Genome-Wide Distribution of Yeast RNA Polymerase II and Its Control by Sen1 Helicase

Eric J. Steinmetz,1 Christopher L. Warren,2 Jason N. Kuehner,1 Bahman Panahi,1 Aseem Z. Ansari,2,3 and David A. Brow1,⁎
1Department of Biomolecular Chemistry
School of Medicine and Public Health
2Department of Biochemistry
3The Genome Center
University of Wisconsin
1300 University Avenue
Madison, Wisconsin 53706

Summary

Functional engagement of RNA polymerase II (Pol II) with eukaryotic chromosomes is a fundamental and highly regulated biological process. Here we present a high-resolution map of Pol II occupancy across the entire yeast genome. We compared a wild-type strain with a strain bearing a substitution in the Sen1 helicase, which is a Pol II termination factor for noncoding RNA genes. The wild-type pattern of Pol II distribution provides unexpected insights into the mechanisms by which genes are repressed or silenced. Remarkably, a single amino acid substitution that compromises Sen1 function causes profound changes in Pol II distribution over both noncoding and protein-coding genes, establishing an important function of Sen1 in the regulation of transcription. Given the strong similarity of the yeast and human Sen1 proteins, our results suggest that progressive neurological disorders caused by substitutions in the human Sen1 homolog Senataxin may be due to misregulation of transcription.

Introduction

The model eukaryote Saccharomyces cerevisiae (brewer’s yeast) has approximately 6000 protein-coding genes, which are transcribed by RNA polymerase II (Pol II). These genes are arranged along 16 chromosomes that total 12 million base pairs in length (Goffeau et al., 1996). Scattered throughout the genome are additional genes coding for untranslated RNAs, which are transcribed by Pol I, Pol II, or Pol III. The efficient and properly regulated expression of all these genes requires strict control of RNA polymerase traffic along the chromosomes. Such control is particularly challenging in a compact genome like that of S. cerevisiae, in which genes are often separated by only a few hundred base pairs or less.

Functional engagement of a eukaryotic RNA polymerase with genomic DNA can be regulated by at least three general mechanisms. First, access to the DNA can be negatively regulated by chromatin structure. Such silencing or repression may be specific to a particular class of RNA polymerase or general to all three classes. Second, Pol I, Pol II, or Pol III can be actively recruited to a gene by distinct preinitiation complexes. Third, transcribing RNA polymerase can be removed from the DNA by the process of termination. The combination of these processes is expected to result in a specific pattern of RNA polymerase distribution across the genome of a cell in a given metabolic and developmental state. The genome-wide distribution of yeast Pol III has been reported (Harismendy et al., 2003; Roberts et al., 2003; Moqtaderi and Struhl, 2004), but the distribution of yeast Pol I and Pol II has not.

Because most genes are transcribed by Pol II, dysfunction of a factor that regulates Pol II engagement with DNA should have a strong effect on global RNA polymerase distribution and thus gene expression. We have identified the Sen1 protein as a Pol II termination factor that functions with the RNA-binding proteins Nrd1 and Nab3 on noncoding RNA genes (Steinmetz and Brow, 1996; Steinmetz et al., 2001). Sen1 is an essential 252 kDa superfamily 1 helicase that is most similar to Upf1/SMG-2 and SDE3/Armitage (see Figure S1 in the Supplemental Data available with this article online). The Sen1 ortholog from S. pombe has been shown to have RNA-RNA, RNA-DNA, and DNA-DNA helicase activity in vitro (Kim et al., 1999). Interestingly, mutations in the gene encoding the human Sen1 ortholog Senataxin can result in the diseases ataxia-ocular apraxia type 2 (AOA2; Moreira et al., 2004) and autosomal dominantly inherited juvenile amyotrophic lateral sclerosis (ALS4; Chen et al., 2004), which are progressive neurological disorders. While Sen1 has been strongly implicated in 3′ end formation of small nucleolar (sno) RNAs (Ursic et al., 1997; Rasmussen and Culbertson, 1998; Steinmetz et al., 2001), a more general role in Pol II termination has not been ruled out. Indeed, Sen1′s involvement in the autoregulation of Nrd1 mRNA level by transcription attenuation indicates that Sen1-dependent termination acts on at least one protein-coding gene (Steinmetz et al., 2001).

Here we determine the genome-wide distribution of Pol II and investigate its alteration by a substitution in the helicase domain of Sen1. We find that the sen1-E1597K mutation causes readthrough of both a snoRNA gene terminator and the poly(A) site/terminator of the protein-coding CYC1 gene, indicating that Sen1 function in 3′ end formation is not restricted to noncoding transcripts. A high-resolution map of the genomic distribution of Pol II in both wild-type and sen1 mutant strains confirms many known aspects of gene regulation but also has some surprising features. In addition to validating the role of Sen1 in snoRNA gene termination, our data suggest that Sen1 functions in termination on at least some short protein coding genes as well. Sen1-dependent attenuation of transcription was identified on several genes in addition to NRD1, including HRP1 and IMD2. Pol II silencing or repression appears to be specific to transcribed regions, suggesting that exclusion of Pol II may require prior transcription by Pol II. Our results provide a high-resolution view of the global distribution of Pol II in yeast and expand the list of genes that are targets of Sen1-dependent termination. Furthermore, our findings suggest a general molecular basis for the progressive neurological diseases AOA2 and ALS4.
Results

The sen1-E1597K Mutant Is Defective for Termination at Both an mRNA and a snoRNA Terminator

We previously isolated a heat-sensitive mutant allele of SEN1, sen1-E1597K, in a selection for readthrough of an artificial Nrd1-dependent terminator (Steinmetz and Brow, 1996; Figure S1). To directly test if the sen1-E1597K strain has a defect in Pol II termination, we performed transcription run-on assays (Steinmetz and Brow, 2003). Yeast cells bearing a reporter plasmid (Figure 1A) were permeabilized and pulse labeled with [α-32P]UTP. When no terminator is present between the two G-less cassettes in the plasmid, their labeling is roughly equal (Figure 1, CYC-ds). Addition of the CYC1 poly(A) site results in 98%-99% efficient termination of transcription in a wild-type strain (Figure 1, CYC-pAmax), consistent with its known function as a terminator (Osborne and Guarente, 1989; Russo and Sherman, 1989). However, in the sen1-E1597K mutant strain, a 10- to 20-fold increase in readthrough of the strong CYC1 terminator is observed. Presence of the terminator from the SNR13 snoRNA gene results in slightly weaker termination in a wild-type strain (Figure 1, CYC-pAmax), consistent with its known function as a terminator (Osborne and Guarente, 1989). However, in the sen1-E1597K mutant strain, a 10- to 20-fold increase in readthrough of the strong CYC1 terminator is observed. Presence of the terminator from the SNR13 snoRNA gene results in slightly weaker termination in a wild-type strain (SNR13-CYCds) and again 10- to 20-fold increased readthrough in the sen1-E1597K mutant strain. Thus, the sen1-E1597K mutation exhibits quantitatively similar effects on termination in response to the CYC1 poly(A) site and a snoRNA terminator.

To confirm that the sen1-E1597K mutation induces readthrough of the CYC1 terminator in a living cell, we used an ACT-CUP fusion reporter gene with the CYC1 poly(A) signal inserted in the intron (Steinmetz and Brow, 2003). Termination within the intron of the ACT-CUP reporter gene renders a SEN1 cup1Δ strain sensitive to 0.4 mM copper (Figure 2B). Strikingly, the same construct in a sen1-E1597K strain confers robust copper-resistant growth, consistent with strong through of the terminator. RNA analysis confirms that readthrough of the intronic CYC1 terminator occurs in the sen1-E1597K strain (Figure 2C). We conclude that function of the CYC1 poly(A) site in the ACT-CUP intron is impaired in the sen1-E1597K mutant strain.

The sen1-E1597K Mutation Alters the Global Distribution of Pol II on Chromosomes

To determine the genome-wide effects of the sen1-E1597K mutation on Pol II distribution, we used chromatin immunoprecipitation (ChIP). ChIP of Pol II is a sensitive measure of terminator readthrough on endogenous genes (Kim et al., 2004). We have expanded this approach to the entire yeast genome by applying the products of ChIP with antibody against the Rpb3 subunit of Pol II to a high-density microarray (ChIP-Chip). We analyzed both wild-type and sen1-E1597K cells grown in rich medium at 30°C, which is a semipermissive temperature for the sen1 mutant.

To allow an initial overview of the entire genome, we averaged the high-resolution data over intervals of 10,000 base pairs (Figure 3). Several features of the wild-type genomic Pol II distribution can be discerned at this resolution. First, it is not homogeneous. Peaks of Pol II enrichment correspond to one or a few genes with high Pol II density (Figure 3A). In the sen1 mutant, Pol II occupancy is reduced at many loci throughout the genome. Nevertheless, the wild-type and sen1 mutant traces are nearly identical at a number of loci (e.g., SED1; Figure 3B), suggesting there has not been an error in normalization (see Experimental Procedures). Thus, it appears that the sen7 mutation decreases Pol II occupancy on many expressed genes. Since Sen1 is a termination factor, this generalized reduction is presumably an indirect effect of the mutation, although we cannot exclude the possibility that Sen1 also functions in transcription initiation.
A second striking feature of the wild-type trace is that it rarely dips below an enrichment ratio equal to 1, and regions where it does often correspond to loci known to be silenced or repressed for Pol II transcription. For example, the FLO11 gene is known to be repressed in rich medium (Kuchin et al., 2002) and is indeed strongly depleted of Pol II (Figure 3C). The most Pol II-deficient expanse in the genome is the tandem repeat of ribosomal RNA genes (rDNA) on chromosome 12, which are transcribed by Pol I and Pol III (Figure 3D). Strikingly, the Pol II enrichment ratio reaches a value of 1 at a single position in the rDNA repeat, which is the location of a bidirectional Pol II promoter, called “E-pro,” that appears to be important for controlling the copy number of the rDNA repeats (Kobayashi and Ganley, 2005). Pol II is also strongly depleted at most telomeres, consistent with the well-established phenomenon of telomeric silencing (Tham and Zakian, 2002). When examined in detail, for example, at the right end of chromosome 15 (Figure 3E), most but not all telomeric genes exhibit far below the genomic average amount of crosslinking to Pol II, while intergenic regions show little or no depletion of Pol II. This pattern is consistent with the variable strength of silencing observed at different positions in natural telomeres (Fourel et al., 1999; Pryde and Louis, 1999) and implies that Pol II exclusion occurs primarily over transcribed regions (see Discussion). Surprisingly, depletion of Pol II is not observed at the silent mating type loci, HML and HMR, in either the low- or high-resolution data, which suggests that transcriptional silencing at these loci may be qualitatively or quantitatively different than at other loci.

Pol II Occupancy Measured by ChIP Correlates Well with Reported Transcription Activity at Most, but Not All, ORFs
To test if Pol II occupancy correlates with transcript synthesis, we compared relative Pol II ChIP enrichment in our wild-type strain with previous estimates of transcription activity for 3852 named open reading frames (ORFs) (Holstege et al., 1998). Linear regression of the data yielded a correlation coefficient of just over 0.5, consistent with a fairly linear relationship between the two data sets (Figure 4). At least some of the outliers can be explained by differences in the genotypes of the two strains (e.g., the strains are of opposite mating type). Interestingly, a large subset of ribosomal protein genes exhibits only moderate Pol II enrichment (1.5-2.5) and yet very high transcription frequency (>85 transcripts per hour). The reason that these genes fall so far off the regression line is unknown but could be due to a specialized mode of transcription that results in high transcript production at relatively low Pol II occupancy. Alternatively, there may have been a difference in growth conditions that specifically affected expression of these genes.

Perhaps the most striking outcome of the regression analysis is that the x intercept, equal to the Pol II enrichment ratio at 0 transcripts per hour, is 1.0. This result indicates that the absence of Pol II enrichment corresponds precisely to the absence of productive transcription. Given that the enrichment ratio is well below 1.0 at regions that are silenced or repressed (Figure 3), this intriguing finding implies that genes that are producing no stable transcripts but are not actively silenced or repressed are associated with significant levels of Pol II.

Surprisingly, the protein-coding gene with the lowest level of Pol II is TRP1 (Figure 4, inset). The strain we used for this analysis is a trp1 mutant, and the cells were grown in rich medium, so the lack of expression of the trp1 gene is expected. However, we know of no mechanism for the very strong repression implied by the striking Pol II depletion.

Readthrough of SNR Gene Terminators in the sen1 Mutant
There are about 80 SNR genes in yeast, five of which code for the spliceosomal snRNAs and the remainder for snoRNAs. Two are transcribed by Pol III, the U6 snRNA gene SNR6 and the snoRNA gene SNR52. Strikingly, SNR6 exhibits a very low level of Pol II association, equivalent to that seen over the rDNA (Figure 5A). Depletion of Pol II is not observed on SNR52 (Figure 5B) or any other known Pol III transcription units. Furthermore, previous studies indicated that SNR6 has a specialized chromatin structure (Kaiser et al. [2004] and references therein). Interestingly, in the sen1 mutant strain, the level of Pol II at SNR6 increases, as is the case for the rDNA.
repeats and most telomeric loci (see above). Thus, it appears that SNR6 is yet another silenced locus.

Of the SNR genes that are transcribed by Pol II, the best-characterized Nrd1/Sen1-dependent terminator is that of SNR13 (Morlando et al., 2002; Steinmetz and Brow, 2003; Carroll et al., 2004). The array data provide clear evidence of readthrough of SNR13 transcripts into the downstream TRS31 ORF in the sen1 mutant (Figure 5C). The Pol II peak that is centered over SNR13 in the wild-type strain shifts downstream and broadens to also cover the TRS31 gene in the sen1 mutant strain, consistent with the bicistronic SNR13/TRS31 transcript observed in strains defective for Sen1-dependent termination (Rasmussen and Culbertson, 1998; Steinmetz et al., 2001). Readthrough is clearly indicated in the red difference trace by a dip below the line, followed by a peak above the line. The SNR3 terminator was previously shown by nuclear run-on analysis to be read through in a nrd1 mutant strain (Steinmetz et al., 2001). There is also clear evidence from the array data for readthrough of the SNR3 terminator in the sen1 mutant (Figure 5D). Additional examples of readthrough on both rightward and leftward transcribed snoRNA genes are shown in Figures 5E–5G.

The peak of Pol II is centered over the downstream end of most snoRNA genes in the wild-type strain. This alignment reflects the fact that snoRNAs are synthesized as 3′-extended precursor transcripts (Allmang et al., 1999; van Hoof et al., 2000), and underscores the remarkable resolution of the ChIP-Chip procedure. Overall, readthrough is clearly observed in the sen1 mutant strain at 33 SNR loci. Readthrough may occur at additional SNR loci but is obscured by decreased total Pol II density and/or closely adjacent mRNA transcription units. Excluding the six SNR genes that are transcribed as part of pre-mRNA introns, and given that the remaining 68 SNR genes comprise only 57 transcription units, some of which are polycistronic, we can state that the majority of SNR loci that are autonomously transcribed by Pol II require Sen1 for proper termination. Interestingly, the longest polycistronic SNR gene, which encodes seven snoRNAs (snR72–78) and exceeds 1 kb
in length, exhibits no apparent readthrough in the sen1 mutant, nor does the longest (>1 kb) snRNA gene, SNR20. This observation suggests that long noncoding genes may not have Sen1-dependent terminators.

A Small Subset of Protein-Coding Genes Are Readthrough in a sen1 Mutant

Given our evidence that Sen1 is important for recognition of the CYC1 poly(A) site/terminator (Figures 1 and 2), we inspected the Pol II ChIP-Chip results to see if the endogenous CYC1 terminator is readthrough in the sen1 mutant strain. Enrichment of Pol II at the CYC1 locus drops sharply in the sen1 mutant strain, presumably due to decreased initiation (Figure 5H). Nevertheless, there is a clear shift of Pol II density downstream, resulting in a positive difference (red) peak. This result is consistent with readthrough of the CYC1 terminator at the endogenous locus in the sen1 mutant strain grown at semipermissive temperature.

To see if other protein-coding genes are similarly dependent on Sen1 for termination of transcription, we examined the Pol II enrichment plots for all 16 chromosomes. Figures S1–S5 show the clearest examples of known protein-coding genes that exhibit a readthrough pattern. Our results suggest that Sen1-dependent terminators are much less common on protein-coding genes than on noncoding genes. In light of the apparent absence of Sen1-dependent termination on the longest SNR genes (see above), it is interesting that the protein-coding genes that most clearly exhibit readthrough in the sen1 mutant are all quite short, with ORFs ranging in length from 204 to 554 base pairs (introns included). The median ORF length in yeast (introns excluded) is 1227 bp, and only 17% of ORFs are shorter than 555 bp.

Identification of Protein-Coding Genes Regulated by Sen1-Dependent Attenuation

We previously showed that the NRD1 gene is regulated by a negative feedback loop through a Nrd1-dependent terminator in its 5′ UTR. Decreased termination due to a mutation in the Nrd1, Sen1, Nab3, Ssu72, or Pti1 genes results in a 5- to 10-fold increase in Nrd1 mRNA (Steinmetz et al., 2001; Steinmetz and Brow, 2003; Dheur et al., 2003; Arigo et al., 2006). This attenuation process is clearly detected in the Pol II ChIP-Chip data (Figure 6A). In the wild-type strain, the peak of Pol II density is over the upstream end of the NRD1 ORF. In the sen1 mutant strain, the amount of Pol II associated with the gene appears unchanged, but the peak shifts so that it is centered over the NRD1 ORF.

Using this characteristic signature, we searched for other genes that are regulated by Sen1-dependent
attenuation. The two clearest examples we found are HRP1 and IMD2 (Figures 6B and 6D). A 3-fold increase in the mRNA from HRP1 in the sen1 mutant strain confirmed Sen1-dependent attenuation (Figure 6C).

*S. cerevisiae* Hrp1 (also called Nab4) is an mRNA-binding protein that shuttles between the nucleus and cytoplasm and influences poly(A) site selection (Henry et al., 1996; Kessler et al., 1997; Chen and Hyman, 1998; Minvielle-Sebastia et al., 1998). Hrp1 binds to AU dinucleotide repeat sequences (Kim Guisbert et al., 2005), which are present as efficiency elements in cleavage and polyadenylation signals (Guo and Sherman, 1996) and in Nrd1/Sen1-dependent snoRNA gene terminators (Steinmetz et al., 2006). Interestingly, Hrp1 mRNA has a long 5' UTR (Zhang and Dietrich, 2005) that is highly conserved in sequence between species of the genus *Saccharomyces* (Cliften et al., 2003; Kellis et al., 2003), consistent with the presence of cis-regulatory elements. Our results suggest that Hrp1 may be a component of the Sen1-dependent termination pathway and/or may use the pathway to autoregulate its expression.

The IMD2 gene codes for the enzyme IMPDH, which catalyzes the first dedicated step of GTP synthesis. Expression of IMD2 correlates inversely with intracellular guanine nucleotide levels (Exinger and Lacroute, 1992; Escobar-Henriques and Daignan-Fornier, 2001; Shaw et al., 2001). Guanine nucleotide regulation of IMD2 depends on a TATA box element that is highly conserved in sequence between species of the genus *Saccharomyces* (Cliften et al., 2003; Kellis et al., 2003), consistent with the presence of cis-regulatory elements. Our results suggest that Sen1 might be involved in regulatory processes other than attenuation, we looked for other types of changes in Pol II distribution in the sen1 mutant strain. Some examples are described here. The CPR1 gene encodes a component of the Set3 complex. We reproduced their results and mapped the precise positions of the transcription start sites in the presence or absence of 6-azauracil (6-AU), which reduces intracellular GTP pools (Exinger and Lacroute, 1992). In the absence of 6-AU, the major start sites are at G residues within 75–135 bp of the TATA box, and only the small RNA is observed (Figure 6E, lane 3, and Figure 6F, lane 1). In the presence of 6-AU, the major start site shifts to an A residue 200 bp downstream of the TATA box, Imd2 mRNA accumulates, and less of the small RNA is detected (Figure 6E, lane 4, and Figure 6F, lane 2). Intriguingly, the sen1 mutation results in accumulation of an elongated Imd2 mRNA that initiates at the upstream G starts (Figure 6E, lane 2 and data not shown). These results suggest that Sen1-dependent attenuation at IMD2 is regulated by alternative start site selection by Pol II "scanning" from the TATA box (Figure 6G). At high intracellular GTP concentration, the TATA box-proximal start sites are used and a Sen1-dependent terminator is recognized, producing short, unstable non-coding RNAs. At low intracellular GTP concentration, the upstream G starts are skipped and the more distal +1A start site is used. The resulting transcript lacks a functional Sen1-dependent terminator, allowing production of full-length mRNA. In the sen1 mutant strain, full-length mRNA can be made from the upstream starts, since the terminator is read through.

Other Possible Examples of Sen1-Mediated Gene Regulation

To investigate if Sen1 might be involved in regulatory processes other than attenuation, we looked for other types of changes in Pol II distribution in the sen1-E1597K strain. Two examples are described here. The CPR1 gene encodes a component of the Set3 complex.
histone methyltransferase/deacetylase complex, which is thought to repress transcription (Pijnappel et al., 2001). In the wild-type strain, a prominent peak of Pol II is present over CPR1, consistent with accumulation of an abundant transcript approximately 800 nucleotides in length (Figure 7A) and an estimated transcription frequency of 65 transcripts per hour (Holstege et al., 1998). A sharp peak of Pol II is also seen upstream of CPR1 in the wild-type strain, which could be due to transcription of the RPA14 gene. However, in the sen1-E1597K strain,
both peaks are absent and a new peak appears between them, and the *CPR1* mRNA is diminished. A possible explanation for this intriguing pattern is that a transcription unit exists antisense to and partially overlapping *RPA14*, here called “upstream CPR1” or uCPR1. uCPR1 uses a Sen1-dependent terminator, and, in the *sen1* mutant strain, readthrough of the terminator causes occlusion of the *CPR1* promoter and downregulation of Cpr1 mRNA. Such a mechanism would be similar to that which regulates expression of *SER3* (Martens et al., 2004). However, we cannot exclude other possible explanations for the complex Pol II distribution pattern at *CPR1*, and our model does not explain the unusually extended association of Pol II downstream of *CPR1*.

The *NPL3* gene encodes an mRNA-binding protein that has been implicated in several steps of mRNA biogenesis (Kim Guisbert et al. [2005] and references therein). Recently, Npl3 was shown to antagonize termination in response to a cleavage and polyadenylation site (Bucheli and Buratowski, 2005). As shown in Figure 7B, in the wild-type strain, Pol II enrichment drops off gradually downstream of the *NPL3* ORF until reaching the next ORF, whereupon the Pol II density drops precipitously. This pattern implies that the Npl3 mRNA may have an extended 3′UTR. Strikingly, in the *sen1* mutant strain, Pol II density drops sharply at the end of the *NPL3* ORF, suggesting increased termination efficiency and a truncated 3′UTR. RT-PCR analysis detects two major poly(A) sites for Npl3 mRNA, one that maps very close to the stop codon, and the other an additional ~200 nucleotides downstream (Figure 7B). The ORF-distal poly(A) site predominates in the wild-type strain, while the two sites are used about equally in the *sen1* mutant strain. Our results suggest that Sen1 facilitates antitermination at the *NPL3* ORF-proximal poly(A) site, which is of particular interest in light of accumulating evidence that Npl3 expression is autoregulated through the 3′UTR of its mRNA (M. Lund and C. Guthrie, personal communication).

Discussion

In the course of studying the function of the Sen1 helicase in transcription termination, we have mapped the location of RNA polymerase II across the entire yeast genome at high resolution, in both wild-type and *sen1* mutant strains. In addition to identifying additional Sen1-dependent Pol II terminators, our results illuminate many other aspects of regulation of Pol II function. As discussed below, our findings have implications for the mechanisms of regulation of Pol II transcription by attenuation and gene repression/silencing, exclusion of Pol II from Pol I and Pol III transcription units, and the influence of “intergenic” transcription on gene expression. To maximize the utility of our ChIP-Chip data for understanding gene expression in yeast, we have deposited the entire data set at the Saccharomyces Genome Database web site (SGD; http://www.yeastgenome.org).

Sen1 Directs Pol II Termination on Short Noncoding and Protein-Coding Genes

Our genome-wide analysis of Pol II occupancy confirms that the *SNR* genes are a primary target of the Sen1-dependent termination pathway. We also show that a small number of protein-coding genes, including the well-studied *CYC1* gene, have Sen1-dependent terminators in their 3′UTRs. Intriguingly, the protein coding
genes that most clearly show Sen1 dependence are all short, as are most snoRNA genes, raising the possibility that the Sen1-dependent termination pathway is most important for termination within 500 base pairs or so from the transcription start site. We hypothesize that this is because Sen1 (Figure S2) and its cofactors Nrd1 and Nab3 (Nedea et al., 2003) associate with the transcription complex at the promoter and so can efficiently recognize terminator elements near the 5’ ends of transcripts. In contrast, cleavage and polyadenylation factors appear to be recruited to the transcription elongation complex hundreds of base pairs downstream of the transcription start site (Buratowski, 2005) and so may be important for termination primarily on protein-coding genes greater than 500 base pairs in length.

Sen1-Dependent Attenuation Regulates Genes Other Than NRD1

Our genome-wide analysis confirmed the previously described Sen1/Nrd1-dependent autoregulation of Nrd1 gene expression (Steinmetz et al., 2001; Arigo et al., 2006) and uncovered two other genes that are regulated by Sen1-dependent attenuation, HRP1 and IMD2. Strikingly, all three genes share an extended and conserved TATA box sequence, the core of which is AAAAGTA TAAA. This sequence is found in few other genes, which suggests that genes regulated by Sen1-dependent attenuation may assemble a specialized preinitiation complex. We have identified substitutions in the Rpb3 and Rpb11 subunits of Pol II, located at the upstream end of the polymerase, that weaken Pol II’s response to Sen1-dependent terminators (Steinmetz et al., 2006). Conceivably, Rpb3 and Rpb11 bind to an extrinsic termination factor, the recruitment of which is facilitated by the extended TATA box. This factor could be Sen1 itself, or possibly Nrd1 or Nab3, all of which appear to be recruited at the transcription start site (see above).

In the case of the IMD2 gene, Sen1-dependent attenuation is coupled with alternative start site selection to confer regulation of IMD2 expression by intracellular GTP concentration. We have not determined the mechanism for alternative start site selection, but the fact that the major upstream start sites correspond to guanosine residues and the sole downstream start site is an adenosine residue strongly suggests that the start site is determined by initiating nucleotide concentration. While a related mechanism mediates UTP-dependent regulation of the codBA operon in E. coli (Qi and Turnbough, 1995), we know of no other example of such control in eukaryotes.

Exclusion of Pol II from Chromatin May Be Transcription Dependent

A striking feature of Pol II exclusion at telomeres is that it is much stronger over ORFs than over intergenic regions (Figure 3E), suggesting that transcriptionally inert Pol II is present at the promoters of silenced genes, consistent with an earlier study (Sekinger and Gross, 2001). The same phenomenon is observed at nontelomeric loci, e.g., at the ENA gene cluster (Figure S3). One interpretation of these results is that Pol II exclusion can only occur in transcribed regions, perhaps because the initiation of silencing requires transcription through the region to be silenced. A corollary of this hypothesis is that a “valley” of unusually low Pol II density in a region with no annotated genes may indicate the presence of an unidentified, silenced gene. Interestingly, downstream of ENA1 and upstream of the adjacent annotated gene, RS1810, there is such a valley of Pol II density (Figure S3), and such unannotated Pol II valleys occur elsewhere in the genome.

Depletion of Pol II over the rDNA repeat was expected because of the known silencing of Pol II transcription units placed in the rDNA (Bryk et al., 1997; Smith and Boeke, 1997). However, once again the nonuniformity of Pol II depletion across the repeat unit is striking. In addition to the E-pro promoter, a less-pronounced peak of Pol II is centered over the middle of the rDNA repeat in the vicinity of a cluster of ORFs (Figure 3D), including the TAR1 gene, which is known to be expressed (Coelho et al., 2002). Conceivably, bidirectional transcription from the E-pro promoter initiates silencing of the adjacent repeats, and terminators in the ORF-rich region in the middle of each repeat halt transcription and thus the initiation of silencing, making the chromatin in the vicinity of the TAR1 gene more accessible to Pol II.

The mechanism for depletion of Pol II over the U6 RNA gene, SNR6, is unknown, but given its magnitude seems likely to share characteristics in common with rDNA silencing. Although transcribed by Pol III, SNR6 is unusual in that it has a perfect consensus TATA box (Eschenlaer et al., 1993). Perhaps silencing is necessary to prevent Pol II transcription of SNR6. Notably, the first characterized Nrd1-binding site was found in the antisense strand of SNR6 (Steinmetz and Brow, 1996, 1998). This fact and the effect of the sen1 mutant on Pol II exclusion from SNR6 (Figure 5A) suggest the involvement of Nrd1/Sen1-dependent termination in silencing at this locus.

Although tRNA genes do not appear to exclude Pol II, Pol II levels are elevated precisely over a number of tRNA genes in the sen1 mutant strain (Figure S4). This result suggests that at least some tRNA genes can recruit Pol II and that recruitment is repressed, directly or indirectly, by Sen1. Engelke and coworkers have shown that active yeast tRNA genes can repress adjacent Pol II transcription units (Hull et al., 1994). They have proposed that such repression is due to association of tRNA genes with the nucleolus, which may exclude Pol II (Thompson et al., 2003; Wang et al., 2005). Interestingly, Culbertson and coworkers have shown that a mutation in SEN1 can disrupt nucleolar structure (Ursic et al., 1995). Conceivably, the increased Pol II we observe over some tRNA genes in the sen1 mutant strain is due to release of the genes from the nucleolus, exposing them to Pol II recruitment. Another possible explanation comes from the recent observation that some tRNA genes bind components of the Set3 complex, which normally acts to repress Pol II transcription (Mou et al., 2006). The decrease in expression of the Set3 complex subunit Cpr1 in the sen1 mutant (Figure 7A) may diminish binding of the Set3 complex to tRNA genes, allowing Pol II to bind.

Identification of Genes by Pol II Association

Our Pol II ChIP-Chip analysis could potentially identify genes that encode unstable RNAs and have thus previously escaped detection. An alternative approach for identifying unstable RNAs is genome-wide transcript
analysis with mutant strains lacking the Rrp6 subunit of the nuclear exosome (Davis and Ares, 2006; Wyers et al., 2005). These studies are likely to be relevant to our work, because the Nrd1/Nab3/Sen1 complex has recently been shown to be physically and functionally associated with the nuclear exosome (Vasiljeva and Buratowski, 2006). We note at least some overlap of our CHIP-Chip results with the exosome mutant expression array results. For example, we detect Pol II peaks corresponding to transcripts NEL025c, gBL04w, and NGR060w of Wyers et al. (2005) (Figures S5A–S5C) and iYCR006 and iYCRC06 of Davis and Ares (2006) (Figure SSD). Furthermore, we identify a Pol II peak corresponding to transcript iYR004C of Torchet et al. (2005), which associates with the H/ACA snoRNP-specific proteins Gar1 and Nhp2 (Figure S5E).

Some of the intergenic transcription units we have identified appear to produce previously unknown stable RNAs (data not shown). Further experiments will be required to confirm the identity and elucidate the function of these transcripts. Nevertheless, it is already apparent from our results that there are many uncharacterized Pol II transcription units in the yeast genome and that some likely code for unstable regulatory RNAs while others encode stable, presumably functional RNAs. Synthesis of members of both classes of RNA is affected by the sen1-E1597K mutation.

Conclusions and Prospects
Our data provide a high-resolution view of Pol II engagement across the entire yeast genome. The complexity of the picture we obtained is, at first, overwhelming. However, integration of these new data with past results from decades of study is expected to provide new insights into the mechanisms of global regulation of gene expression. One thing is apparent: Pol II distribution does not match our simple expectations. It will be of interest to determine how Pol II is excluded from loci like TRP1 and SNR6 and why it is present at numerous loci not annotated as genes.

The profound effect on Pol II distribution brought about by a single amino acid substitution in the Sen1 protein is striking and suggests that the symptoms of inherited disorders caused by mutations in the human Sen1 ortholog Senataxin are due to misregulation of gene transcription, perhaps via aberrant synthesis of regulatory noncoding RNAs. Determining the mechanism by which Sen1 influences Pol II distribution, including the relationship between transcription and gene silencing, is an important goal in developing an understanding of normal and pathological gene expression.

Experimental Procedures
Yeast Strains and Plasmids
The cup1.1 sen1-1E1597K (nrzd-1) strain and its cup1.1 sen1-1 parent 46a were described previously (Steinmetz and Brow, 1996). 46s is isogenic to 46a but has the opposite mating type. EJS101-9c is a NRZD haploid progeny from the diploid strain EJS101 (Steinmetz and Brow, 1996). Strain CD4KO (rpm2::kanMX, imd3::his3, imd3::kanMX) was provided by M. Ares and was described previously (Davis and Ares, 2006). CD4KO was transformed with pRS316 to provide a functional copy of the URA3 gene, which enhanced the response to 6-AU.

ACT-CUP reporter plasmids pGAC24-JUTR and pGAC24-CYC83F-JUTR were derived from pGAC24 and pGAC24-CYC83F (Steinmetz and Brow, 2003) by replacement of all CUP1 3′-flanking sequences downstream of the stop codon with a Sall restriction site as described (Steinmetz et al., 2006). The PGK1 poly(A) site remains intact.

Transcription Run-On Analysis
Transcription run-on analysis using dual G-less cassette constructs was performed as described previously (Steinmetz and Brow, 2003), except that cultures of wild-type and sen1-1E1597K mutant cells were grown continuously at 30 °C in -leu media.

RNA Analysis
Total cellular RNA was prepared as described (Steinmetz and Brow, 1996). For northern blots, 10–20 μg/ lane of RNA was run on a 1% agarose/0.8 M formaldehyde gel and blotted to Zeta-probe membrane (Bio-Rad). The HRP1 probe was a random-primed 800 bp DNA. The scR1 probe is an oligonucleotide complementary to positions 75–92. The single-stranded IMD2 probe was generated as described (Davis and Ares, 2006). The CPR1 probe was a random-primed DNA fragment.

Primers extension of 10 μg of RNA was done with 32P-labeled oligonucleotides complementary to nucleotides 72–92 of the Imd2 5′UTR and 56–75 of scR1 RNA. Reaction conditions were as described in Eschenhauer et al. (1993), except that annealing was at 45 °C. Sequencing ladders were generated using the Sequitherm EXCEL II DNA sequencing kit (Epicentre). The cDNA products were electrophoresed on 6% polyacrylamide, 8.3 M urea gels. Gels were visualized with a Storm Phosphorimager (Amersham Biosciences).

RT-PCR was performed using 5 μg of RNA from 46a and nrzd-1 strains and SuperScript II Reverse Transcriptase (Invitrogen) according to the manufacturer’s protocol. First-strand cDNA synthesis used 100 pmol of T16-BSG1 primer (Steinmetz et al., 2006). The cDNA was amplified by PCR for 30 cycles using Taq DNA Polymerase (Invitrogen), T16-BSG1 reverse primer, and forward primers complementary to nucleotides 522–545 or 1027–1048 of the NPL3 ORF.

CHIP-Chip Analysis of Pol II Distribution
Crosslinked chromatin from SEN1 (46s) and sen1-1E1597K (nrzd-1) cultures grown in YEPD at 30 °C was immunoprecipitated with monoclonal anti-Rpb3 1Y26 (Neoclone) as described in Supplemental Material. First-strand cDNA synthesis was performed as described previously (Steinmetz and Brow, 2003) by replacement of all dATP, dCTP, and dGTP, yielding 4–5 μg modified DNA, 80% of which was then coupled to input DNA and Cy5 coupled to input DNA. Hybridization was performed with the dye labels swapped; i.e., with Cy3 coupled to input DNA and Cy5 coupled to input DNA. Hybridization was carried out for 16 hr at 45 °C.

Microarrays were synthesized by NimbleGen Systems, Inc., and contain ~194,000 different 24-mer oligonucleotides that “tile” the Saccharomyces cerevisiae S288C genome. Sequences were spaced every 60–120 base pairs along the genome, alternating between the Watson and Crick strands. Probe sequences were chosen within these 60 bp range by NimbleGen to maximize each probe’s genomic uniqueness. The microarray was based on SGD genomic sequence as of June 1, 2003. Analysis of the Chip-Chip data converted those probe sequences to SGD version 11/1/05 of the genomic sequence and annotation.
Microarray Data Analysis

Microarrays were scanned using the GenePix 4000B scanner and extracted using GenePix 4.1 software. Cy3 and Cy5 values for each array were normalized by multiplying each feature’s Cy5 fluorescence intensity by the ratio of the median Cy3 value to the median Cy5 value. For each microarray feature, the average ratio of IP/input fluorescence (after normalization) from the two hybridizations was determined, and the natural log of this ratio was calculated. Data were ordered by genomic position, and values averaged over a sliding window of nine adjacent probes were plotted on the center probe position. Differences between the wild-type and sen1 mutant samples were expressed as ln(IP/input)sen1 − ln(IP/input)wt and were averaged and plotted as above.

The low-resolution average density plot (Figure 3) was calculated by linearly interpolating both the wild-type and sen1 ChIP-Chip ratios for every base pair based on the ChIP-Chip ratio of the immediately preceding probe and succeeding probe (except for base pairs preceding the position of the first ChIP-Chip probe on a given chromosome or succeeding the last ChIP-Chip probe). The average intensity for each 10 kb section of each chromosome was then determined, and the natural log of this ratio was calculated. Data for every base pair based on the ChIP-Chip ratio of the immediately preceding probe and succeeding probe (except for base pairs preceding the position of the first ChIP-Chip probe on a given chromosome or succeeding the last ChIP-Chip probe). The average intensity for each 10 kb section of each chromosome was then averaged and plotted as above.

The scatter plot of ORF Pol II density versus ORF transcription frequency (Figure 4), Pol II density for a given ORF was calculated as the average interpolated data value for all base pairs in the ORF.

Supplemental Data

Supplemental Data include supplemental text, five figures, and Supplemental References and can be found with this article online at http://www.molecule.org/cgi/content/full/24/5/735/DC1/.

Acknowledgments

We thank M. Kim and S. Buratowski for sharing ChIP procedures and PCR primer sequences; K. Arndt for PCR primer sequences; A. Roopra for advice on chromatim sonication procedures; N. Thompson and R. Burgess for Pol II antibody; David Frisch, Yu Qiu, and Elenita Kanin for assistance in DNA microarray procedures; M. Ares and D. Reines for strains; M. Ares, D. Engelke, C. Fox, R. Gourne, C. Guthrie, J. Keck, and M. Lund for discussions; and C. Fox and R. Landick for comments on the manuscript. This work was supported by NIH-PHS grant GM44665 (to D.A.B.), grants from the March of Dimes Foundation and USDA-Hatch (to A.Z.A.), and NIH-PHS grant R44HG02193 to NimbleGen for microarrays. C.L.W. was supported in part by a Petermann Fellowship from the UW-Madison Department of Biochemistry.

Received: May 19, 2006
Revised: August 12, 2006
Accepted: October 13, 2006
Published: December 7, 2006

References


Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., Mozoni, I. (2002). Functional analysis of yeast snoRNA and snRNA 3′-end formation by the essential yeast protein Nrd1 protein. Biochemistry 41, 571–574. (Note: this is a citation error. The correct citation is likely missing.)


Accession Numbers

Microarray pole positions and data were deposited at http://www.ncbi.nlm.nih.gov/geo under accession numbers GPL4563 and GSE6293, respectively.