### Citation

### Published Version
[doi:10.1371/journal.pbio.0060277](http://dx.doi.org/10.1371/journal.pbio.0060277)

### Accessed
March 31, 2017 5:50:21 AM EDT

### Citable Link
[http://nrs.harvard.edu/urn-3:HUL.InstRepos:11181061](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](http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA)

(Article begins on next page)
Chromatin- and Transcription-Related Factors Repress Transcription from within Coding Regions throughout the *Saccharomyces cerevisiae* Genome

Vanessa Cheung¹, Gordon Chua²a, Nizar N. Batada¹nb, 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.

**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-
type strain but becomes activated in an spt6 mutant [23]. Evidence suggested that in spt6 mutants, the failure to reassemble nucleosomes in the wake of elongating RNA polymerase II (RNAPII) allowed transcription initiation factors to bind to and activate cryptic promoters [23].

Several transcription factors are required to repress cryptic promoters in *S. cerevisiae*. An early study revealed that several different mutants allow cryptic initiation [23]. Subsequent analysis has suggested that the level of histone modifications in coding regions, as regulated by the Set2 histone methyltransferase and the Rpd3S histone deacetylase complex, also controls cryptic initiation [28–30] and that set2Δ mutations allow cryptic initiation in a large set of genes [31]. Additional work has identified other mutants that allow cryptic initiation, including asf1 and cik1, [32,33], as well as particular combinations of double mutants, revealing roles for other elongation factors, including the Paf1 complex, Bur1-Bur2, the HIR complex, Spr2, and Elf1 [34–36]. These studies suggest that the repression of cryptic promoters requires a variety of factors that play roles in transcription elongation and chromatin structure. These factors appear to be entirely distinct from those that suppress cryptic intergenic transcripts [37].

In this paper, we present the results of genome-wide approaches to comprehensively study cryptic transcription from within open reading frames (ORFs) in *S. cerevisiae*. First, we used both spontaneous mutant selection and a synthetic genetic array (SGA) screen to identify new mutations that allow cryptic transcription. These mutations have varying effects on the expression of a set of cryptic transcripts, suggesting the existence of different classes of cryptic promoters and mechanisms for their activation. Second, we used microarray analysis to identify cryptic transcripts throughout the *S. cerevisiae* genome that are activated in spt6 and spt16 mutants. These experiments showed that cryptic transcription is widespread, occurring in at least 1,000 genes (17% of all genes). We have also investigated the possibility of a physiological role for cryptic transcription, as it is not understood whether it represents unwanted transcription from fortuitous promoters that are activated only in mutants in which chromatin structure has been altered, or whether it serves a biological role in some cases, possibly to express different gene products. Here, we demonstrate that a number of cryptic transcripts expressed in an spt6 mutant are translated into corresponding short proteins. In addition, we show that some cryptic transcripts are modestly activated in wild-type (*SPT6*) strains upon a nutritional shift and that this activation is dependent upon Ras2. Taken together, our results show that cryptic transcription from ORFs can occur in a widespread fashion throughout the *S. cerevisiae* genome and suggest that some cryptic promoters may normally serve to express alternative genetic information during environmental changes.

**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
Most Cryptic Transcription Mutants Have Normal Levels of Histone H3 K36 Methylation

Recent results have shown that Set2-dependent methylation of histone H3 at K36 plays a role in the repression of cryptic transcription [28,47,29]. Furthermore, both H3 K36 dimethylation and trimethylation have recently been shown to be defective in \textit{spt6} and \textit{set1} mutants, as well as in \textit{set2} mutants [28,48]. Therefore, we tested whether this histone H3 K36 methylation defect might be a common phenotype among cryptic transcription mutants. Our results show that, of 50 mutants tested, only five showed a significant decrease in total H3 P38 di- and trimethylation (\textit{spt6-1004}, \textit{set2-1}, \textit{cik1A}, \textit{cik2A}, and \textit{ctk3A}) (Figure 3, Table S2). The histone H3 K36 methylation defects in these five mutants have been previously reported [28,33,47–49]. We note that under our growth conditions, the \textit{spt16-197} mutant had wild-type levels of H3 K36 di- and trimethylation, in contrast to a previous report [48], yet still showed a high level of cryptic transcription. These results show that the majority of the cryptic transcription mutants regulate at a step other than H3 K36 methylation. At least 1,000 Cryptic Transcripts are Produced in \textit{spt6} and \textit{spt16} Mutants

Previous studies of cryptic initiation in an \textit{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 \textit{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 \textit{spt6-1004} mutant (Figure S1; Table S3). As detailed in Materials and Methods, this method may unavoidably be biased towards

Transcriptional Analysis Suggests Different Classes of Cryptic Promoters

To test whether the mutants we identified activate cryptic transcription from multiple genes, we performed northern analysis on 14 cryptic initiation mutants, examining transcription of \textit{FLO8}, \textit{SPB4}, and \textit{STE11}, three genes previously shown to have cryptic promoters [23]. Our results show that there are different patterns of cryptic promoter activation among the mutants (Figure 2B). Most of the mutants express the \textit{FLO8} short transcript, with the exceptions of \textit{hir1A} and \textit{ehl1A} (Figure 2B, lanes 10 and 14; also see [35,36]), suggesting that in some cases, the \textit{FLO8-HIS3} reporter is more sensitive in detecting cryptic initiation than northern analysis. Conversely, an \textit{spt16-197} mutant appeared weakly His+ with the \textit{FLO8-HIS3} reporter, whereas northern analysis indicated high levels of expression of the \textit{FLO8} short transcript (Figure 2A and 2B, lane 3). This effect with \textit{spt16-197} may be due to the slow growth of the \textit{spt16-197} mutant. In addition, short transcripts could be detected for \textit{SPB4} and \textit{STE11} for most of the mutants, indicating that cryptic initiation was not specific to \textit{FLO8}. However, there were differences in the pattern of cryptic transcription among the mutants tested. For example, \textit{spt6-1004}, \textit{eaf3A}, and \textit{rtt106A} confer distinct patterns of activation of \textit{FLO8}, \textit{SPB4}, and \textit{STE11} cryptic transcripts (Figure 2B, compare lanes 2, 8, and 13). These distinct patterns suggest that there are distinct classes of cryptic promoters and different mechanisms for their repression. Other evidence suggesting differential expression of cryptic transcripts has recently been described [47].
transcripts and found that all five indeed produce short cryptic transcripts, we used northern analysis to 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 < 2.2 × 10^{-16}). The majority of genes with cryptic transcription also have lower transcriptional frequencies (for *spt6-1004*, average = 2.46 mRNA/hour [p-value < 2.2 × 10^{-16}] and for *spt16-197*, 1.93 mRNA/hour [p-value < 2.2 × 10^{-16}]) 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*, TATAA(T/A)(T/A)(T/A) [51]. We found that genes with at least one TATA element in their coding region are three times more likely to have a cryptic transcript in the *spt6-1004* mutant than genes without a TATA box (p-value < 2.2 × 10^{-16}). We see an even stronger enrichment for the *spt16-197* mutant (p-value < 2.2 × 10^{-16}) (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]. We found that cryptic transcripts are two times (Fisher exact test, p-value = 2 × 10^{-12}) and 2.4 times (Fisher exact test, p-value = 2 × 10^{-16}) more likely to be from genes with natural TATA-less promoters than from genes with TATA-containing promoters for the *spt6-1004* and *spt16-197* mutants, respectively, after correction for gene expression levels (Figure S4).

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

<table>
<thead>
<tr>
<th>Category</th>
<th>Gene</th>
<th>Mutant Phenytope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histones</td>
<td>HHT1</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>HHT2</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>HTA1-HTB1</td>
<td>++</td>
</tr>
<tr>
<td>Histone gene regulation</td>
<td>SPT10</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>SPT21</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>HIR1</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>HIR2</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>HIR3</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>HPC2</td>
<td>++</td>
</tr>
<tr>
<td>Chromatin assembly and remodeling</td>
<td>SPT6</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>SPT16</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>RTT106</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>ASF1</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>CHD1</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>ISWI</td>
<td>++</td>
</tr>
<tr>
<td>Rpd3S-mediated histone deacetylation</td>
<td>CTK1</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>CTK2</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>CTK3</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>SET2</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>RPD3*</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>SIN3</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>RCO1</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>EAF3</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>UME1</td>
<td>+</td>
</tr>
<tr>
<td>Histone deacetylation</td>
<td>HST4</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>HDAC2</td>
<td>+</td>
</tr>
<tr>
<td>Transcription elongation factors</td>
<td>BUR1</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>BUR2</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>SPT4</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>SPT5</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>CDC73</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>RTF1</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>SPT2</td>
<td>+</td>
</tr>
<tr>
<td>Mediator components</td>
<td>SIN4</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>SBR8</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>SBR9</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>SBR10</td>
<td>+</td>
</tr>
<tr>
<td>Proteasome function</td>
<td>PRE9</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>UMP1</td>
<td>+</td>
</tr>
<tr>
<td>DNA replication and repair factors</td>
<td>HUR1</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>PMS1</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>RNR4</td>
<td>+</td>
</tr>
<tr>
<td>Other categories</td>
<td>ATG11</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>DIG1</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>KEQ2</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>MAP1</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>PIR1</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>PMR1</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>SSD1</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>YBR246W</td>
<td>+</td>
</tr>
</tbody>
</table>

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 rd3 mutant was not tested with FLO8-HIS3; its score is based on northern analysis. doi:10.1371/journal.pbio.0060277.S001

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).
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.
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 produce a protein from the cryptic transcript. The conditions tested included starvation for carbon, nitrogen, phosphate, or sulfate, as well as heat shock, high salt concentration, or exposure to different drugs such as 3AT or menadione. Of the 20 different growth conditions tested, one of them, a shift from rich medium (YPD) to minimal medium (SD), caused modest expression of cryptic transcripts in three of the 16 genes tested, CHS6, FLO8, and SPB4 (Figure 6, lanes 3–6). For these genes, cryptic transcripts were detectable by 30 min after the shift, and for two of the genes, CHS6 and FLO8, it was transient, no longer detectable by 2 h after the shift. In all cases, the level of the short transcript was clearly less than observed in the spt6-1004 mutant, indicating that an spt6 mutant represents an extreme condition for cryptic initiation in the genome, relative to what may be seen in a wild-type strain under different growth conditions.

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

Whole-cell extracts were prepared from cells grown at 30 °C, except for strains FY2425 and FY347, which were shifted to 37 °C 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. WT, wild type.
doi:10.1371/journal.pbio.0060277.g003
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}^{D} and \textit{ras2}^{D} mutants during a nutritional shift. Although \textit{gcn4}^{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{ras2A} 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}-null mutant, the expression of the \textit{CHS6} and \textit{FLO8} cryptic transcripts is strongly Ras2-dependent. However, the expression of the \textit{SPB4} cryptic transcript is largely Ras2 independent, 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.
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

---

**Figure 5. Analysis of Short Protein Expression in an spt6 Mutant**

All Western analyses used whole-cell extracts prepared from cells after an 80-min temperature shift from 30 °C to 37 °C. Probes for the western analyses were generated with an antibody recognizing the TAP epitope tag. Pgk1 was used as a loading control. Arrows indicate full-length proteins, and asterisks indicate short proteins. WT, wild type.

(A) Western analysis of short protein expression in an *spt6-1004* mutant.

(B) Western analysis of short-protein expression in wild-type and *spt6-1004* strains containing *APM2* and *PUS4* ATG mutations. The lower molecular weight bands in lane 6 are likely to be degradation products.

doi:10.1371/journal.pbio.0060277.g005
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 \textit{set2A} 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 \textit{set2A} mutants would be a subset of those found in \textit{spt6} and \textit{spt16}, only 45% of those found in \textit{set2A} were found in \textit{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 \textit{set2A} mutants compared to \textit{spt6} and \textit{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].

Other evidence suggests that promoters within coding regions occur on a wider scale than indicated by our microarrays of \textit{spt6} and \textit{spt16} mutants. One study, that examined the \textit{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 \textit{spt6-1004} microarrays to express short transcripts. This small overlap is expected, as our experiments were designed to identify cryptic promoters activated specifically in \textit{spt6} mutants. In addition, our \textit{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 \textit{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 [22,65–66]. 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 \textit{S. cerevisiae} genes with cryptic transcription and \textit{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 [\textit{PSI}⁺], which affects the fidelity of translational termination and thus allows for the possible production of novel protein products [69–71]. Strains containing [\textit{PSI}⁺] can acquire complex phenotypic traits distinct from [\textit{psi}–] strains, and when outcrossed to wild-type strains, these phenotypic traits can sometimes be maintained even after treatment to remove [\textit{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 \textit{PRY3} of \textit{S. cerevisiae} serves to repress \textit{PRY3} expression during mating [21]. In this case, transcription

![Figure 6](https://example.com/figure6.png)

**Figure 6.** Northern Analysis of Cryptic Initiation during a Nutritional Shift from YPD Rich Medium to SD Minimal Medium for the Indicated Times. Lanes 1 and 2 contain RNA isolated from cells shifted from 30 °C to 37 °C for 80 min as indicated. All other RNA was isolated from cells grown at 30 °C. SNR190 was used as a loading control. Asterisks indicate short transcripts. SNR190 served as a loading control. doi:10.1371/journal.pbio.0060277.g006
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 “noise,” 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 Trf4, 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 strain FY1274 [91]. Strains carrying deletion mutations, either by crosses or by transformations [75], the sip6-1004 [23], sip16-197 [76], his1-his1::ADE2 [77], sap24::kanMX [35], sip21::kanMX [78], sip4::URA3 [79], and RAS2::ADE2 [80] alleles have been previously described. The can1-101::STE2p::LEU2 allele was generously provided by the Boone lab [81]. The FLO8::URA3, HIS3::STE2p::LEU2, and ura3-52::kanMX alleles were added to SC-His plates at the concentrations described below. Microarray experiments, and strains used to transform the APM2-TAP::His3MX reporter, a 2-kb cassette (ATG start codon), to create strain FY2174. The allele contains a point mutation in the in-frame ATG codon at base pair 1 ATG and the most-3 probe. Based on the microarray results for five genes with a 3 ratios (close to 2.5) were predicted to encode short transcripts in either the sip6 or the sip16 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 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 transcripts was estimated by calculating 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 sip6 mutant (FLO8, SPB4, STE11, RAD18, and VPS72) [23], a 3’5’ ratio threshold was set at 2.5, with a 3’5’ ratio greater than 2.5 in either all four sip6 microarray experiments or both sip16 microarray experiments were predicted to express a short transcript. Using this criterion, 960 genes were predicted to express short transcripts in an
sp66 mutant, and 1,130 genes were predicted to express short transcripts in a sp66 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 signal from the 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 the detection of these transcripts by northern analysis, 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 analyses 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 × 10⁶ cells/ml). For temperature shift experiments, cells were grown in YPD to mid-log phase at 30 °C and then moved to 42 °C for 80 min. For media 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 FLW9 (+1,360 to +2,320), SPT16 (+1,605 to +1,812), STE11 (+1,860 to +2,126), DDQ7 (+1,449 to +1,785), 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 hybridize to either sense or antisense transcripts. The base pair 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). Equal amounts 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 [79]. Antibodies were used that recognize 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 (1–5 × 10⁶ cells/ml) and then shifted to 37 °C for 80 min. Cells were washed twice with wash buffer (20 mM Tris-HCl, 150 mM NaCl [pH 7.5]) and resuspended in 400 μl of lysis buffer (50 mM Heps-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 5432 for 40 min at 4 °C. The cell lysate was spun out through a hole punctured 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 14,000 rpm. 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 antiserum (Molecular Probes) that was generously provided by A. E. Johnson’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′-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 these 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 sp66-1004 Transcription
Clustergram showing normalized sp66-1004/ wild-type ratios (as median-subtracted asinh values; [98]) for individual probes from the 900 genes we classified as having short transcripts in an sp66-1004 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 the abundance of cryptic transcripts does not scale with that of the full transcript. The ratio of 3′/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) sp66-1004 mutant; (B) sp66-197 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 in transcription in that gene is 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) sp66-1004 mutant; (B) sp66-197 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 shorter 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 transcript 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) sp66-1004 mutant; (B) sp66-197 mutant.

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

Figure S5. Examples of Proteins Made in sp66-1004 Mutants from Cryptic Transcripts
Shown are proteins found to be expressed from cryptic promoters (Table S8). The gray boxes designate the portions that are 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 Found at doi:10.1371/journal.pbio.0060277.s002 (20 KB XLS).

Table S3. 5p6-1004 Microarray Results Found at doi:10.1371/journal.pbio.0060277.s003 (5.62 MB XLS).

Table S4. 5p16-197 Microarray Results Found at doi:10.1371/journal.pbio.0060277.s004 (5.62 MB XLS).

Table S5. TATA Consensus Sequences in Coding Regions Found at doi:10.1371/journal.pbio.0060277.s005 (22 KB DOC).

Table S6. Translation of Coding Regions in Cryptic Transcripts Found at doi:10.1371/journal.pbio.0060277.s006 (1.01 MB XLS).

Table S7. Genes Identified That Express Short Proteins in an spt6-1004 Mutant Found at doi:10.1371/journal.pbio.0060277.s007 (22 KB DOC).

Table S8. Domains Lost in Potential Translation Products in spt6-1004 Mutants Found at doi:10.1371/journal.pbio.0060277.s008 (1.34 MB XLS).

Table S9. S. cerevisiae Strains Found at doi:10.1371/journal.pbio.0060277.s009 (33 KB DOC).

References


