Chromatin- and Transcription-Related Factors Repress Transcription from within Coding Regions throughout the Saccharomyces cerevisiae Genome

Citation
Cheung, Vanessa, Gordon Chua, Nizar N. Batada, Christian R. Landry, Stephen W. Michnick, Timothy R. Hughes, and Fred Winston. 2008. Chromatin- and Transcription-Related Factors Repress Transcription from within Coding Regions throughout the Genome. PLoS Biology 6: e277.

Published Version
doi:10.1371/journal.pbio.0060277

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:11181061

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Chromatin- and Transcription-Related Factors Repress Transcription from within Coding Regions throughout the Saccharomyces cerevisiae Genome

Vanessa Cheung1, Gordon Chua2*,a, Nizar N. Batada1†c, Christian R. Landry3, Stephen W. Michnick3, Timothy R. Hughes2, Fred Winston1*†

1 Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America, 2 Banting and Best Department of Medical Research, University of Toronto, Toronto, Canada, 3 Département de Biochimie, and Centre Robert-Cedergren, Bio-Informatique Génomique, Université de Montréal, Montréal, Canada

Previous studies in Saccharomyces cerevisiae have demonstrated that cryptic promoters within coding regions activate transcription in particular mutants. We have performed a comprehensive analysis of cryptic transcription in order to identify factors that normally repress cryptic promoters, to determine the amount of cryptic transcription genomewide, and to study the potential for expression of genetic information by cryptic transcription. Our results show that a large number of factors that control chromatin structure and transcription are required to repress cryptic transcription from at least 1,000 locations across the S. cerevisiae genome. Two results suggest that some cryptic transcripts are translated. First, as expected, many cryptic transcripts contain an ATG and an open reading frame of at least 100 codons. Second, several cryptic transcripts are translated into proteins. Furthermore, a subset of cryptic transcripts tested is transiently induced in wild-type cells following a nutritional shift, suggesting a possible physiological role in response to a change in growth conditions. Taken together, our results demonstrate that, during normal growth, the global integrity of gene expression is maintained by a wide range of factors and suggest that, under altered genetic or physiological conditions, the expression of alternative genetic information may occur.

Citation: Cheung V, Chua G, Batada NN, Landry CR, Michnick SW, et al. (2008) Chromatin- and transcription-related factors repress transcription from within coding regions throughout the Saccharomyces cerevisiae genome. PLoS Biol 6(11): e277. doi:10.1371/journal.pbio.0060277

Introduction

Several recent studies have demonstrated that transcription occurs across large eukaryotic genomes in a much more widespread and complex pattern than previously imagined. The recent findings of the ENCODE project, which analyzed transcription of 1% of the human genome [1], demonstrated the use of multiple transcription start sites and transcription across most sequences, including intergenic regions (reviewed in [2]). Many other recent studies have also identified extensive transcription across human sequences, including antisense transcription (reviewed in [3–5]). Similarly, in Drosophila melanogaster, recent studies estimate that 85% of the genome is transcribed, with extensive intergenic transcription and multiple transcription start sites [6]. Although the function of most of this pervasive transcription is currently not understood, there is evidence that a significant amount of it is regulated, raising the possibility that it is required for previously unknown modes of regulation or that it allows the expression of previously undetected genetic information [3–5]. Strong precedents exist for regulatory roles for intergenic transcription (for example, [7,8]; see [4,9] for recent reviews).

In Saccharomyces cerevisiae, similar to larger eukaryotes, several recent genome-wide studies have demonstrated widespread transcription across coding and noncoding regions [10–15]. In a small number of cases in S. cerevisiae, intergenic transcription [16–18], antisense transcription [19,20], and initiation within coding regions [21,22] have been shown to play biological roles. In addition to transcriptional events that occur in wild-type strains, other studies have revealed that transcription initiation can be activated from within coding regions in particular mutants [23,24]. Such initiation was originally observed in strains containing mutations in SPT6 and SPT16, which encode conserved, essential transcription factors believed to be involved in nucleosome disassembly and assembly [23–27]. In an spt6 mutant, the use of a transcription start site within the FLO8 gene was shown to be dependent upon a consensus TATA element within the FLO8 coding sequence, suggesting the existence of a cryptic promoter within FLO8 that is normally repressed in a wild-
Recent studies have shown that much more of the eukaryotic genome is transcribed into RNA than previously thought. In *Saccharomyces cerevisiae*, when particular factors are defective, cryptic promoters within several coding regions become active and produce shorter transcripts corresponding to the 3' portions of genes. (Transcription proceeds from the 5' end of genes to the 3' end.) A comprehensive analysis of cryptic transcription identified the factors that normally repress this event. We find that at least 50 factors, many involved in chromatin structure and transcription, are required to repress cryptic transcription. Other results suggest that the potential for cryptic transcription is widespread, initiating from at least 1,000 locations across the *S. cerevisiae* genome. In mutants in which cryptic transcripts are produced, some of the transcripts are translated into proteins not normally made in unmodified, wild-type cells. Finally, in wild-type cells, a subset of cryptic transcripts is transiently induced following a nutritional shift, suggesting a possible role for cryptic transcription. Taken together, our results demonstrate that the normal pattern of gene expression is maintained by a wide range of factors and suggest that, under altered genetic or physiological conditions, the expression of alternative genetic information may occur.

**Results**

**Comprehensive Identification of Mutants Permissive for Cryptic Transcription**

Previous results have shown that cryptic promoters are active in several mutants that impair transcription and chromatin structure. However, no systematic isolation of cryptic initiation mutants has been performed. To comprehensively identify factors that regulate cryptic promoters, we first constructed a reporter to allow easy detection of activation of the FLO8 cryptic promoter. In this reporter, we replaced the region of *FLO8* 3' of the cryptic transcription start site with the HIS3 coding sequence (Figure 1A; Materials and Methods). The HIS3 coding sequence was inserted out-of-frame with respect to the FLO8 coding sequence, using the first ATG within FLO8 that follows the cryptic start site. As this ATG is in the +2 reading frame, functional HIS3 mRNA can only be made by transcription initiation at the FLO8 cryptic start site (Figure 1A). In one version of this reporter, the normal FLO8 promoter was replaced with the GAL1 promoter to allow regulation of full-length FLO8-HIS3 transcription by growth on different carbon sources and in a second version, the wild-type FLO8 promoter was maintained. Both growth assays on plates lacking histidine and northern analysis demonstrated that the FLO8-HIS3 fusion constitutes a sensitive reporter for mutants that allow cryptic initiation (Figure 1B and 1C).

Using FLO8-HIS3, we employed two methods to identify mutants that are permissive for cryptic initiation: direct selection and a screen of the *S. cerevisiae* nonessential deletion set (Materials and Methods). Direct selection was valuable for identification of strong mutations that are not in the deletion set, in particular, mutations in histone genes, described below. The deletion set screen allowed systematic testing of all nonessential genes. Overall, we identified mutations in 50 genes that allow cryptic initiation at FLO8-HIS3 (Table 1). These 50 mutants are permissive for the FLO8 cryptic promoter to varying degrees and several are dependent upon expression from the upstream GAL1 promoter in the FLO8-HIS3 reporter (Figure 2A). Overall, the majority of genes identified encode histones, regulators of histone gene expression, histone chaperones, and other factors implicated in transcriptional control.

Among this large collection of mutants, histone H3 mutants are of particular interest as some identify previously unstudied changes in H3 that may play roles in transcription.
At least 1,000 Cryptic Transcripts are Produced in *spt6* and *spt16* Mutants

Previous studies of cryptic initiation in an *spt6-1004* mutant identified only a few genes with cryptic promoters [23]. However, the frequency at which they were found among a small set of genes tested suggested that cryptic promoters may be widespread. To test this possibility, we assayed for cryptic transcription within ORFs on a genome-wide scale by microarray analysis. In these experiments, we compared mRNA from a wild-type strain to that from an *spt6-1004* mutant, using microarrays with six probes across each coding region (Materials and Methods). Using a stringent threshold (Materials and Methods), our results suggest that out of the 5,689 ORFs represented on the microarray, at least 960 genes (17%) have active cryptic transcription in the *spt6-1004* mutant (Figure S1; Table S3). As detailed in Materials and Methods, this method may unavoidably be biased towards
identifying cryptic transcripts from genes with lower transcription levels, likely resulting in an underestimate of the actual number of cryptic transcripts (Materials and Methods; Figure S2). In support of the ability of the microarrays to identify genes with cryptic transcripts, we used northern analysis to test five genes predicted by the microarrays to have cryptic transcripts and found that all five indeed produce short transcripts (Figure 4A).

To test whether another mutant permissive for cryptic transcription allows production of the same large set of cryptic transcripts, microarray analysis was performed on the temperature-sensitive spt16 mutant, spt16-197. These experiments identified approximately 1,130 genes predicted to have cryptic transcripts in the spt16-197 mutant (Table S4). Between the spt6-1004 and spt16-197 results, there is a striking overlap (correlation coefficient \( r = 0.83 \), Figure 4B and 4C), indicating that these two mutants affect cryptic transcription similarly at most genes. Taken together, these results strongly suggest that approximately one sixth of all S. cerevisiae genes produce detectable cryptic transcripts in spt6 and spt16 mutants.

To determine whether the genes that produce cryptic transcripts share any particular traits, we examined several different characteristics of the genes that we identified in the spt6 and spt16 microarray experiments as having cryptic transcripts. With respect to the length of coding regions, the average length of the genes with cryptic transcription in both spt6 and spt16 mutants is 2.4 kb, significantly longer than the average length of the 5,869 genes on the microarray (1.5 kb; Wilcoxon rank-sum test, \( p \)-value \( < 2 \times 10^{-10} \)). The majority of genes with cryptic transcription also have lower transcriptional frequencies (for spt6-1004, average = 2.46 mRNA/hour [\( p \)-value \( < 2 \times 10^{-10} \)] and for spt16-197, 1.93 mRNA/hour [\( p \)-value \( < 2 \times 10^{-10} \)]) when compared with the whole genome (average = 7.57 mRNA/hour) [50]. The enrichment for longer genes with lower transcriptional frequencies was expected, as these two characteristics correlate, and our method for detection of cryptic transcripts enriched for genes with lower transcription levels.

In addition, we focused on TATA elements, both within coding regions and in 5’ noncoding regions. Since cryptic initiation within the FLO8 coding region depends on the presence of a TATA element [23], we first tested whether genes showing cryptic transcription are enriched for those with TATA motifs in their coding sequence. We searched for the TATA consensus sequence in S. cerevisiae, TATA(A/T)(A/T)(A/T) (Table S5). Given that our set of genes was enriched for those that are longer, we also examined whether these findings were still significant when corrected for gene size (longer genes are more likely to contain TATA motifs by chance) and found that they were (Figure S3). Thus, the genes with cryptic promoters identified by the spt6-1004 and spt16-197 microarray results suggest that cryptic transcription tends to be located in coding regions that contain TATA consensus sequences. We also classified the normal promoters of genes with cryptic promoters as to whether they have a TATA element or not. Genes with TATA elements tend to display more cell-to-cell and strain-to-strain variation in expression [52–57].

### Table 1. Genes Required for Repression of Cryptic Promoters

| Category                        | Gene | Mutant Phenotype | Phoebe Phenotype |
|---------------------------------|------|-----------------|-----------------|
| Histones                        | HHT1 | +++             |                 |
|                                 | HHT2 | +++             |                 |
|                                 | HTA1-HTB1 | +++         |                 |
| Histone regulation              | SPT10 | +++            |                 |
|                                 | SPT21 | +++            |                 |
|                                 | HIR1 | +++            |                 |
|                                 | HIR2 | +++            |                 |
|                                 | HIR3 | +++            |                 |
|                                 | HPC2 | +++            |                 |
| Chromatin assembly and remodeling| SPT6 | +++            |                 |
|                                 | SPT16 | +++           |                 |
|                                 | RTT106 | +++       |                 |
|                                 | ASP1 | +++            |                 |
|                                 | CHD1 | +              |                 |
|                                 | ISW1 | +              |                 |
| Rpd3S-mediated histone deacetylation | CTK1 | +++           |                 |
|                                 | CTK2 | +++            |                 |
|                                 | CTK3 | ++             |                 |
|                                 | SET2 | ++             |                 |
|                                 | RPD3 | +              |                 |
|                                 | SIR3 | +              |                 |
|                                 | RCO1 | +              |                 |
|                                 | EAF3 | +              |                 |
|                                 | UME1 | +              |                 |
| Histone deacetylation           | HST4 | +              |                 |
|                                 | HD2A | +              |                 |
| Transcription elongation factors | BUR1 | ++             |                 |
|                                 | BUR2 | ++             |                 |
|                                 | SPT4 | +              |                 |
|                                 | SPT5 | +++            |                 |
|                                 | CDC73 | +            |                 |
|                                 | RTF1 | +              |                 |
|                                 | SPT2 | +              |                 |
| Mediator components             | SIN4 | +              |                 |
|                                 | SRI8 | +              |                 |
|                                 | SRI9 | +              |                 |
|                                 | SRI10 | +         |                 |
| Proteasome function             | PRE9 | +              |                 |
|                                 | UMP1 | +              |                 |
| DNA replication and repair factors | HUR1 | +            |                 |
|                                 | PMS1 | +              |                 |
|                                 | RNR4 | +              |                 |
| Other categories                | ATG11 | ++            |                 |
|                                 | DIG1 | +              |                 |
|                                 | KEK2 | +              |                 |
|                                 | Map1 | +              |                 |
|                                 | PIR1 | +              |                 |
|                                 | PIR2 | +              |                 |
|                                 | SSD1 | +              |                 |
|                                 | YBR246W | +    |                 |

Scoring of mutant phenotypes using FLO8-HIS3 is as follows: +++ indicates mutants with a strong His\(^+\) phenotype (His\(^+\) after 2 d of growth); ++, mutants with a medium His\(^+\) phenotype (His\(^+\) after 4 d of growth); and +, mutants with a weak His\(^+\) phenotype (His\(^+\) after 5 d or more of growth).

An rpd3 mutant was not tested with FLO8-HIS3; its score is based on northern analysis. doi:10.1371/journal.pbio.0060277.t001
Many Cryptic Transcripts Expressed in an spt6 Mutant Are Translated

Given the large number of cryptic transcripts, it seemed likely that many of them would have the potential to encode proteins. We examined the potential for cryptic transcripts to be translated by mapping all ATGs in the three reading frames downstream of the 5' most limit of transcription initiation established in the spt6-1004 microarray analysis. For
Strains upon a Nutritional Shift

A Subset of Cryptic Transcripts Are Expressed in Wild-Type Strains upon a Nutritional Shift

The expression of the cryptic transcripts that we have identified is normally repressed in wild-type strains when cells are grown in rich medium. If some of the cryptic transcripts serve a biological function, however, they might be expressed in a wild-type background under particular growth conditions. To screen for such an effect, we used northern analysis to assay the transcription of 16 genes with cryptic transcripts under 20 different growth conditions. Most of these genes were selected from those shown to express proteins from cryptic promoters are translated to produce alternative, shorter proteins. The functions of these proteins are likely to be different from the full-length proteins because they often lose predicted protein domains.

By western analysis of proteins that have cryptic transcripts, we screened 146 genes that are predicted to encode short peptides, while, as expected, the full-length proteins if the cryptic transcripts were to be translated: 820, 825, and 731 ORFs in frames +1, +2, and +3, respectively. However, the two alternative reading frames primarily encode short peptides, while, as expected, the +1 frame encodes much longer sequences.

To test directly whether genes with cryptic transcripts express proteins, we screened 146 genes that are predicted to have cryptic promoters and that have at least one internal ATG codon in the +1 frame located 3’ of the predicted cryptic start site. To screen these strains, we used the tandem affinity purification (TAP)-tagged set of S. cerevisiae strains in which each ORF is fused at its 3’ end to a sequence encoding the TAP epitope tag [58]. The TAP-tagged strains corresponding to the 146 selected genes were crossed to an spt6-1004 strain to obtain TAP-tagged versions in both SPT6+ and spt6-1004 backgrounds. These strains were then screened by western analyses using an antibody recognizing the TAP tag to determine whether any altered proteins are made in the spt6-1004 strains. We note that this method will only detect proteins produced by translation in the same reading frame as the full-length protein, because it requires that the TAP epitope tag be expressed. Our results show that 20 of the 146 genes tested produced a detectable shorter protein in the spt6-1004 mutant but not in the SPT6+ strain (Table S7, examples shown in Figure 5A). The short proteins were all in the size range predicted by the microarray results, and several of them encode domains with known activities lacking their normal amino-terminal sequences (Figure S5). Northern analysis of these genes verified that corresponding short transcripts of the appropriate sizes were indeed expressed in the spt6-1004 mutant (Figure 4A; unpublished data).

To verify that the short proteins were produced by translation initiation from their corresponding short transcripts and were not simply degradation products of the full-length proteins, we analyzed the expression of short proteins made from two genes, APM2 and PUS4. For each of these genes, we constructed and analyzed mutations that alter the initiation codon for both the normal, full-length protein and for the shorter protein and analyzed each by western analysis. Our results show that mutation of the normal ATG initiation codon eliminated expression of the full-length protein, but had no effect on expression of the short protein expression (Figure 5B, lanes 7 and 15). Furthermore, mutation of the internal ATG specifically abolished expression of the short protein (Figure 5B, lanes 6 and 14). We also observed that this mutation in APM2 resulted in apparent degradation products (Figure 5B, lane 6), perhaps due to the amino acid change in the mutant protein. This mutation in APM2 also causes increased expression of the full-length Apm2 protein specifically in the spt6-1004 mutant and may be due to changes in either mRNA or protein stability. Taken together, these results demonstrate that at least a subset of transcripts expressed from cryptic promoters are translated to produce alternative, shorter proteins. The functions of these proteins are likely to be different from the full-length proteins because they often lose predicted protein domains.

Figure 3. Western Analysis of Histone H3 K36 Methylation Levels in Cryptic Initiation Mutants

Whole-cell extracts were prepared from cells grown at 30 ºC, except for strains FY2425 and FY347, which were shifted to 37 ºC for 80 min as indicated. Probes for the western analyses were generated with antibodies specific for total histone H3, dimethylated H3 K36, and trimethylated H3 K36, and western analyses using an antibody recognizing the TAP tag to determine whether any altered proteins are made in the spt6-1004 strains. We note that this method will only detect proteins produced by translation in the same reading frame as the full-length protein, because it requires that the TAP epitope tag be expressed. Our results show that 20 of the 146 genes tested produced a detectable shorter protein in the spt6-1004 mutant but not in the SPT6+ strain (Table S7, examples shown in Figure 5A). The short proteins were all in the size range predicted by the microarray results, and several of them encode domains with known activities lacking their normal amino-terminal sequences (Figure S5). Northern analysis of these genes verified that corresponding short transcripts of the appropriate sizes were indeed expressed in the spt6-1004 mutant (Figure 4A; unpublished data).

To verify that the short proteins were produced by translation initiation from their corresponding short transcripts and were not simply degradation products of the full-length proteins, we analyzed the expression of short proteins made from two genes, APM2 and PUS4. For each of these genes, we constructed and analyzed mutations that alter the initiation codon for both the normal, full-length protein and for the shorter protein and analyzed each by western analysis. Our results show that mutation of the normal ATG initiation codon eliminated expression of the full-length protein, but had no effect on expression of the short protein expression (Figure 5B, lanes 7 and 15). Furthermore, mutation of the internal ATG specifically abolished expression of the short protein (Figure 5B, lanes 6 and 14). We also observed that this mutation in APM2 resulted in apparent degradation products (Figure 5B, lane 6), perhaps due to the amino acid change in the mutant protein. This mutation in APM2 also causes increased expression of the full-length Apm2 protein specifically in the spt6-1004 mutant and may be due to changes in either mRNA or protein stability. Taken together, these results demonstrate that at least a subset of transcripts expressed from cryptic promoters are translated to produce alternative, shorter proteins. The functions of these proteins are likely to be different from the full-length proteins because they often lose predicted protein domains (Table S8).
Previous studies have shown that a nutritional shift from rich to minimal media causes other transient effects with very similar kinetics to what we have observed. Among these effects is the induction of translation of the transcription factor Gcn4 [59–61], which occurs in a Ras2-dependent fashion [62]. We therefore tested whether either Gcn4 or Ras2 plays a role in the expression of cryptic transcripts that we observe by assaying \textit{gcn4} \textit{D} and \textit{ras2} \textit{D} mutants during a nutritional shift. Although \textit{gcn4} \textit{D} did not affect cryptic transcript levels (unpublished data), our results showed that the expression of the \textit{CHS6} and \textit{FLO8} cryptic transcripts upon the nutritional shift was strongly Ras2-dependent, whereas the expression of the \textit{SPB4} cryptic transcript appeared to be largely Ras2 independent (Figure 6, lanes 7–10). These results also suggest that the cryptic initiation induced at \textit{CHS6} and \textit{FLO8} after the nutritional shift is not simply the result of the increased expression of the full-length transcript seen for both genes following the media shift. Even though full-length expression of \textit{CHS6} and \textit{FLO8} is still greatly increased following the shift in the \textit{ras2} \textit{D} mutant, cryptic transcripts are not expressed, indicating some form of regulation of the cryptic promoters under these conditions. Thus, our results suggest that a subset of cryptic promoters can be specifically activated upon a nutritional shift in a Ras2-dependent fashion.

**Discussion**

In this work, we have investigated cryptic transcription and its consequences in \textit{S. cerevisiae} on a genome-wide scale. Our results have established that a large number of chromatin- and transcription-related factors are required to repress widespread cryptic transcription from within coding regions throughout the \textit{S. cerevisiae} genome. Most of the cryptic transcripts contain ORFs, and our results suggest that when these cryptic transcripts are expressed, such as in an \textit{spt6}
mutant, many of them are translated to produce proteins that are not normally made. Thus, loss of Spt6 causes a dramatic change in the mRNAs and proteins produced genome-wide. Furthermore, a small subset of cryptic transcripts have been shown to be modestly expressed in wild-type strains during a nutritional shift. Taken together, these results demonstrate the widespread existence of cryptic transcription and the expression of alternative genetic information in *S. cerevisiae*.

Several results strongly suggest that multiple mechanisms control the expression of cryptic transcripts. Below, we discuss these possible mechanisms in terms of distinct classes of cryptic promoters. We note that our microarray results have established widespread cryptic transcription, but have not demonstrated that these transcripts all arise from cryptic promoters. However, based on our earlier studies of the *FLO8* and *SPB4* genes ([23] and unpublished data), we think it is likely that most or all of the cryptic transcripts identified are the result of activation of cryptic promoters. Testing this possibility will be the focus of future investigations. First, the mutants identified in this study vary greatly in their strength of cryptic initiation, based both on the *FLO8-HIS3* reporter and on northern analysis. Second, one of the most permissive mutants for cryptic initiation, *spt6-1004*, is known to impair at least two features of normal transcription elongation that individually contribute to repression of cryptic promoters: histone H3 K36 methylation [28,47,48] and the recruitment of the transcription factor Spt2 [35]. Consistent with this observation, both *set2A*, which abolishes histone H3 K36 methylation, and *spt2A* are less permissive for cryptic initiation than is *spt6-1004* (our results and [28,31,34,35]). In addition to these effects, *spt6-1004* likely causes other effects on chromatin structure [23,26]. Third, our results also showed that most mutations that allow cryptic initiation do not impair H3 K36 di- or trimethylation; therefore, loss of this histone modification is not the sole mechanism by which cryptic promoters are derepressed. This conclusion is consistent with recent studies that showed enhanced cryptic initiation in double mutants that lack Set2 and another factor, indicating that mechanisms other than histone H3 K36 methylation play an important role in this regulation [34]. Fourth, previous analysis has identified cases in which mutations that impair distinct aspects of transcription can combine to cause strong effects on cryptic initiation [34–36]. Finally, assay of a small set of cryptic promoters showed that they were activated in distinct patterns among different cryptic initiation mutants. For example, the pattern of cryptic initiation in mutants that impair Rpd3-mediated histone deacetylation was different from cryptic initiation in mutants affecting histone assembly (Figure 2B). Thus, cryptic promoters may be similar to normal promoters in terms of the complexity of regulation by distinct sets of factors, raising the possibility that additional transcription factors may regulate specific subsets of cryptic promoters. Consistent with this idea, our analysis of the *FLO8* cryptic promoter has shown that it requires a UAS-like element as well as a TATA element (V. Cheung and F. Winston, unpublished data).

The microarray experiments that we have performed suggest that there are at least 1,000 cryptic promoters in the *S. cerevisiae* genome that are activated in *spt6* or *spt16* mutants. The similarity between these two mutants suggests
that they serve similar roles in normally repressing cryptic initiation, likely by helping to establish or maintain a repressive chromatin structure across coding regions [23,25]. Another recent set of microarray studies examined cryptic initiation in set2Δ mutants [31] and identified 621 genes with cryptic transcription on the sense strand. That study also identified 494 antisense transcripts, something not measured in our analysis. Similar to our results, the genes identified by Li et al. [31] were enriched for long genes transcribed at low level. Although we would expect that the cryptic promoters activated in set2Δ mutants would be a subset of those found in spt6 and spt16, only 45% of those found in set2Δ were found in spt6. This degree of overlap, while still quite significant, was likely affected, at least in part, by differences in the microarrays and analysis of the datasets. The smaller number of cryptic promoters in set2Δ mutants compared to spt6 and spt16 fits with our results that mechanisms beyond histone H3 K36 methylation control cryptic initiation. The possible role of antisense transcripts is unknown, although recent studies have demonstrated roles in transcriptional regulation [19,20].

Another evidence suggests that promoters within coding regions occur on a wider scale than indicated by our microarrays of spt6 and spt16 mutants. One study, that examined the S. cerevisiae transcriptome in a wild-type strain by serial analysis of gene expression (SAGE), identified 384 genes with transcription start sites located within the 3' half of the coding region [12]. Only 55 of these 384 genes (14.3%) were identified in our spt6-1004 microarrays to express short transcripts. This small overlap is expected, as our experiments were designed to identify cryptic promoters activated specifically in spt6 mutants. In addition, our spt6-1004 microarrays were designed to detect short transcripts only from the sense strand, while the SAGE analysis was able to detect both sense and antisense short transcripts. More thorough microarray and transcriptome analysis of additional cryptic initiation mutants and other growth conditions will provide a more comprehensive map of cryptic promoters in the yeast genome.

The question still remains as to why so many cryptic promoters are found in the S. cerevisiae genome and what role they serve, if any. We can envision at least four possible roles for cryptic promoters, none of which are mutually exclusive, as all are possible for different subsets. First, some cryptic promoters may direct the expression of gene products that carry out specific functions, being expressed in response to particular environmental changes. In this way, use of cryptic promoters would be analogous to other mechanisms of expressing different genetic information, such as alternative splicing or use of internal ribosome entry sites. Although our results have not demonstrated a function for a product of cryptic initiation, precedent exists for using an internal promoter to express an alternative protein, sometimes under particular growth conditions [22,63–66]. In mammalian cells, the use of alternative promoters has been shown to have numerous roles in normal gene expression and in disease-associated genes [67]. Other results have also shown the potential to express shorter gene products in response to an environmental change [68]. Our results, showing that many cryptic transcripts are translated and that some cryptic promoters are activated by a nutritional shift, also fit with this possibility. We note that we did test for evidence of conservation between S. cerevisiae genes with cryptic transcription and S. bayanus orthologs, but did not detect any significant reduction in either synonymous or nonsynonymous changes in genes with cryptic transcription when compared to genes without cryptic transcription, but of similar length (unpublished data). Second, the information expressed from cryptic promoters may provide the potential for an adaptive mechanism in which, under appropriate selective conditions, expression of such products would enable improved growth or survival, thereby facilitating evolutionary genetic changes. Such an idea was previously suggested for the yeast prion [PSI+], which affects the fidelity of translational termination and thus allows for the possible production of novel protein products [69–71]. Strains containing [PSI+] can acquire complex phenotypic traits distinct from [psi–] strains, and when outcrossed to wild-type strains, these phenotypic traits can sometimes be maintained even after treatment to remove [PSI+] [70,71]. A possible role for intergenic RNAs has also been previously suggested [72]. Third, some cryptic promoters may serve to regulate transcription or control chromatin structure without producing a functional gene product. A previous study demonstrated that a promoter within PRY3 of S. cerevisiae serves to repress PRY3 expression during mating [21]. In this case, transcription
from the internal promoter does not appear to play any functional role. In other cases, the act of transcription may alter chromatin structure in some beneficial way, as previously suggested [68]. Finally, some cryptic promoters may be “noisy,” existing as one of many transcriptional events that serve no apparent biological role [73]. In such a scenario, a significant role of the genes we identified in our screen would be to minimize such “noise,” similar to that of Trn4, Air1, Air2, and components of the exosome in the removal of cryptic intergenic transcripts [37]. Given the very large number of cryptic promoters in S. cerevisiae, it seems reasonable to speculate that all of these reasons and others may turn out to be true. The analysis of specific cryptic promoters will likely yield additional insights into their roles and into previously unknown aspects of gene expression.

Materials and Methods

S. cerevisiae strains and media. All S. cerevisiae strains are listed in Table S9. Strains with the prefix “FY” are isogenic with a GAL2 derivative of FY2717 [74]. Strains were transformed by standard methods, either by crosses or by transformations [75]. The spt6-1004 [23], spt16-197 [76], hta1-hib1A::LEU2 [77], spt2a::kanMX [35], spt21::kanMX [78], spt4a::URA3 [79], and RAS2::ura3 [80] alleles have been previously described. The can1A::STE2pr::LEU2 allele was generally provided by the Bloomington Stock Center [81]. The cdc3A::his3 [82], spt16::ura3, set2a::kanMX, esaA::kanMX, vro1A::kanMX, asA1::kanMX, rtt106::kanMX, cah1A::kanMX, ced73A::kanMX, sfr1A::kanMX, ras2A::kanMX, and ygr177A::kanMX deletion mutations are from the S. cerevisiae haploid nonessential genome deletion library [82], and the deletions were verified to be correct by PCR. The cdc3A::ura3, cdc73A::ura3, cah1A::his3, hsa1A::leu2, and hts1A::kanMX deletion mutations were constructed by replacing the ORFs with URA3, HIS3, LEU2, or kanMX by standard methods [83–85]. The spt10-302::URA3 allele consists of a Tn10-LUK transposon [86] inserted at the SPT10 locus. The [SPT10::URA3]::his3 duplication allele was constructed by integrating plasmid pFW216 (a derivative of pRS306 containing SPT10 and URA3) at the ure3-52 genomic locus [87,88]. To construct the hh2-s257p allele, a 1.9-kb Smal-EcoRI fragment containing HHT2 and HHT2f from plasmid pDM18 was ligated into vector pRS306 (URA3) [88,89]. The TCT1 codon (Ser) at base pair position +172 of HHT2 (relative to the +1 ATG start codon) was mutated to a CCT codon (Pro) by site-directed mutagenesis (QuikChange kit, Stratagene). The resulting plasmid was used to replace the wild-type HHT2 allele in strain FY2716 by two-step gene replacement. The presence of the mutation in the hh2-s257p allele in the genomic context was determined by sequencing strain FY2724 containing the mutation in HHT2 changing the GAA codon (Glu) at base pair +316 to an AAA codon (Lys) and was created by UV mutagenesis of strain FY2724 and verified by sequencing.

The kanMX-GAL1pr-flo8-HIS3 reporter, a 2-kb cassette containing the kanMX marker and the GAL1 promoter was amplified by PCR from plasmid pA6a-kanMX6-PGA1 [90]. This cassette was used to transform strain FY2425 by integration at the FLO8 promoter, replacing base pairs −1,147 to −1 (relative to the FLO8 +1 ATG start codon), to create strain FY2174. The HIS3 ORF (663 bp) was amplified by PCR from plasmid pRS313 [84] and transformed into strain FY2174 at the genomic FLO8 locus, replacing the 3’ end of the FLO8 ORF and the first 105 bp of the 3’ UTR (base pairs +1,727 to +2,505 of FLO8 relative to the +1 ATG start codon). Successful transformants were selected on SC-His medium and verified by PCR. The HIS3 ORF was inserted out-of-frame with respect to the FLO8 ORF and is inserted 3’ of both the internal FLO8 TATA element (+1,626 to +1,651) and the cryptic transcription initiation sites of the FLO8 short transcript (+1,679 to +1,685) [25]. To construct the flo8-HIS3 reporter, the HIS3 ORF was transformed into strain FY2425 and inserted at the FLO8 locus as described above. The AP2M2-TAP-His3MX, DDC1-TAP-His3MX, OMS1-TAP-His3MX, PUS4-TAP-His3MX, SYF1-TAP-His3MX alleles are from the S. cerevisiae TAP-tagged library [58]. The pam2-1-TAP-His3MX allele contains a point mutation in the in-frame ATG codon at base pair position +1 (relative to FLO8 +1, where +1 is the first ATG start codon and changing it to a TTG codon (Leu). The pam2-2-TAP-His3MX allele contains three point mutations at the +1 ATG start codon of AP2M2, changing it to a GTG codon. The pam2-3-TAP-His3MX allele contains both the +1 ATG and the +147 ATG mutations in PUS4. The pam2-2-TAP-His3MX allele contains a point mutation in the in-frame ATG codon at base pair position +1 of the FLO8 ORF (relative to FLO8 +1, where +1 is the first ATG start codon) changing it to a TTG codon (Leu). The pam2-2-TAP-His3MX allele contains three point mutations at the +1 ATG start codon of PUS4, changing it to a 4-GT1 codon. The pam2-3-TAP-His3MX allele contains both the +1 ATG and the +147 ATG mutations in PUS4. All ATG mutations were made using a two-step gene replacement using a previously described method [91] and verified by sequencing.

For liquid cultures, strains were grown in either YPD rich medium (0.67% yeast extract; 2% bactopeptone; 2% glucose) or 2% galactose-based medium (0.67% yeast extract; 2% galactose; 2% ammonium sulfate; 2% glucose) as indicated. Synthetic complete media plates (SC and SC-His) were prepared as previously described [75]. SC + Gal plates and SC-His + Gal plates were made using 2% galactose instead of glucose as the carbon source. For diploid selection (S3AT) a 100 μM 3AT was added to SC-His plates at the concentrations described below.

Isolation of cryptic initiation mutants. Cryptic initiation mutants were isolated using the following three methods: spontaneous mutant selection, synthetic genetic array (SGA) selection with the S. cerevisiae genome nonessential deletion set [82,92], and direct testing of candidate genes. Spontaneous mutant selection was performed using the parental wild-type strains FY2393, FY2713, FY2717, and FY2718, each containing the kanMX-GAL1pr-flo8-HIS3 reporter. Parental strain FY2393 was selected by growth on SC-His plates washed twice in water, and then either 1 × 10⁵ cells or 1 × 10⁶ cells from each culture were plated on SC-His media plates containing 0, 1, 2, 3, 4, 5, or 10 mM 3AT. Plates were either UV-irradiated (5,000 μJ/cm²) or left untreated, and then grown at 30 °C to select for His+ mutants. Potential His+ cryptic initiation mutants were single-colony purified and verified to be His+ by growth on SC-His media plates. The resulting clones were identified by diploid complementation, plasmid complementation, linkage analysis, and cloning by plasmid complementation with an S. cerevisiae genomic library [93]. A total of 254 different mutants were isolated, and 22% of them were identified as belonging to three previously described mutant groups: SPT21, SPT10, HTA1-HTB1, HHT1, HHT2, HIR1, HIR2, HIR3, HPC2, and mutations linked to the kanMX-GAL1pr-flo8-HIS3 reporter. SGA analysis was performed as previously described [92], using the query strain L1102 and screening for deletion mutants that allowed growth on SC-His media plates. Potential positive candidates from SGA analysis were individually crossed with strain FY2506, and their His phenotype was verified by tetrad analysis.

Microarray design, hybridization, and analysis. Probe sequences corresponding to 5,869 ORFs of the S. cerevisiae genome were submitted to Agilent Technologies for microarray production. Each ORF was represented by six 60-mer probes spaced evenly along its coding sequence, with the most-5’ probe beginning at base pair position +1 (relative to the +1 ATG start codon) and the most-3’ probe ending at the final stop codon. Strains FY80 and FY2425 were used for four independent sp66 microarray experiments, and strains FY2713 and FY2717 were used for four independent microarray experiments. Experimental pairs were performed in dye reversal. Wild-type and mutant cells were grown in YPD medium at 30 °C to mid-log phase (1–5 x 10⁶ cells/ml), shifted to 37 °C for 80 min, and then harvested as previously described [94] Sample preparation, labeling, and hybridization of microarrays were performed as previously described [94,95]. Microarray images were acquired and spots quantified with a GenePix 400B microarray scanner and 3.0 software, respectively (MDS Scienx). Spatial detrending and variance stabilization normalization of raw microarray data were performed as previously described [95]. Genes were detected as expressing short transcripts in either the sp66 or the sp16 mutant using the following criteria. The mutant/wild-type ratio was calculated for each probe on the microarray using the normalized spot fluorescent intensity values. For each ORF, the 3’/5’ ratio was calculated by dividing the mutant/wild-type ratio of the most-3’ probe by the mutant/wild-type ratio of the most-5’ probe. Genes with high 3’/5’ ratios were predicted to express short transcripts, whereas genes with low 3’/5’ ratios (close to 1.0) were predicted to not express short transcripts. The location of the internal transcription start site for genes generating short transcript was estimated to be the 3’ end of the probe. The ratio of each probe in the gene relative to the corresponding ratio of the most-5’ probe. Based on the microarray results for five genes previously known to express short transcripts in an sp66 mutant (FLO8, SPB4, STE11, RAD18, and VPS72) [23], a 3’/5’ ratio threshold was set at 0.05 with a 3’/5’ ratio of greater than 2.5 in either all four sp66 microarray experiments or both sp16 microarray experiments were predicted to express a short transcript. Using this criterion, 960 genes were predicted to express short transcripts in an
sp6 mutant, and 1,130 genes were predicted to express short transcripts in a sp6 mutant. It is likely, though, that even more cryptic promoters exist, as the method of calculation likely and unavoidably discriminates against the identification of cryptic transcripts from highly transcribed genes. This discrimination arises from the fact that the hybridization signal from the 3' probe is the sum of the signals of both full-length and cryptic transcripts. The genes with a high level of the full-length transcript, the level of a cryptic transcript would need to also be high to be detectable. The 3'/5' ratio from the microarray results are shown plotted according to expression levels in Figure S2. In support of a greater number of cryptic transcripts, when a more relaxed threshold was used (3/5' ratio of 2.0 rather than 2.5), 620 additional genes were predicted to express cryptic transcripts. When five genes were randomly selected from these 620 genes, four of them expressed short transcripts as detected by northern analysis (unpublished data). However, it is clear that not all genes that produce cryptic transcripts, as northern analysis of ten other genes at random showed that only one produced a detectable cryptic transcript (I. Ivanovska, J. Pamment, and F. Winston, unpublished data).

Northern analysis. mRNA preparation and northern hybridization analysis were performed as previously described [96]. Unless otherwise indicated, RNA was prepared from cells grown in YPD at 30°C to mid-log phase (1–3 × 10^7 cells/ml) for 80 min. For northern blot experiments, cells were grown in YPD to mid-log phase at 30°C and then shifted to 37°C for 80 min. For RNA in situ shift experiments, cells were grown in YPD to mid-log phase at 30°C, washed twice with SD, and then grown in an equivalent volume of SD at 30°C for the indicated times. Double-stranded northern probes were amplified by PCR from genomic DNA and were designed to hybridize to the 3' ends of FLO8 (+1,502 to +2,320), SNR190 (+1,605 to +1,812), STE11 (+1,861 to +2,137), FMP2 (+1,445 to +1,780), DDC1 (+1,768 to +1,879), OMS1 (+1,084 to +1,351), PUS4 (+861 to +1,134), SYF1 (+2,032 to +2,525), and CHS6 (+1,917 to +2,295). A probe for SNR190 (+1 to +190) was used as a loading control for all northern analyses. Because the probes are double stranded, they could anneal to either sense or antisense transcripts, and thus, the base pair position given for each probe is relative to the +1 ATG start codon of the respective gene.

Western analysis. For Western analysis of histone H3 and H3 K36 methylation, whole-cell protein extracts were prepared as previously described [79]. The protein concentration of extracts was determined by Bradford assay (Bio-Rad). Fifty micrograms of whole-cell extracts were separated on 15% acrylamide SDS-PAGE gels, transferred to immobilon-P membrane (Millipore), and analyzed by immunoblotting as previously described [48]. Antibodies were used that recognized total histone H3 (1:5,000 dilution, Abcam), dimethylated H3 K36 (1:10,000 dilution, Upstate), and trimethylated H3 K36 (1:10,000 dilution, Abcam). Antibodies were detected by chemiluminescence (PerkinElmer).

For western analysis of TAP-tagged proteins, whole-cell protein extracts were prepared as follows: 30 ml of cells were grown in YPD at 30°C to mid-log phase (1–3 × 10^7 cells/ml) and then shifted to 37°C for 80 min. Cells were washed twice with wash buffer (20 mM Tris-Cl, 150 mM NaCl, [pH 7.5]) and resuspended in 400 μl of lysis buffer (50 mM Hepes-KOH [pH 7.5], 150 mM NaCl, 10% glycerol, 0.5% NP-40, 1 mM EDTA, 1 mM PMSF, 2 μg/ml Leupeptin, 2 μg/ml Pepstatin A). One milliliter of glass beads was added, and cells were lysed by vortexing in an Eppendorf multthead shaker 5492 for 40 min at 4°C. The cell lysate was spun out through a hole punched in the bottom of the tube, by spinning for 2 min at 1,000 rpm. The lysate was spun for 5 min at 14,000 rpm, and the supernatant was saved and spun again for 15 min at 37°C. The supernatant was saved from this final spin and used for western analysis. Total protein concentration of extracts was determined by Bradford assay (Bio-Rad). Equal amounts of whole-cell extracts were separated on 8% acrylamide SDS-PAGE gels, transferred to immobilon-P membrane (Millipore), and analyzed by immunoblotting as previously described [79]. The TAP tag was detected by chemiluminescence (PerkinElmer Life Sciences) using the peroxidase anti-peroxidase antibody (1:5,000 dilution, Sigma). Pgk1 was used as a loading control and visualized with anti-Pgk1 antisera (Molecular Probes) that was generously provided by Angela Nison’s laboratory.

Analysis of open reading frames. To examine which protein domains are present and lost, we obtained data on proteins from SGD (ftp://genome-ftp.stanford.edu/pub/yeast/sequence similarity/ domains, last updated on September 25, 2007) and mapped them onto the proteins encoded by genes with cryptic transcription initiation. We considered the first ATG after the most-5’ limit of the cryptic transcript as a conservative limit for the length of the short protein being produced; i.e., cases in which a minimal number of residues would be lost. A domain was called to be absent if the position of the ATG was downstream of the domain start site. Using published data on protein domains that are at the physical interface of the interacting partners [97], we also examined whether these lost domains are known to mediate physical interaction among proteins. Finally, in order to estimate how common these domains are among yeast proteins, we tabulated how many proteins in the genome have the domains.

Accession number. The microarray data, accession number GSE12272, can be found at GEO (http://www.ncbi.nlm.nih.gov/geo).

Supporting Information

Figure S1. Clustergram Analysis of sp6-1004 Transcription

Clustergram showing normalized sp6-1004 wild-type ratios (as median-subtracted asinh values; [98]) for individual probes from the 906 genes we classified as having short transcripts in an sp6-1004 mutant. The color scale shown spans from asinh-2 to +2. Numbers shown next to the color bar indicate corresponding ratios in the linear domain and are rounded to the nearest integer.

Found at doi:10.1371/journal.pbio.0060277.sg001 (49 KB PDF).

Figure S2. Possible Underestimation of Cryptic Transcripts

Cryptic transcripts may be less likely to be detected for genes with abundant transcriptional abundance. The ratio does not scale with that of the full transcript. The ratio of 3' to 5' of hybridization intensity is shown as a function of transcript abundance. The ratio was calculated as defined in the text. In grey are ratios of intensities for probe 6 and 5 and in black for probes 6 and 1. The red line indicates the cutoff for calling the presence of a clear cryptic transcript (2.5). (A) sp6-1004 mutant; (B) sp616-197 mutant.

Found at doi:10.1371/journal.pbio.0060277.sg002 (7.41 MB PDF).

Figure S3. TATA Motifs in ORFs and Cryptic Transcription

The presence of a TATA motif in the coding sequence of a gene increases the probability of cryptic initiation of transcription in that gene, independent of the size of the gene. Genes were separated by size classes corresponding to each of the ten intervals corresponding to the ten quantiles of the size distribution. The expected number of genes was calculated as the product of the fraction of genes with a TATA motif in this size interval times the number of genes with cryptic transcripts in that interval. The observed number represents the fraction of genes with a TATA motif in the coding sequence that produce cryptic transcript. (A) sp6-1004 mutant; (B) sp616-197 mutant.

Found at doi:10.1371/journal.pbio.0060277.sg003 (466 KB PDF).

Figure S4. TATA Motifs in Promoters and Cryptic Transcription

Genes with a TATA box in their promoter are less likely to produce cryptic transcripts, and the occurrence of a TATA box occurs independently from transcript abundance. Genes were separated by expression classes corresponding to each of the ten intervals corresponding to the ten quantiles of the distribution. The expected number of genes with a TATA promoter was calculated as the product of the fraction of genes with a TATA promoter for this expression level interval times the number genes with cryptic transcription in that interval. If the presence of a TATA promoter was independent from the production of a cryptic transcript, the fraction of genes that produce a cryptic transcript and that have a TATA promoter should be proportional to the fraction of genes with a TATA promoter in this size interval. (A) sp6-1004 mutant; (B) sp616-197 mutant.

Found at doi:10.1371/journal.pbio.0060277.sg004 (517 KB PDF).

Figure S5. Examples of Proteins Made in sp6-1004 Mutants from Cryptic Transcripts

Shown are proteins found to be expressed from cryptic promoters (Table S8). The gray boxes designate the portions in the shorter proteins. The orange line represents the full-length protein that is made from the wild-type transcript.

Found at doi:10.1371/journal.pbio.0060277.sg005 (172 KB PDF).

Table S1. Histone H3 Mutants That Allow Cryptic Initiation

Found at doi:10.1371/journal.pbio.0060277.s001 (22 KB DOC).
Table S2. Histone H3 Di- and Trimethylation Levels in Mutants

Table S3. sp6-1004 Microarray Results

Table S4. spel6-197 Microarray Results

Table S5. TATA Consensus Sequences in Coding Regions

Table S6. Translation of Coding Regions in Cryptic Transcripts

Table S7. Genes Identified That Express Short Proteins in an sp6-1004 Mutant

Table S8. Domains Lost in Potential Translation Products in sp6-1004 Mutants

Table S9. S. cerevisiae Strains

References

1. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. (2004) Identification and analysis of functional elements in the human genome by the ENCODE pilot project. Nature 447: 799–816.

2. Gingeras TR (2007) Origin of phenotypes: genes and transcripts. Genome Res 17: 682–690.

3. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, et al. (2005) The transcriptional landscape of the mammalian genome. Science 309: 1559–1563.

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

5. Prasanth KV, Spector DL (2007) Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. Genes Dev 21: 11–42.

6. Manak JR, Dike S, Semenchenko V, Kapranov P, Biemar F, et al. (2005) Biological function of unannotated transcription during the early development of Drosophila melanogaster. Nat Genet 38: 1151–1158.

7. Ashe HL, Monks J, Wijgerde M, Fraser P, Proudfoot NJ (1997) Intergenic transcription and transduction of the human beta-globin locus. Genes Dev 11: 2549–2569.

8. Schmitt S, Prestel M, Paro R (2005) Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev 19: 697–708.

9. Yagaz O, Krebs JE (2007) Noncoding but non expendable: transcriptional regulation by large noncoding RNA in eukaryotes. Biochem Cell Biol 85: 484–496.

10. David L, Huber W, Granzovskia M, Toedling J, Palm GJ, et al. (2006) A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci U S A 103: 5320–5325.

11. Davis CA, Ares M Jr (2006) Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 103: 3262–3267.

12. Miura F, Kagawachi N, Sese J, Toyota A, Hattori M, et al. (2006) A largescale full-length cDNA analysis to explore the budding yeast transcriptome. Proc Natl Acad Sci U S A 103: 17846–17851.

13. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, et al. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320: 1344–1349.

14. Samanta MP, Tongprasit W, Sethi H, Chin CS, Stolc V (2006) Genome-wide distribution of yeast RNA polymerase II and its control by the Bur1 and Paf1 complexes. Mol Cell 24: 735–746.

15. Martens JA, Laprade L, Winston F (2004) Intergenic transcription is required to repress the Saccharomyces cerevisiae SERS5 gene. Nature 429: 571–574.

16. Uhler JP, Hertel C, Sweijstrou QJ (2007) A role for noncoding transcription in activation of the yeast PHO5 gene. Proc Natl Acad Sci U S A 104: 8011–8016.

17. Camblong J, Iglésias N, Fickentcher C, Diepgois G, Stutz F (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. Cell 131: 796–797.

18. Ono B, Futase T, Honda W, Yoshida R, Nakano K, et al. (2005) The TATA consensus sequences in coding regions. Cell 127: 735–745.

19. Pickell KS, Morris DR (2006) Role of the transcription activator Ste12p as a repressor of PRY3 expression. Mol Cell Biol 26: 7901–7912.

20. Chu Y, Simic R, Warner MH, Arndt KM, Prelich G (2007) Regulation of transcriptional landscape of the mammalian genome. Science 309: 1559–1563.

21. Martinez JA, Laprade L, Winston F (2005) Sp6 expression elongation factors repress transcription initiation from cryptic sites. Science 301: 1096–1099.

22. Mason PB, Struhl K (2005) The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. Mol Cell Biol 25: 8325–8333.

23. Belotserkovskaya R, Oh S, Bondarenko VA, Orphanides G, Stoddsky VM, et al. (2005) FACT facilitates transcription-dependent nucleosome alteration. Science 310: 1090–1093.

24. Bortvin A, Winston F (1996) Evidence that Sp4p controls chromatin structure by a direct interaction with histones. Science 272: 1473–1476.

25. Hartogza G, Wada T, Handa H, Winston F (1998) Evidence that Sp4p, Sp5p, and Sp6p control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes Dev 12: 357–369.

26. Carrozza MJ, Li B, Floros L, Suganuma T, Swanson SK, et al. (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123: 591–601.

27. Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, et al. (2005) Cotranscriptional set2 methylation of histone H3 lysine 46 recruits a repressive Rpd3 complex. Cell 125: 593–606.

28. Li B, Gogol M, Carey M, Pattenden SG, Seidel C, et al. (2007) Infrequently transcribed long genes depend on the Set2/Rpd3 pathway for accurate transcription. Genes Dev 21: 1422–1430.

29. Schwalbe MA, Struhl K (2006) Afl1 mediates histone eviction and deposition during elongation by RNA polymerase II. Mol Cell 22: 415–422.

30. Xiao T, Shibata Y, Rao B, Laribee RN, O'Rourke R, et al. (2007) The RNA polymerase II kinase Cdk1 regulates positioning of a 5'-histone methylation boundary along genes. Mol Cell 27: 721–731.

31. Chou Y, Simic R, Warner MH, Arndt KM, Prelich G (2007) Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. EMBO J 26: 4646–4656.

32. Shimizu D, Nishimura H, Yasuda T, Nishiyama N, Nishimura H, et al. (2007) Domain II and III of TAF11 contribute to transcription elongation factor in Saccharomyces cerevisiae. Mol Cell 25: 10122–10135.

33. Prather D, Krogan NJ, Emili A, Greenblatt JF, Winston F (2005) Identification and characterization of the TAF11 subunit of the FLAIR complex. Mol Cell 21: 712–727.

34. Frater D, Krogan NJ, Emili A, Greenblatt JF, Winston F (2005) Identification and characterization of the TAF11 subunit of the FLAIR complex. Mol Cell 21: 712–727.

35. Hyland EM, Congrove MS, Molina H, Wang D, Pandey A, et al. (2005) Insights into the role of histone H3 and histone H4 core modifiable residues in Saccharomyces cerevisiae. Mol Cell Biol 25: 10960–10970.

Acknowledgments

We thank Miles Trohse for assistance with microarray design, Charlie Boone for a yeast strain, Angelika Amon for antibodies, and Takashi Ito for sharing data. We are grateful to Mark Hickman for help with analysis of the microarray data. We also thank Lisa Laprade and the Schwartzfahr for help in construction of the FLOWHS3 reporter. We also thank Alan Hinnebusch and Kevin Struhl for helpful discussions, and Karen Arndt for helpful comments on the manuscript.

Author contributions. VC, GC, NNB, CRL, SWM, TRH, and FW conceived and designed the experiments. VC, GC, NNB, and CRL performed the experiments. VC, FW, and CRL wrote the paper with contributions from GC, NNB, CRL, SWM, and TRH.

Funding. GC was supported by a Charles H. Best Postdoctoral Fellowship, NNB was funded by a postdoctoral fellowship from Canadian Institute of Health Research (CIHR), and CRL is a National Sciences and Engineering Research Council postdoctoral fellow in evolutionary biology and genome sciences. This work was supported by Genome Canada through the Ontario Genomics Institute to TRH and by National Institutes of Health grant GM32967 to FW.

Competing interests. The authors have declared that no competing interests exist.
