3' UTR is 231 bp. Additional features, GO TERM classification and discussion of UTRs (including overlapping convergent 3' UTRs) are

provided in Supplementary Table 2, Supplementary Figure 5 and Supplementary Note online.

According to current annotation, introns reside in 43% of genes and average ~81 bases¹³. Thus, we interrogated most introns by one probe fully within the intron-exon boundary (intron probes) and two probes that overlap the boundary (boundary probes). Our data verified the notion that splicing is efficient in *S. pombe*, as intron probes showed a mean 32-fold decrease in signal intensity relative to exon probes (Fig. 2c). The current annotation of predicted introns seemed largely accurate; 94% of annotated introns examined showed a pronounced decrease in signal intensity. However, of the remaining 6%, 5% were expressed at a level similar to that of the surrounding gene, suggesting either inefficient splicing or possible misannotation. Notably, the remaining 1% showed a signal intensity greater than eightfold that observed at the surrounding gene (Fig. 2d), including three potential ncRNAs residing within introns of annotated protein-coding genes, which we verified by strand-specific qPCR—rps3, SPBC1734.10c and SPAC6F6.03c (Supplementary Fig. 6 and Supplementary Table 3 online). Unique transcripts processed from introns are rare; only four are annotated in *S. pombe*, and all are snoRNAs (Fig. 2d). As mature snoRNAs are structured and lack a poly(A) tail, they are highly underrepresented in other transcriptome formats³, whereas they are exceptionally prominent features in the HybMap datasets. Finally, tRNAs are structured RNAs that provide exceptionally high signal (Fig. 2e), although we note that their similarity precludes their precise mapping.

The abundance of non-annotated transcripts

We observed large numbers of non-annotated transcription units (fragments), termed ‘transfrags’¹². For example, the recently identified^{14,15} telomerase ncRNA (TER1) was easily observed in our dataset (Fig. 3a). We identified transfrags on the basis of the following criteria: (i) two or more consecutive probes with a pseudo median value at least 3.5-fold above baseline and (ii) separation by at least 180 bp from any annotated element. Using these thresholds, we defined 7,093 previously unknown transfrags in rich medium, with increasing thresholds yielding moderately lower transfrag numbers (Fig. 3b). Of these, 38% reside in intergenic regions (Fig. 3c) and 35% are antisense to a known transcript (Fig. 3d). The remainder reside in intergenic–UTR boundaries and are a mixture of independent transfrags and long unannotated UTRs. Twenty transfrags were tested by both RT-PCR (Supplementary Table 3) and strand-specific quantitative PCR, and all were verified (Supplementary Fig. 6). We then analyzed the genome-wide occupancy profile of RNA Pol II (ref. 16) and found that most highly transcribed transfrags, including SPAC6F6.03c (noted above) show levels of Pol II well above the genome average (Fig. 3c-e and data not shown). Taken together, HybMap and the Pol II ChIP data reveal large numbers of potential new transcripts.

Transcriptomes for alternative growth and stress conditions

We then derived transcriptomes for alternative growth and stress conditions: minimal medium, heat stress and DNA damage (methyl methanesulfonate, MMS). As expected, GO analyses showed changes in gene classes that related to the conditions imposed (Supplementary Tables 4-9 online)¹⁷. Here, we discuss only key general features, showing the signal intensity from rich medium in a bar format and that from the alternative condition as a superimposed colored line (Fig. 4). For heat stress, the heat shock gene *hsp16* is moderately transcribed at 30 °C but upregulated more than tenfold during heat stress (Fig. 4a). Likewise, genes for thiamine biosynthesis (*thi4*) and phenazine metabolism (*ase1*) are upregulated in minimal medium or DNA damage, respectively (Fig. 4c,e). Full datasets are available at the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) database (see Accession codes section in Methods).

Transfrags upregulated during alternative growth conditions

To identify candidate functional ncRNAs for each growth condition, we defined the transfrags present in each condition that increase or decrease substantially relative to their values in rich medium, termed ‘diff frags’ (differential transfrags). To quantify diff frags, we first selected particular transfrags—those upregulated >3.5-fold from baseline in their initial condition and present in intergenic regions >180 bp away from an annotated gene (total ~7,200, Fig. 3b). We then determined (for each condition) the percentage of these transfrags whose signal intensity differed more than twofold from their levels in rich medium. By these criteria, we found that the percentage of transfrags that are ‘diff frags’ is 24% in heat stress, 12% in minimal media and 15% in the presence of MMS. Thus, a substantial fraction of the non-annotated transcriptome is altered when growth conditions are changed. Examples of notable diff frags include an ncRNA arising on an antisense strand during heat stress (Fig. 4b), and two arising from intergenic regions near a 5′ end of a gene, either in minimal medium or MMS (Fig. 4d,f).

Conservation of transcribed loci

We then determined whether RNAs newly identified through HybMap are conserved in distantly related yeast species. We selected twenty newly identified RNAs that spanned multiple probes (typically >5 probes) and were expressed >10-fold above baseline, and then interrogated two recently sequenced distant relatives of *S. pombe*, *S. japonicus* and *S. octosporus*. Using a stringent cut-off (*E* value of 10^{-4}), we found that 11 of these RNAs were conserved in *S. octosporus*, and 3 in the more distantly related *S. japonicus*. This analysis suggests that HybMap is indeed revealing new functional transcripts. To determine the potential for these RNAs to encode proteins, we carried out a 3-frame translation and searched for ORFs with a codon bias specific for *S. pombe*; 8 of the 20 RNAs revealed such an ORF. Finally, our intersection of the conservation and ORF searches identified two RNAs (transfrags Q and AA) that share both codon bias and sequence conservation within their predicted longest ORF (discussed further below and in Supplementary Table 10 online). Thus, HybMap identifies transfrags with significant potential for function, either as ncRNAs or small proteins.

Properties of antisense transcripts

Antisense transcription occurs at particular loci in the mammalian genome¹⁸ and in the *S. pombe*¹⁹ genome. Here, we characterize its extent genome-wide, its properties and its source (Fig. 1a and Fig. 5). First, we observed a significant correlation between signals from sense-antisense probe pairs (+0.595, *P* value $< 2 \times 10^{-16}$), and between antisense signal and RNA Pol II occupancy (+0.50, *P* value $< 2 \times 10^{-16}$). Sense-antisense correlation is especially clear in dynamic transcriptome comparisons: genes upregulated >fourfold during heat shock show greatly increased antisense transcription, and repressed genes show a substantial decrease (Fig. 5a). Of note, antisense transcription is slightly negatively correlated with histone H3 occupancy (−0.20; *P* value $< 2 \times 10^{-16}$)¹⁶, but slightly positively correlated with H3K36me3 (+0.21; *P* value $< 2 \times 10^{-16}$), a mark found on Pol II-transcribed regions. Together, these results raise the possibility that areas that are highly transcribed and slightly histone deficient are susceptible to increased antisense transcription; that is, that strong sense transcription begets antisense transcription.

Notably, our examination of poly(A)-selected RNA samples showed that antisense transcripts were highly reduced (Fig. 5b and Fig. 6a), suggesting that most antisense transcripts lack polyadenylation. Furthermore, the deflection of signal from background at genes was often greater, and the distinction between annotated genes was clearer as a result of the lack of antisense signal (Fig. 6a). However, we did find a few locations where antisense signal is high along the entire gene in our poly(A) datasets, suggesting that certain

antisense transcripts, including those derived from genes encoding certain transcription factors, are indeed polyadenylated (Fig. 5c and Supplementary Table 11 online). Additional features of antisense transcripts at the different RNA PolI/II/III classes are described in the Supplementary Note.

Induced divergent transcription

A particularly marked feature of our datasets was the prevalence of ‘induced divergent transcription’, in which transcripts were found on the opposite strand directed away from the promoter of a highly transcribed annotated gene. For example, during heat stress, a transcript diverges away from the *hsp16* promoter on the opposite strand, generating an antisense transcript that encompasses most of the flanking gene (Fig. 4a); this was confirmed by strand-specific qPCR (data not shown). Another confirmed instance is the *cnl2-thi4* locus in minimal medium (Fig. 4c and data not shown).

Previously unknown divergent poly(A) RNAs were also common in our static poly(A) datasets, suggesting that divergent transcription is a feature of basal transcription at many loci. For example, we found that poly(A) transcripts diverge from *hsp60*, *cip2* and *SPAC630.07c* (Fig. 6a). Two of these divergent transcripts, *hsp60* and *cip2*, were tested and verified by strand-specific qPCR (data not shown). These unannotated poly(A) transcripts are obscured in the total RNA dataset as a result of the high levels of antisense (which is generally not polyadenylated), but are obvious transfrags in the poly(A)-selected RNA dataset (Fig. 6a). In addition, we found a clear strand-specific polyadenylated transcript corresponding to the two transcripts (transfrag Q and AA, Supplementary Table 10) that are highly conserved among the three related yeast species and that share an *S. pombe* codon bias (Fig. 6b,c). The total RNA and poly(A)-selected RNA datasets are thus useful for uncovering different aspects of the transcriptome.

A strand-specific drop in signal before the TSS

One feature we commonly observed at tandem genes is a reduction in signal intensity of sense probes just before the true transcription start site (TSS). A clear example of this is shown in Figure 2a, where ‘dips’ are seen selectively in the sense strand just adjacent to the 5’ end (note that these genes are on the bottom/reverse strand). Although common, it is not a constant feature. At present, we do not understand the basis for this common and curious feature.

Intergenic transcription measures telomeric chromatin

The *S. pombe* genome contains three types of heterochromatic regions: telomeres, centromeres and the silent mating locus. In yeasts, repressive telomeric heterochromatin propagates from the telomere end (which bears repetitive elements) into the chromosome arm, and is opposed by chromatin modifications associated with transcription^{19,20}, by transcription itself, or by Pol III genes (TFIIIC binding sites)²¹. In *S. pombe*, heterochromatic regions bear H3K9me (ref. 22). A notable feature of our data is the extent to which intergenic transcription gradually falls from baseline levels to near background levels as examination progresses from the gene-rich arm toward the telomere (Fig. 7a); intergenic transcript signals drop on both strands in synchrony as H3K9me is encountered. We suggest that the level of opportunistic transcription within intergenic regions serves as a proxy measurement for the extent of chromatin silencing.

Transcriptional features of centromeres

The centromeres of *S. pombe* are complex. For brevity, we will separate them into four general regions: (i) a partially unique central region that bears high levels of the histone H3

variant CENPA23,24, flanked by (ii) an innermost repeat that bears sets of tRNAs, which may help form a boundary^{21,25} between the central region and (iii) regions of repetitive DNA elements (*dg*, *dhl*, others), which are silenced through an RNAi-dependent mechanism (involving H3K9me and histone deacetylation) and are typically flanked by (iv) tRNAs of various numbers (chr. 1, right side excepted) present at the boundary between the silent repeat region (heterochromatin) and expressed genes²⁶⁻³¹. With repetitive elements, the signal represents a class-average sum, and in certain repeat regions, there is little if any transcription when copy number is considered. However, certain sections of the centromere (that is, those flanking *dhl*) are clearly transcribed and have observable signal from both strands, consistent with earlier reports³², and our data helps inform the locations and magnitude of that transcription (Fig. 7b and Supplementary Fig. 7 online). Results for an additional heterochromatic region (the MAT locus, Supplementary Fig. 8 online) as well as the rDNA locus (Supplementary Fig. 9 online) are also presented.

tRNAs and strand-specific transcription in heterochromatin

One notable feature of the centromere datasets is the effect of tRNAs residing in the imr repeats. We note that signal from tRNAs themselves cannot be uniquely attributed to a map location, given their near identity. The tRNA transcript itself, however, is not the feature of interest. Rather, we observed that the region surrounding the tRNA is selectively transcribed on the strand opposite the tRNA (Fig. 7b, Supplementary Fig. 7 and data not shown). In most cases, these flanking sequences are unique, allowing definitive attribution. Furthermore, this pronounced strand-specific effect is not generally observed with tRNAs in euchromatin (Fig. 2a,e). Thus, by an unknown mechanism, either Pol III transcription or machinery seems to create an 'island' within heterochromatin that enables opportunistic transcription on the strand opposite the tRNA.

DISCUSSION

Here, a method for high-density strand-specific transcriptome profiling is developed and applied to *S. pombe*. We find that HybMap is a simple, highly reproducible, accurate and relatively inexpensive technique that should be widely applicable to other organisms. As HybMap uses total RNA without amplification or labeling, structured and non-polyadenylated RNAs are retained and clearly detected. Although our studies were done with a single array platform (Agilent Technologies) the simplicity of the technique should allow its application to other array platforms, although we have not tested this directly. Our use of a large set of internal control probes (from zebrafish) enabled effective normalization of arrays and provided a measure of the nonspecific background signal, a feature important for distinguishing noise from true transcription. The sensitivity and wide dynamic range of the technique allowed us to examine new properties of intergenic and antisense transcription, as well as other features of the transcriptome.

First, we provide extensive information useful for improving the annotation of the genome, including large numbers of previously unknown ncRNAs at various threshold criteria. Of particular interest are the selective changes at non-annotated loci (diffraqs) in alternative growth conditions. A challenge for future studies is determining what proportion of these transcripts contributes to general or condition-dependent fitness. Here, our separate datasets for both total RNA and poly(A)-selected RNA should prove useful for determining whether these transcripts are more likely new coding or ncRNAs.

We provide evidence for extensive transcription of intergenic regions in *S. pombe*. Although we have not evaluated the purpose for this transcription, we hypothesize several possible uses. First, this transcription is likely antagonistic to repressive chromatin and may help in chromatin dynamics by helping to increase histone turnover, thus providing searchable

space for transcription factors and other DNA-binding proteins. Alternatively, it might be a mechanism for locating DNA damage, with RNA polymerase recruiting factors for transcription-coupled repair to damaged sites³³. Finally, we emphasize its possible utility in genome evolution: with opportunistic transcription allowing the cell to survey genome sequence space for new useful coding or ncRNAs and thereby transform opportunistic transcription into functional transcription. By this mechanism, an ncRNA or small ORF is provided the opportunity to evolve (through mutations or insertions, for example) from one of minor benefit to a more important contributor. Transcripts derived from intergenics were not generally polyadenylated, but may still be at levels sufficient for generating low levels of peptides whose composition and abundance could be modified through mutation and selection. A particularly notable feature was basal and induced divergent transcription at certain strong promoters, which were often polyadenylated (Fig. 6a). One interpretation is that these transcripts might arise from a misregulation of directional fidelity, with basal factors (such as TBP) and RNA polymerase simply assembling and initiating in the 'wrong' direction, perhaps as a result of the lack of sequence stringency for TBP binding. However, we speculate that this process might provide a means for evolving coordinated expression of coding or ncRNAs; through mutation, rearrangement or insertions, a more useful divergent promoter and transcript can evolve from this initial version. Thus, a transcriptome that might initially seem wasteful and inaccurate may, in fact, be a vehicle to assist in adaptation.

We observe extensive antisense transcription in genes, which is shown to generally lack polyadenylation. Thus, antisense transcription may originate without the full use of the factors (basal and chromatin) for site-specific promoter initiation and processivity, although this remains to be formally tested. We observed increased antisense transcription at highly transcribed and histone-deficient genes, attributes consistent with an opportunistic origin in open chromatin. At present, it is not clear whether antisense transcription within genes has a functional purpose in *S. pombe*. We note that both sense and antisense transcription are very low (near background) from inter-genics within heterochromatin (telomeres; Fig. 7a). We suggest that intergenic transcription levels serve as a proxy measurement of the chromatin state of a large region, with polymerized heterochromatin preventing opportunistic initiation.

One key issue is how opportunistic sense and antisense transcription is insulated from the *S. pombe* RNAi pathway, as double-stranded RNAs can lead to siRNA silencing^{32,34-36}. First, as HybMap does not measure the level of cellular dsRNA³⁷, it remains unknown whether the complementary sense/antisense transcripts are indeed double-stranded *in vivo*. However, the positive feedback loop for siRNA production is known to utilize nascent poly(A) RNA, whereas our data shows that the vast majority of antisense transcripts lack polyadenylation. Thus, a lack of polyadenylation may serve to insulate opportunistic dsRNA from entering the RNAi pathway.

One particularly notable feature was a sense strand-specific 'dip' adjacent to the 5' end, most commonly observed at tandem genes. One speculative possibility (among many) is that a DNA loop may form between an enhancer and the pre-initiation complex, and that the enhancer-binding protein could capture an opportunistic RNA polymerase traveling from the upstream gene and then pass Pol II directly to the pre-initiation complex, thus preventing it from passing through and transcribing the loop region in one particular direction. An additional strand-specific feature was the presence of 'islands' of transcription on the antisense strand around tRNAs in heterochromatin. Here, it will be of interest to determine how this occurs and whether this transcription helps form the boundary between the very different chromatin of the central region (bearing CENPA) and the pericentric repeat region. This feature may be mechanistically similar to the previously demonstrated transcription near heterochromatin boundary elements (TFIIIC binding sites) at the MAT locus²¹. These

are some of the many prominent and unexpected features of the *S. pombe* transcriptome worthy of exploration in future studies.

METHODS

Yeast strains and treatment conditions

Yeast growth, manipulations and molecular biology were done according to standard protocols. We used a prototrophic *Schizosaccharomyces pombe* *h90* strain for all experiments. For rich medium, the strain was grown on yeast extract with glucose (2%) supplemented with histidine, uracil, leucine and adenine at 32 °C to an optical density at 600 nm of 1.0, unless otherwise noted. We harvested the cells and, after mechanical pulverization, we stored the cell material at –80 °C. For the heat shock transcriptome, cells were cultured to an optical density at 600 nm of 1.0 and transferred to flasks that were preheated in a 40 °C water bath. Cells were incubated, with shaking, for 15 min and harvested. For the minimal medium transcriptome, cells were cultured as described in media lacking amino acids and containing 2% glucose. For the methyl methanesulfonate transcriptome, cells were cultured to an optical density at 600 nm of 0.6 and treated with MMS (0.02%) for 1 h. All samples were grown and collected in duplicate.

Nucleic acid purification

We extracted total RNA using the Ambion *mirVana* kit and isolated poly(A) RNA with the Qiagen Oligotex PolyA Purification kit.

Design of tiling array

We tiled the entire *S. pombe* genome, including the mitochondrial genome, with 60-mer oligos every 55 base pairs. We generated the reverse complement of each probe on the forward strand in order to tile the reverse strand for a total of 458,566 probes. In addition, we added 500 probes from a zebrafish array with no homology to *S. pombe* as background controls. The zebrafish probes have the same median GC content (37%) and melting temperature (71 °C) as the *S. pombe* probes. We aligned each probe to the genome using the BLAST algorithm to acquire copy number information. The set of two 244K arrays was manufactured by Agilent Technologies.

Array hybridization and antibody detection

RNA samples (15 µg total RNA or 1 µg poly(A) RNA) were fragmented and combined with Agilent Hi-RPM GE Hybridization kit reagents. We then hybridized these samples to the array slides at 65 °C for 17 h. The slides were washed in 6× SSPE, 0.005% *N*-lauroylsarcosine for 1 min at 20 °C and in 0.06× SSPE, 0.005% *N*-lauroylsarcosine for 1 min at 31 °C. We then incubated the slides with 0.025 mg/ml primary monoclonal mouse antibody to RNA-DNA hybrid complexes, available in a hybridoma cell line from ATCC (HB-8730). The antibody was diluted in 500 µl of SB+BSA (100 mM MES, pH 6.6, 1 M NaCl, 0.05% Tween 20, and 1 mg/ml BSA). The 60 min incubation is at 25 °C. After this incubation, we washed the slides four times for 3 min in NSWB (6× SSPE, 0.01% Tween 20). The secondary Cy3-labeled goat antibody to mouse (KPL, 078–18–061) was used at a concentration of 3 µg/ml in 500 µl of SB+BSA. We incubated the slides with the secondary antibody for 60 min at 25 °C. Finally, we washed the slides four times for 3 min in NSWB at room temperature and scanned them in an Agilent G2505B Microarray Scanner using extended dynamic range software. Feature extraction of the scanned images was done using Agilent Feature Extraction Software version 9.5.1.

Computational analytical methods

All of the software used in this analysis is open source and available from the TiMAT2 project site. We obtained *S. pombe* annotation and genomic sequence from the Wellcome Trust Sanger Institute (referred to here as the April 2007 build). These files, as well as the low- and high-level analysis files, can be downloaded from our bioserver. See below for relevant URLs.

Transcriptome analysis

For static maps, processing of the Agilent microarray transcriptome data was done in three basic steps: data normalization, sliding window summaries and enriched region identification. For each condition (growth and or RNA fraction) and strand, we extracted the median unadjusted signal intensities from the cy3 channel. We then mapped probes to the April 2007 build and retained those with <100 exact matches. Two biological replicate samples were isolated for each condition, and between replicate pairs the datasets were highly similar (heat stress, $r = 0.96$; minimal medium, $r = 0.96$; DNA damage, $r = 0.97$). We subject the paired datasets for each condition to quantile normalization³⁸ and two different ways of scaling. In the first, we scaled the data such that the median of the zebrafish control probes was one. In the second, we scaled the data such that the median of the trimmed intergenic region probes was one. Trimmed intergenic probes were defined as those >240 bp (4 probes) from any known annotation. We calculated probe level ‘oligo’ summaries by taking the $\log_2$ of the mean of the two biological replica probes for each condition. Window level summaries were generated by identifying windows of 60 bp containing two or more oligo start positions and calculating a $\log_2$ pseudo median on the associated values. This window summary score was assigned to the center position of the window “Pse” or represented as heat map “PseHM” data. Lastly, extended regions of high scoring windows, called “intervals,” were identified by merging windows that exceed a set threshold and abut or overlap. We picked several different thresholds on the basis of the relative transcription above the trimmed intergenic or zebrafish control median value.

Dynamic difference maps

To identify regions of change between different RNA fractions (total and poly(A)) and different growth conditions (rich, minimal, DNA damage and heat shock), we used a similar approach as that used in generating the static maps. The two biological replicas for each condition were quantile normalized and scaled such that the median of the zebrafish controls was one. We calculated probe level summaries by taking the $\log_2$ ratio between the mean treatment and mean control. The control data was defined as either the RNA derived from the rich medium condition or the total RNA. We calculated window level summaries by identifying 60-bp windows that contain 2 or more oligos. These windows were scored by first calculating all the relative difference pairs between the treatment and the control replica probes, and second, by calculating the pseudo median of these relative difference pairs. In some cases, the pseudo median relative difference scores were converted to $\log_2$ (ratios). Lastly, overlapping high- or low-scoring windows were merged into intervals and ranked by their best window score.

Strand-specific quantitative PCR analysis. See Supplementary Methods

URLs—TiMAT2 project site, <http://sourceforge.net/projects/timat2>; *S. pombe* annotation, <ftp://ftp.sanger.ac.uk/pub/yeast/pombe/GFF> (*S.pombe.gff* 3/16/07); *S. pombe* genome sequence, ftp://ftp.sanger.ac.uk/pub/yeast/pombe/Chromosome_contigs (*chromosomeX.contig* 4/23/07); Supplementary datasets, http://bioserver.hci.utah.edu/SupplementalPaperInfo/2008/Dutrow_NatGen_PombeTranscriptome and <http://>

bioserver.hci.utah.edu:8080/DAS2/das2; IGB, <http://bioserver.hci.utah.edu/BioInfo/index.php/Software:IGB>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank S. Leppla (US National Institutes of Health) for generously providing the S9.6 antibody and Bob Schackmann (University of Utah) for oligo synthesis. The work was supported by the Howard Hughes Medical Institute (B.R.C., D.H., T.J.P., and supplies), US National Institutes of Health Genetics Training Grant T32 GM007464 (N.D.), the Huntsman Cancer Institute (D.A.N.), and CA24014 (for core facilities).

References

1. Birney E, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007; 447:799–816. [PubMed: 17571346]
2. Okazaki Y, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*. 2002; 420:563–573. [PubMed: 12466851]
3. David L, et al. A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. USA*. 2006; 103:5320–5325. [PubMed: 16569694]
4. Washietl S, Hofacker IL, Lukasser M, Huttenhofer A, Stadler PF. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat. Biotechnol.* 2005; 23:1383–1390. [PubMed: 16273071]
5. Kampa D, et al. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.* 2004; 14:331–342. [PubMed: 14993201]
6. Steigle S, Huber W, Stocsits C, Stadler PF, Nieselt K. Comparative analysis of structured RNAs in *S. cerevisiae* indicates a multitude of different functions. *BMC Biol.* 2007; 5:25. [PubMed: 17577407]
7. Struhl K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat. Struct. Mol. Biol.* 2007; 14:103–105. [PubMed: 17277804]
8. Perocchi F, Xu Z, Clauder-Munster S, Steinmetz LM. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res.* 2007; 35:e128. [PubMed: 17897965]
9. Hu Z, Zhang A, Storz G, Gottesman S, Leppla SH. An antibody-based microarray assay for small RNA detection. *Nucleic Acids Res.* 2006; 34:e52. [PubMed: 16614443]
10. Boguslawski SJ, et al. Characterization of monoclonal antibody to DNA:RNA and its application to immunodetection of hybrids. *J. Immunol. Methods*. 1986; 89:123–130. [PubMed: 2422282]
11. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.* 2007; 8:413–423. [PubMed: 17486121]
12. Kapranov P, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*. 2007; 316:1484–1488. [PubMed: 17510325]
13. Wood V, et al. The genome sequence of *Schizosaccharomyces pombe*. *Nature*. 2002; 415:871–880. [PubMed: 11859360]
14. Leonardi J, Box JA, Bunch JT, Baumann P. TER1, the RNA subunit of fission yeast telomerase. *Nat. Struct. Mol. Biol.* 2008; 15:26–33. [PubMed: 18157152]
15. Webb CJ, Zakian VA. Identification and characterization of the *Schizosaccharomyces pombe* TER1 telomerase RNA. *Nat. Struct. Mol. Biol.* 2008; 15:34–42. [PubMed: 18157149]
16. Gordon M, et al. Genome-wide dynamics of SAPHIRE, an essential complex for gene activation and chromatin boundaries. *Mol. Cell. Biol.* 2007; 27:4058–4069. [PubMed: 17371846]
17. Chen D, et al. Global transcriptional responses of fission yeast to environmental stress. *Mol. Biol. Cell*. 2003; 14:214–229. [PubMed: 12529438]

18. Katayama S, et al. Antisense transcription in the mammalian transcriptome. *Science*. 2005; 309:1564–1566. [PubMed: 16141073]
19. Nicolas E, et al. Distinct roles of HDAC complexes in promoter silencing, antisense suppression and DNA damage protection. *Nat. Struct. Mol. Biol.* 2007; 14:372–380. [PubMed: 17450151]
20. Wiren M, et al. Genomewide analysis of nucleosome density histone acetylation and HDAC function in fission yeast. *EMBO J.* 2005; 24:2906–2918. [PubMed: 16079916]
21. Noma K, Cam HP, Maraia RJ, Grewal SI. A role for TFIIC transcription factor complex in genome organization. *Cell*. 2006; 125:859–872. [PubMed: 16751097]
22. Volpe TA, et al. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*. 2002; 297:1833–1837. [PubMed: 12193640]
23. Allshire RC, Javerzat JP, Redhead NJ, Cranston G. Position effect variegation at fission yeast centromeres. *Cell*. 1994; 76:157–169. [PubMed: 8287474]
24. Takahashi K, et al. A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. *Mol. Biol. Cell*. 1992; 3:819–835. [PubMed: 1515677]
25. Scott KC, White CV, Willard HF. An RNA polymerase III-dependent heterochromatin barrier at fission yeast centromere 1. *PLoS ONE*. 2007; 2:e1099. [PubMed: 17971862]
26. Baum M, Ngan VK, Clarke L. The centromeric K-type repeat and the central core are together sufficient to establish a functional *Schizosaccharomyces pombe* centromere. *Mol. Biol. Cell*. 1994; 5:747–761. [PubMed: 7812044]
27. Partridge JF, Scott KS, Bannister AJ, Kouzarides T, Allshire RC. cis-acting DNA from fission yeast centromeres mediates histone H3 methylation and recruitment of silencing factors and cohesin to an ectopic site. *Curr. Biol.* 2002; 12:1652–1660. [PubMed: 12361567]
28. Steiner NC, Hahnenberger KM, Clarke L. Centromeres of the fission yeast *Schizosaccharomyces pombe* are highly variable genetic loci. *Mol. Cell. Biol.* 1993; 13:4578–4587. [PubMed: 8336703]
29. Clarke L, Amstutz H, Fishel B, Carbon J. Analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. *Proc. Natl. Acad. Sci. USA*. 1986; 83:8253–8257. [PubMed: 3464952]
30. Nakaseko Y, Kinoshita N, Yanagida M. A novel sequence common to the centromere regions of *Schizosaccharomyces pombe* chromosomes. *Nucleic Acids Res.* 1987; 15:4705–4715. [PubMed: 3601654]
31. Nakaseko Y, Adachi Y, Funahashi S, Niwa O, Yanagida M. Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast. *EMBO J.* 1986; 5:1011–1021. [PubMed: 15957216]
32. Motamedi MR, et al. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell*. 2004; 119:789–802. [PubMed: 15607976]
33. Lindsey-Boltz LA, Sancar A. RNA polymerase: the most specific damage recognition protein in cellular responses to DNA damage? *Proc. Natl. Acad. Sci. USA*. 2007; 104:13213–13214. [PubMed: 17684092]
34. Cam HP, et al. Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat. Genet.* 2005; 37:809–819. [PubMed: 15976807]
35. Verdel A, et al. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science*. 2004; 303:672–676. [PubMed: 14704433]
36. Reinhart BJ, Bartel DP. Small RNAs correspond to centromere heterochromatic repeats. *Science*. 2002; 297:1831. [PubMed: 12193644]
37. Kato H, et al. RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science*. 2005; 309:467–469. [PubMed: 15947136]
38. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003; 19:185–193. [PubMed: 12538238]

Figure 1.

Extensive transcription of the *S. pombe* genome. **(a)** *S. pombe* probes were categorized according to genomic feature and classified into sense and antisense probes on the basis of gene orientation. **(b)** The percentage of the genome that is transcribed (*y* axis) at different signal thresholds (*x* axis), shown for two different units of measurement, genes or base pairs.

Figure 2.

Features of transcription and splicing in euchromatin. High-resolution views of different loci are depicted. The physical map is presented in a strand-specific manner, with currently annotated exons (gray boxes) and introns (gray lines). RNA signal intensities for probes along the top/forward (blue) and bottom/reverse (orange) strands are provided (probe centered). The background (dashed line) and the baseline (solid line) are shown for each strand. (a) A locus that generally conforms to current annotation. Brackets denote regions where signal drops below baseline on only one strand, near the TSS. (b) A locus that deviates from current annotation at several locations. Unannotated UTRs and antisense transcription are evident. (c) A typical locus, where signal intensity drops in accordance with annotated introns. (d) *snoR54* signal reflects its transcription and processing from an intron within *gua1*. *snoRNAs* are highly stable and typically accumulate to levels well beyond their host RNAs. (e) tRNA genes are multicopy and highly transcribed, resulting in a saturated signal. We note that because of the high sequence conservation among tRNAs, their signal cannot be uniquely attributed.

Figure 3.

Quantitation of transcription fragments and potential noncoding RNAs. Transfrags are quantified on the basis of threshold criteria. New transfrags identified by HybMap are denoted by hatched boxes. **(a)** Recently identified ncRNA TER1 is observed as a transfrag in our dataset. **(b)** Genome-wide analyses yield similar high numbers of transfrags in each condition. **(c-e)** Data from poly(A)-enriched RNA (overlaid blue line) and a Pol II ChIP (green line) provide evidence for the existence of potential ncRNAs. Transfrags are found in three general genomic environments: in intergenic regions **(c)**, in the antisense strand of known genes **(d)** and in intronic regions **(e)**.

Figure 4. Transcriptomes and loci derived from alternative growth conditions. Signals from alternative growth conditions (colored line) are overlaid on standard conditions (bars). (**a-f**) Under heat stress (**a,b**), minimal medium (**c,d**) or MMS treatment (**e,f**); the expression of particular genes (**a,c,e**) and transfrags (**b,d,f**) is induced.

Figure 5.

Features of antisense transcription. **(a)** Correlation of sense and antisense transcription. The 432 genes that differ at least fourfold in heat shock from standard conditions are plotted. For each gene, the median change for the sense strand was plotted against the antisense strand. **(b,c)** Signal in standard conditions (bars) overlaid with poly(A)-enriched data (overlaid blue line). Antisense transcription may be generally divided into two classes comprised of genes with negligible polyadenylated antisense transcripts (the vast majority) **(b)** and rare genes with polyadenylated antisense transcripts **(c)**.

Figure 6.

Poly(A) RNA shows divergent transcription and previously unknown RNA transcripts. **(a)** We depict the locus in Figure 2b overlaid with poly(A) data (solid purple line). Brackets denote three short transcripts diverging from *hsp60*, *cip2* and SPAC630.07c. Notably, these transcripts would not be distinguishable in total RNA, as a result of the prevalence of nonadenylated antisense. **(b,c)** A polyadenylated transcript clearly corresponds to each of the two transfrags most conserved among the related yeast species.

Figure 7.

Transcriptional features of heterochromatic loci and flanking regions. Signals derived from the top/forward and bottom/reverse strands are overlaid, to enable direct comparison. A colored solid line (underneath) indicates the locations of unique (green) and non-unique (purple) probes. **(a)** Transcriptional repression near telomeres. The physical map of the subtelomere of chromosome 1 is accompanied by maps of the heterochromatic mark H3K9me2 (ChIP data, black) and a mark correlated with transcribed ORFs, H3K36me3 (ChIP data, green). Notably, as the signal from H3K9me2 increases, intergenic transcription approaches the background level. **(b)** Transcription within the centromere of chromosome 1. Notably, both strands are transcribed in the regions flanking the *dh1* repeats. The opposite strand flanking the highly expressed tRNAs is transcribed. Results for chromosomes 2 and 3 are provided in Supplementary Figure 7.