
Meiosis-induced alterations in transcript architecture and noncoding RNA expression in *S. cerevisiae*

KAREN S. KIM GUIBERT,¹ YONG ZHANG,^{2,5} JARED FLATOW,^{3,6} SARA HURTADO,¹ JONATHAN P. STALEY,⁴ SIMON LIN,³ and ERIK J. SONTHEIMER^{1,7}

¹Department of Molecular Biosciences, Northwestern University, Evanston, Illinois 60208, USA

²Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA

³Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA

⁴Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, Illinois 60637, USA

ABSTRACT

Changes in transcript architecture can have powerful effects on protein expression. Regulation of the transcriptome is often dramatically revealed during dynamic conditions such as development. To examine changes in transcript architecture we analyzed the expression and transcript boundaries of protein-coding and noncoding RNAs over the developmental process of meiosis in *Saccharomyces cerevisiae*. Custom-designed, high-resolution tiling arrays were used to define the time-resolved transcriptome of cells undergoing meiosis and sporulation. These arrays were specifically designed for the *S. cerevisiae* strain SK1 that sporulates with high efficiency and synchrony. In addition, new methods were created to define transcript boundaries and to identify dynamic changes in transcript expression and architecture over time. Of 8407 total segments, 699 (8.3%) were identified by our algorithm as regions containing potential transcript architecture changes. Our analyses reveal extensive changes to both the coding and noncoding transcriptome, including altered 5' ends, 3' ends, and splice sites. Additionally, 3910 (46.5%) unannotated expressed segments were identified. Interestingly, subsets of unannotated RNAs are located across from introns (anti-introns) or across from the junction between two genes (anti-intergenic junctions). Many of these unannotated RNAs are abundant and exhibit sporulation-specific changes in expression patterns. All work, including heat maps of the tiling array, annotation for the SK1 strain, and phastCONS conservation analysis, is available at <http://groups.molbiosci.northwestern.edu/sontheimer/sk1meiosis.php>. Our high-resolution transcriptome analyses reveal that coding and noncoding transcript architectures are exceptionally dynamic in *S. cerevisiae* and suggest a vast array of novel transcriptional and post-transcriptional control mechanisms that are activated upon meiosis and sporulation.

Keywords: tiling array; noncoding RNA; transcript architecture; meiosis; sporulation

INTRODUCTION

The decision for a single cell to enter meiosis and faithfully duplicate, recombine, and segregate its DNA content requires a complex and elegant coordination of regulated events. An error in meiosis can result in death for the cell or disaster for the next generation. Consequently, the cell has numerous mechanisms in place to ensure precision. Meiosis is constructed as a series of feed-forward loops that act as one-way gates that drive the cell toward successful

creation of daughter cells (Mitchell 1994; Clancy 1998; Vershon and Pierce 2000). These loops are balanced by checkpoints that enforce the correct completion of each step before the drive toward the next phase is initiated.

A crucial nexus in the regulation of cell fate is the control of RNA content. Protein-coding regions of the genome have long held dominance as a research focus, yet noncoding regions also have a profound influence on cell physiology. The noncoding transcriptome encompasses noncoding regions associated with every coding transcript in addition to intergenic and antisense RNAs that lack obvious protein-coding potential. Understanding the complete transcriptome is essential to our understanding of cell-fate decisions.

The prominence of noncoding RNA, termed the “dark matter” of the genome, has exploded in recent years (Jacquier 2009). Even though the purpose of this rampant transcription is still under debate, a growing number of previously unannotated RNAs are being recognized for

⁵Present address: Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China.

⁶Present address: Nokia Research Center, Palo Alto, CA 94304, USA.

⁷Corresponding author.

E-mail erik@northwestern.edu.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.030510.111>.

playing critical roles in cell physiology or development. The functions of virtually all but a handful of these unannotated transcripts in budding yeast remain unknown (Hongay et al. 2006; Camblong et al. 2007; Berretta et al. 2008; Camblong et al. 2009).

Annotated genes also have dynamic potential beyond mere expression level through the control of transcript architecture (i.e., the physical nucleotides that comprise the transcript). Changes in architecture include extension or regression of the untranslated regions (UTRs) at the 5' or 3' ends, or at introns. These changes can exert control over the function and expression of the messenger RNA. Internal changes to transcript architecture manifested by splicing are the best-known alterations in architecture. In fact, alternative splicing is a common mechanism in higher eukaryotes to regulate and expand the transcriptome (Nilsen and Graveley 2010). Likewise, changes in the terminal UTRs can also have profound influences on cell fate. A classic example is evident in immunoglobulin genes during B-cell activation in which a change in the poly(A) cleavage site occurs as a competition between the 3'-end processing apparatus and the splicing machinery. Without the change in the 3' end, the B-cell would be unable to express a secreted form of immunoglobulin and the organism would lack a critical component of the adaptive immune response (Takagaki et al. 1996; Takagaki and Manley 1998). Interestingly, shortening of 3' UTR length has been observed in many proliferating or cancerous cell lines (Sandberg et al. 2008; Mayr and Bartel 2009) emphasizing that control of transcript architecture has profound implications for both normal and disease states. At the 5' end, alteration in the architecture of *CIK1* mRNA, differing only in the 5' end, renders one form of the resulting protein sensitive and the other resistant to ubiquitination by the anaphase-promoting complex (Benanti et al. 2009). Both forms are required for normal cellular function. Functionally distinct changes in architecture frequently occur in response to signaling cascades. Therefore, identification of dynamic changes to transcript architecture often requires the study of a dynamic process.

Saccharomyces cerevisiae is an ideal organism in which to delve into the complexities of the complete transcriptome during meiosis. The expression of annotated coding regions has already been characterized via open reading frame (ORF) array analysis (Chu et al. 1998; Primig et al. 2000). A series of well-ordered and highly regulated transcriptional cascades begins with the sensing of environmental cues that trigger meiosis. It then progresses to the initiation of meiosis I (MI), DNA replication, and recombination in the early phase; meiosis II (MII) and DNA segregation in the middle phase; and finally to encapsulation and spore formation in the late phase (Clancy 1998; Vershon and Pierce 2000).

Although *S. cerevisiae* may be the most highly studied eukaryotic organism, knowledge of its noncoding transcriptome has lagged behind that of ORF expression. The process of meiosis must be exquisitely regulated in order to function

properly. Fundamental to understanding meiosis is to define the complete transcriptome, both coding and noncoding, that governs this process. Toward this end, we have designed ultra-high-density tiling microarrays to define the expression of nearly every nucleotide of the yeast genome throughout the multiple stages of meiosis. Our results reveal the existence of numerous unannotated RNAs as well as a surprisingly widespread dynamism in transcript architecture in known mRNAs.

RESULTS

Tiling arrays designed to detect changes in both expression and architecture

To identify high-resolution, strand-specific variations in expression patterns, custom-designed tiling arrays were developed specifically for *S. cerevisiae*. Given that the primary target of the analysis was to identify developmentally induced changes to the transcriptome, the sporulation program of *S. cerevisiae* was chosen for in-depth analysis. Sporulation encompasses both meiosis and spore formation.

Sporulation in many common laboratory strains is inefficient. The currently sequenced and annotated *S. cerevisiae* genome (available on the *Saccharomyces* Genome Database) is based upon strains derived from S288c, which sporulates with low synchrony and efficiency ($12 \pm 2\%$) (Ben-Ari et al. 2006). Therefore, we used the SK1 strain of *S. cerevisiae* that can sporulate synchronously and with high efficiency ($92 \pm 5\%$) (Ben-Ari et al. 2006). Although S288c and SK1 are strains of the same species, optimal array analysis required a new tiling design based upon the SK1 sequence (generously provided by Ed Louis and Richard Durbin as part of the “*Saccharomyces* Genome Resequencing Project” at the Sanger Institute). Our design is based upon the sequence release that was current in January 2008.

To precisely define unannotated transcript expression and architecture changes in an unbiased manner, probes were designed to cover 96% of both strands of the entire genome. Each array, consisting of 2.1 million features, was designed and manufactured in collaboration with Nimblegen. Probes were designed to be T_M balanced and ~ 60 nt in length. Probes on the same strand were offset by 10 nt from each other, whereas probes on opposite strands were offset by 5 nt. Therefore, any given nucleotide is represented in up to six different probes providing 10-nt resolution. Any given double-stranded region could be assayed with up to 12 different probes at 5-nt resolution with this design. Further information about array design as well as statistical information about the quality and distribution of our probe set is available at <http://groups.molbiosci.northwestern.edu/sontheimer/sk1meiosis.php>.

To analyze sporulation, seven evenly spaced time points were taken from 0 h (immediately after induction in sporulation media) to 12 h from three biological replicates that each exhibited a minimum of $>87\%$ sporulation efficiency. All

cultures taken for array analysis were in approximate synchrony with each other as determined by a microscopy count of the number of DAPI-staining bodies per cell. The increase of meiosis-specific transcripts and the decrease of mitosis-specific transcripts, as compared with vegetatively grown samples, were confirmed by quantitative reverse-transcription–polymerase chain reaction (qRT–PCR) analysis before array analysis. Haploid and diploid log phase cells grown in rich media were also harvested for comparison. In addition, we used genomic DNA in independent microarray hybridizations to help establish quality control and parameters for the segmentation analysis. Total RNA from each sample was used to generate labeled cDNA by anchored oligo(dT) priming. Actinomycin D was included in the reverse transcription (RT) to suppress mispriming (Perocchi et al. 2007).

SK1-centric annotation and genome browser

In order to analyze the tiling array data, we first had to annotate the SK1 genome. The current annotated version of the *S. cerevisiae* genome available on the *Saccharomyces* Genome Database is based upon the S288c background. We first built a whole-genomic syntenic alignment between S288c and SK1 by following the chaining and netting pipeline of UCSC (Kent et al. 2003; Rhead et al. 2010). Then, given this alignment, we lifted (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>) the SGD (Engel et al. 2010) annotations of S288c to SK1. Notably, out of 6717 protein-coding gene models of S288c, 6411 (95.4%) could be mapped to the SK1 assembly with the default parameters of liftOver. As for the remaining 306 genes (4.6%), 178 entries lost the start or the end base for at least one exon and 128 entries are partially deleted.

To facilitate downstream analysis a server was created to incorporate the SK1-based annotation with the conservation and tiling array analyses. The server is publicly available via <http://groups.molbiosci.northwestern.edu/sontheimer/sk1meiosis.php>. The design is based upon the UCSC Genome Browser platform (Rhead et al. 2010) and can be customized based upon user inputs.

Definition of expressed segment boundaries

Evaluation of expression patterns from the same locus across time requires the definition of the same segment boundaries to compare in each time point. To establish such boundaries, a single expressed segment was defined as any contiguous array feature that displayed signals above background (Fig. 1A,B;

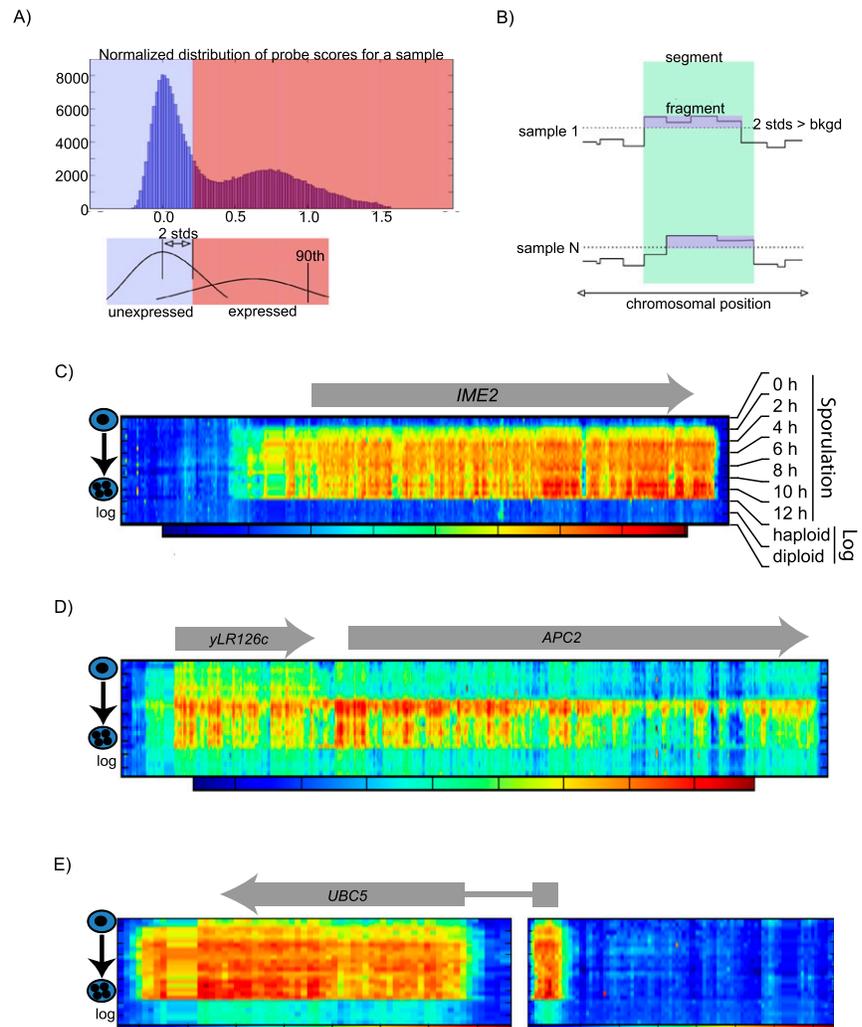


FIGURE 1. Segmentation pipeline. (A) For each array, the background was calculated from the histogram of normalized probe scores on all probes on that array (top). The signal was assumed to come from either expressed or unexpressed distributions. The background cut-off was taken as two standard deviations from the median of the unexpressed distribution (bottom). This first step of the pipeline establishes “fragments” of expression for each array. (B) To determine a constant segment of expression across samples, the union of the fragments was created as the second step of the pipeline. This established the segment boundaries and allowed an analysis of architecture changes across meiosis. (C) Heat map of a segment corresponding to a conventional gene (*IME2*) that is expressed specifically during meiosis. The x axis represents the genomic coordinates with the start and end of the arrow indicating the approximate location of the annotated start and stop codons, respectively. The array signals (three biological replicates for each time point) are stacked along the y axis with the beginning of meiosis at the top. The final six layers are from log-phase haploid and diploid cells. The heat map is generated for each segment individually, with the coloring indicated in the bar directly below the heat map (red and blue denote high and low signal, respectively). (D) Heat map of a single computationally defined segment containing *yLR126c* and *APC2* as an example of a segment that contains more than one annotated ORF. (E) Heat map of a single annotated ORF (*UBC5*) that was split between two segments.

Materials and Methods) in any of our array experiments. This approach allowed a rational analysis of the expressed segment across time. However, such a broad definition of an expressed segment increased the difficulty of identifying the precise 5' and 3' ends of transcripts. A server displaying the heat maps for each expressed segment is available for browsing via <http://groups.molbiosci.northwestern.edu/sontheimer/sk1meiosis.php>.

A total of 8407 expressed segments were identified from all of the array experiments, with an average segment length of 1702 nt. A total of 4496 of the expressed segments contained all or part of an annotated ORF; 3911 (46.5%) displayed a signal above background with no known overlapping annotation. Of the 3911 unannotated segments, 3064 (36.4% of total segments, 78% of unannotated segments) were antisense to a known ORF, while 847 were apparently intergenic transcripts.

Only 30 annotated, verified ORFs were never detected. The remainder of the annotated ORFs distributed into three classes: (1) a single segment containing a single annotated ORF, (2) multiple annotated ORFs contained within a single segment, or (3) one annotated ORF split between more than one segment. The majority of segments (3367 out of 4496, or 74.9%) fell into the first class (e.g., Fig. 1C). Regarding the second class, operon-like transcription units have been identified in other tiling array studies (David et al. 2006). Therefore, the multiple ORFs within a single segment may represent a bona fide biological phenomenon. However, since we adopted a relatively low-stringency definition of an expressed segment, some or all of these could be distinct transcripts that abut or overlap each other. An example of a segment containing multiple ORFs is shown in Figure 1D. The third class, single ORFs split between multiple segments, may either represent truly distinct transcripts or may reflect post-transcriptional processing events. For example, *UBC5* is an intron-containing gene that was split into two expressed segments by our algorithm (Fig. 1E). The splitting of *UBC5* in our algorithm most likely reflects the highly efficient processing of this transcript, leading to below-background signal in the intron. The vast majority (92%) of intron-containing genes that were detected on the array and annotated in SK1 were correctly identified as a single expressed segment in our algorithm. Only 14 intron-containing genes were split between two expressed segments. Recent work has demonstrated that the majority of splicing occurs before termination and coincident with polymerase pausing (Alexander et al. 2010; Brody et al. 2011). However, unspliced polyadenylated RNAs have long been detected (e.g., Tardiff et al. 2006). Therefore, our arrays were sensitive enough to detect the relatively low expression levels of the average intron (i.e., those that are spliced less efficiently than *UBC5* mRNA) before splicing and after polyadenylation.

Like other annotations of S288c, the EST track was remapped to the SK1 assembly as part of the SK1 annotation described above. In total, there are 32,100 items

in the EST track. A total of 3017 alignments were excluded due to low-quality or ambiguous alignments, as defined previously (Zhang et al. 2007). The filtered EST annotation track and the 8406 segments generated in this project were then intersected based on chromosomal coordinates and strandedness. As expected, given that public EST data are seldom derived from cells undergoing meiosis, only 3792 (45%) of the segments overlap with at least one EST sequence. Since ESTs are usually partial and thus cannot be viewed as complete transcripts, we can only use them to infer whether our segmentation algorithm mistakenly breaks one gene into more than one segment, i.e., whether one EST overlaps with different segments. The overwhelming majority (3555 or 94%) of the 3792 segments appears correct by this test. We further focused on 3747 unannotated segments and only 659 (18%) overlap with at least one EST. Again, a majority (556 or 84%) of the 659 segments were unlinkable with other segments via bridging ESTs. Such a result suggests that our segmentation algorithm generally succeeds in calling individual genes as individual segments.

Novel analytical methodology to detect meiosis-induced changes in transcript architecture

Once the boundaries of all of the expressed segments were defined, the expression patterns within the boundaries were examined for more complex changes in the architecture of the expressed transcript. Non-negative matrix factorization (NNMF) (Brunet et al. 2004) was used to identify segments containing more than one “transcript unit” (TU) (Fig. 2A). In brief, each segment was factored to contain the minimum number of transcription units that captures as much expression information as possible. The analysis was done in an iterative fashion. To start with, the segment was factored into one transcription unit and then Principal Component Analysis was performed on the error matrix. If the energy of the error matrix was contained within the first principal component, then the presence of another transcription unit was postulated. The segment was then factored with two transcription units, and the error matrix examined again. This process was repeated until there was no structure left over in the error matrix. Further description along with pseudocode is available in the Supplemental Material.

The NNMF analysis can be visualized as a map function that computes the minimal non-negative matrix factorization for each of the segments (Fig. 2A). For each segment identified, a detailed heat map can be seen next to SK1 genome coordinates with the NNMF trace below the heat map.

The NNMF analysis recognized statistically significant changes in architecture for 699 expressed segments, representing >8% of the total. A total of 398 of these 699 segments were associated with more than one ORF annotation.

Of the expressed segments displaying changes with only a single annotated ORF, no functional categories were over-

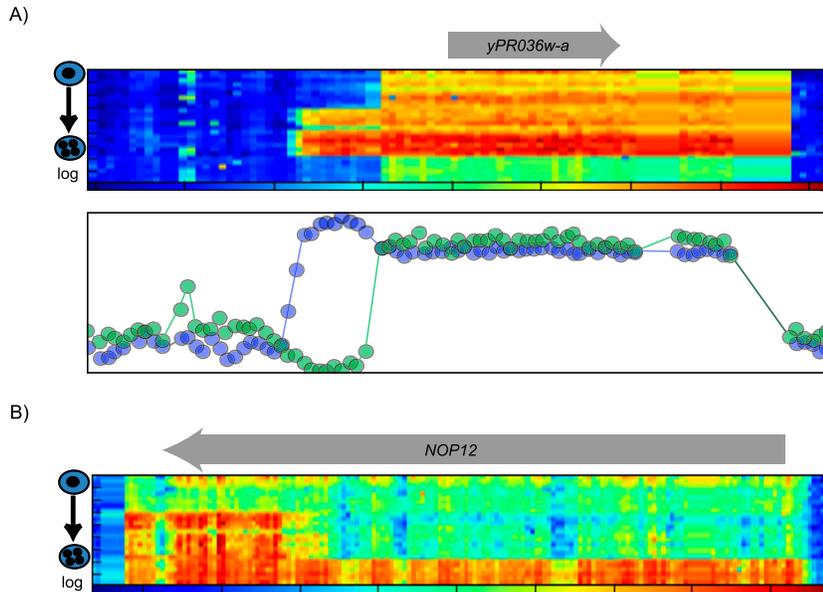


FIGURE 2. Factorization of segments to identify architecture changes at the 5' end. (A) Each segment identified was subjected to NNMF analysis to identify potential changes in transcript architecture. An example of an architecture change at the 5' end is shown at the *yPR036w-a* locus. The factorization can be visualized as a map function seen directly below the heat map of the segment. All segment heat maps and NNMF maps can be viewed at <http://groups.molbiosci.northwestern.edu/sontheimer/sk1meiosis.php>. (B) An example of extreme 5' end regression at the *NOP12* locus.

represented by GO term analysis. Any location within a transcript could be subject to a change in architecture. These changes can be broken down into three classes relative to the transcript detected in log phase or early meiosis: 5' changes, 3' changes, or internal changes to the body of the transcript.

Transcript architecture changes at the 5' end

The 5' end of a transcript is determined by multiple proteins that establish transcription initiation by RNA Polymerase II (RNAPII). While the 3' end is determined primarily by RNA sequences in the transcript itself, the 5' end is most often determined by upstream sequences in the promoter region. Among the set of transcripts with architecture changes at the 5' end, extension and regression were equally likely during sporulation. However, in contrast to the 3' end, whose change does not involve coding regions, several regressed ends invade into the coding sequence, thereby changing the encoded protein associated with this sequence. To our knowledge, nearly all examples of alterations at the 5' end presented here have not been previously reported.

The most significantly recessed transcript is apparent at the *NOP12* locus (Fig. 2B). The *NOP12* gene is annotated to be 1380 nt long. During meiosis, a drastic change occurs in mid-to-late meiosis, yielding a form that is more than

1000 nt shorter at the 5' end of the transcript. The shorter transcript is as abundantly expressed during meiosis as the long form is during log-phase growth. Both forms were validated by sequencing of 5' RACE fragments. *NOP12* is a nucleolar protein involved in pre-25S rRNA processing and contains a single RRM (Wu et al. 2001). Interestingly, the shortened form is predicted to initiate shortly after the RRM and contains an in-frame start codon, indicating that a short form of Nop12 could exist that does not have the capability to bind RNA.

5' extension was also observed. For example, the *yPR036w-a* locus displays an abrupt increase in 5' UTR length at mid-meiosis (Fig. 2A). Interestingly, the extension is predicted to add an upstream ORF (uORF) (Hinnebusch 1993; Vilela and McCarthy 2003) to the 5' UTR. Both the short and long forms were confirmed to exist by 5' RACE and sequence analysis. If the uORF is functional, this represents the first report, to our knowledge, that couples a change in yeast transcript architecture to uORF

control and, presumably, translational regulation. Alternatively, the uORF may not function to regulate the downstream ORF, but may have an independent function. The 5' extension occurs during middle meiosis and persists through the end of meiosis and spore formation. Whether the alternate isoform is important for completion of meiosis, spore formation, or germination is unknown. Two additional ORFs, *yLR307c-a* and *yOR020w-a*, also small and with unknown function, display 5' UTR extensions with apparent uORF addition. In each case, the function of the ORF, the mechanism of the architecture change, and the function of the potential uORF remain unknown.

Transcript architecture changes at the 3' end

The 3' end of an RNAPII transcript is created by a multi-component cleavage and polyadenylation machinery that must first define the site at which the transcript is severed from the elongating polymerase (Keller and Minvielle-Sebastia 1997; Millevoi and Vagner 2010). The same machinery must then create a nontemplate-derived poly(A) tail. The cleavage site is known to be subject to regulation, and changing the cleavage site can have a profound impact on the stability, processing, and regulation of the transcript.

During sporulation, both shortening and lengthening of 3' UTRs was observed (Fig. 3A,B). However, 3' UTR

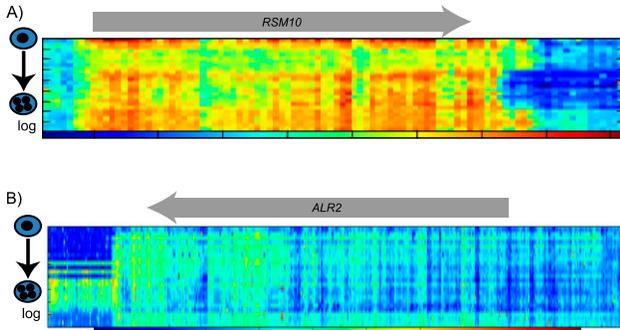


FIGURE 3. Architecture changes at the 3' end. (A) Example of 3' end regression during meiosis at the *RSM10* locus. (B) Example of 3' end extension during meiosis at the *ALR2* locus.

shortening was more commonly observed than lengthening. For example, the *RSM10* locus displays a shortening of the 3' UTR at mid-meiosis through the end of meiosis. *RSM10* encodes one of the few mitochondrial ribosome proteins that are essential for viability. For annotated genes, the 3' terminus of the shortened transcript was never observed to invade the ORF during sporulation. Although not as common, 3' UTR lengthening was also observed, and an example (the *ALR2* locus) is shown in Figure 3B. *ALR2* encodes a nonessential, predicted Mg^{2+} transporter (MacDiarmid and Gardner 1998). Interestingly, the extension occurs at middle meiosis and partially extends across a predicted autonomous replicating sequence (ARS). At middle meiosis, the genome should have already duplicated. Whether the extension is correlated to replication or the block to re-replication is unknown.

In this analysis, 3' ends exhibit more heterogeneity than 5' ends. The boundaries between expressed and non-expressed regions were often less clear at the 3' end than at the 5' end. However, discrete changes in the termini are more prevalent at the 5' ends than at the 3' ends of the detected segments.

Architecture changes within the body of a transcript

The tiling array data set with the NNMF analysis can be used to detect candidate alternative splicing events if the alternative splice sites lie >50 nt from each other. Unique regulation of an intron-containing transcript was uncovered at the *MRK1* locus (Fig. 4). The presence of the intron and the correct splicing of the intron-containing form has been confirmed (Davis et al. 2000). Strikingly, the array signal reveals primarily exon 1 expression in early meiosis, with very little signal from exon 2. Conversely, exon 2 signal is very strong during late meiosis, with very little expression of exon 1. 5' RACE analysis confirms the presence of an exon-2-only form in late meiosis as well as correctly spliced isoforms at 0 h and in log phase. *MRK1* is the only intron-containing GSK3 β paralog in

S. cerevisiae. Sequence analysis of the RACE products predicts that the exon 2-only form retains the reading frame of the mature mRNA, leading to the prediction of possibly three functional proteins produced from this single locus. To our knowledge, this is the first example of differential, modular regulation of exon expression in *S. cerevisiae*.

Architecture changes in unannotated segments

A total of 116 segments displayed multiple transcript units (TUs) with no previous annotation. Two (*GDH2* and *DMC1*) of the 116 were known ORFs that had escaped annotation in the lift-over package designed to reannotate SK1. Three additional segments were annotated to snoRNAs. Since the cDNAs used to hybridize to the array were generated solely from anchored oligo(dT) primers, detection of the snoRNAs was surprising. snoRNAs are polyadenylated during pre-snoRNA processing and during quality control as a mark for degradation (LaCava et al. 2005; Grzechnik and Kufel 2008; Lemay et al. 2010). However, polyadenylated snoRNAs are thought to be short-lived, committed to either completion of pre-snoRNA processing or degradation. Since all experiments here were performed in a wild-type strain (with intact TRAMP and exosome complexes), the accumulation of polyadenylated forms at these loci is noteworthy. From the array experiments alone, it is not possible to determine whether the detected snoRNA forms are intermediates in processing/biogenesis or decay, or if a stable polyadenylated pool of snoRNA exists.

Only 28 of the 114 multiple TU segments displayed no overlap with an annotated segment on the opposite strand. Eighty-six of the 114 multiple TU segments display at least 50 nt of overlap with a segment on the opposite strand that does contain annotation, potentially indicating an antisense partner for these annotated ORFs.

The unannotated transcriptome

A total of 3910 of the 8407 identified expressed segments were not annotated; 3064 were antisense to known ORFs, and 847 were not associated with any known annotation on either strand (i.e., these segments appear to be intergenic

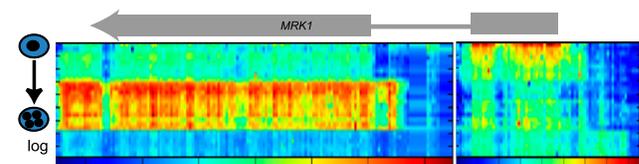


FIGURE 4. Architecture changes internal to the transcript: a novel exon expression pattern at the *MRK1* locus. Note that exon 1 is predominantly expressed during early meiosis and exon 2 is predominantly expressed during late meiosis.

transcripts). Similar to the transcripts associated with annotated regions, the antisense and isolated transcripts also display dynamic expression and architecture patterns. Although dynamic 5' and 3' architecture changes were observed, no transcripts in the unannotated transcriptome displayed any internal changes in architecture. Note that these are not “cryptic unstable transcripts” (CUTs) (Wyers et al. 2005; Neil et al. 2009; Xu et al. 2009), as the exosome was intact in the strain used in our analyses. The average length of the unannotated regions was 1081 nt, with a median value of 753 nt. By comparison, the average length of segments containing annotation was 2555 nt with a median value of 2118.

Interestingly, antisense transcripts were often detected on the opposite strand from introns, and we have termed these antisense transcripts “anti-introns” (Fig. 5A). Fifty-nine intron-containing segments displayed a signal on the opposite strand. Most often, the signal on the opposite strand was contained within the intron and did not extend into the exonic regions. While the antisense signal was often low, it usually showed higher expression levels than the corresponding sense-strand intron. The most intriguing examples of this occur at ribosomal protein genes (RPGs), which are induced at late meiosis, yet show lower intronic signal than during log phase. Since the exonic regions display a large increase in expression, it seems unlikely that the intronic antisense transcript acts to decrease ORF expression levels. Alternatively, the antisense transcript may assist (or at least correlate with) active splicing by an unknown mechanism, potentially accounting for the decreased intronic signal during peak meiotic expression. The intronic signal in log phase most likely comes from unspliced pre-mRNA, not excised intron, because the cDNAs for the array were generated with anchored oligo(dT) primers, and excised introns are not known to be stably polyadenylated. Therefore, the decreased intronic signal during meiosis is unlikely to result from increased clearance of the excised intron, suggesting instead that it results from increased splicing efficiency, increased cotranscriptional splicing, or increased pre-mRNA turnover.

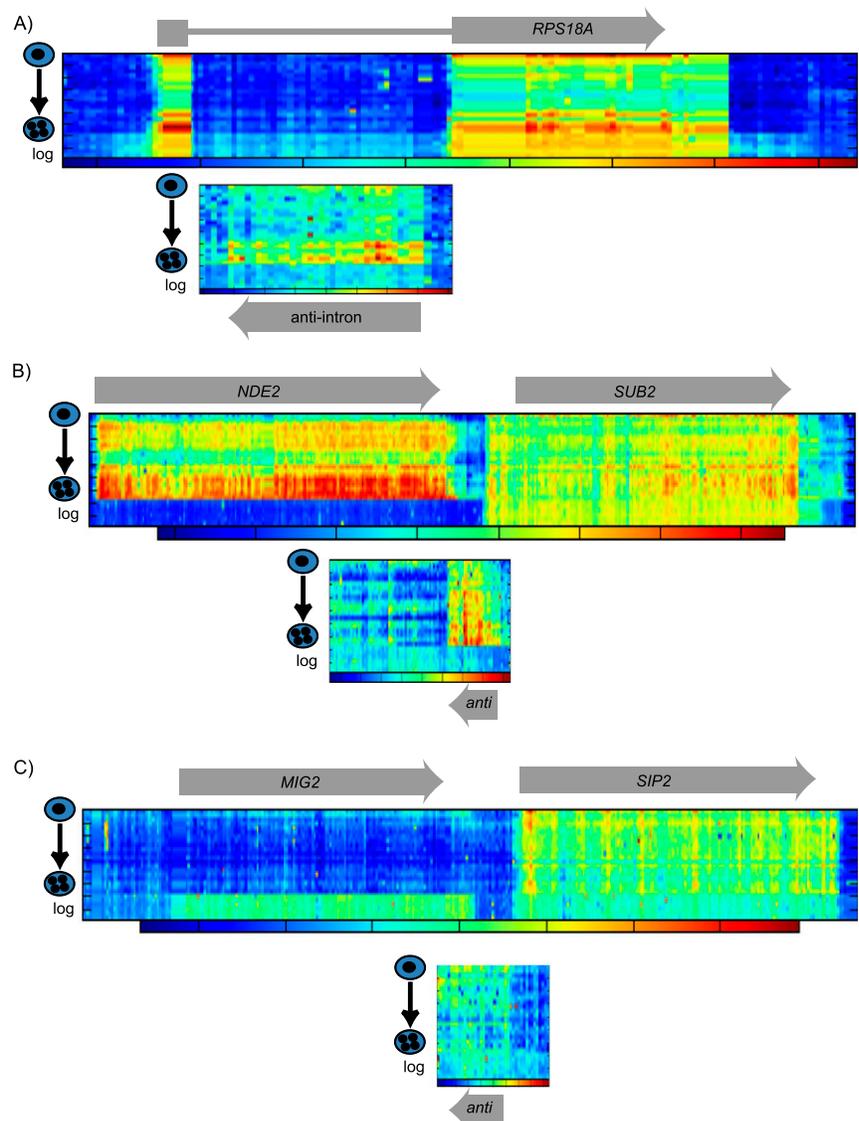


FIGURE 5. The unannotated transcriptome. (A) An anti-intron transcript at the *RPS18A* locus. A select group of intron-containing genes displayed signal on the opposite (antisense) strand that correlated to the position of the intron. The segment opposite the intron has been termed the anti-intron. (B) An anti-intergenic junction transcript at the *NDE2/SUB2* locus. A select group of closely spaced genes displayed signal on the opposite strand in the gap between the annotated genes. In this example, expression of the anti-intergenic junction transcript correlates to a greater degree with expression from the upstream transcript, *NDE2*. (C) An anti-intergenic junction transcript at the *MIG2/SIP2* locus. In contrast to B, this anti-intergenic junction segment correlates to a greater degree with expression from the downstream transcript, *SIP2*.

Isolated unannotated segments showed an average length of 1265 nt. Conservation analysis was undertaken to further characterize the isolated transcripts. Starting from the SK1 *S. cerevisiae* sequence, phastCONS analysis was performed to identify conserved regions across the genome (Siepel et al. 2005). Details of the phastCONS implementation are available at our website, <http://groups.molbiosci.northwestern.edu/sontheimer/sk1meiosis.php>.

Some isolated transcripts are located at intriguing locations across from the interval between adjacent genes

on the opposite strand, and we have termed these “anti-intergenic junction” transcripts (Fig. 5B,C). We observed instances of this arrangement in which the unannotated transcript correlates with either the upstream (Fig. 5B) or the downstream gene (Fig. 5C). Therefore, regulation of this class of unannotated transcripts is unlikely to share a single common mechanism. Unannotated transcript expression that correlates with the divergently transcribed annotated gene may reflect the promoter-proximal transcripts (PROMPTs) observed previously (Preker et al. 2008). However, unannotated transcripts with expression profiles that correlate with the convergently transcribed annotated gene cannot reflect such a mechanism. It is possible that the transcription of the unannotated anti-intergenic junction transcript promotes termination of the upstream gene and limits inappropriate readthrough to the second gene in the segment.

DISCUSSION

We have analyzed the noncoding transcriptome over the conserved process of sporulation in *S. cerevisiae* using custom-designed, high-resolution tiling arrays. Remarkably, a large number of expressed segments were identified that undergo a change in architecture during meiosis. This work identifies transcript architecture changes on a genome-wide scale across a developmental transition in an unbiased manner. We have also identified a large number of previously unidentified, unannotated transcripts without artificially limiting endogenous RNA turnover pathways. A significant number of these unannotated transcripts were induced or showed a change in expression or architecture during meiosis. This study highlights the dynamic potential of transcriptome regulation during development for both annotated and unannotated transcripts.

Plasticity of transcript architecture during meiosis

We anticipated discovery of novel alterations of RNA expression patterns, but the depth and breadth of transcriptome regulation was truly striking. Not only do yeast cells commonly control when a transcript is expressed, they also can add or subtract hundreds of nucleotides from their termini in a developmentally specific manner at a large number of genomic loci. We consider it highly unlikely that these architecture changes reflect modifications of existing transcripts. A more likely scenario is that transcription factors or 3' end processing factors drive the generation of new transcripts with different architecture from the same locus. These changes are not limited to the UTRs, but also include invasion into the coding region as well.

It is important to note that our analysis pipeline excels at identifying expressed segments with potential changes in transcript architecture. Categorization of the architecture changes were made upon subjective analysis of the array

heat maps, and quantitative analysis of architecture changes based on array data alone is not within the capabilities of this technology. Although we cannot quantitatively determine the total number and type of architecture changes that manifest during meiosis, we can state that the dramatic architecture changes do occur and that the transcriptome, both coding and noncoding, appears to be unexpectedly dynamic in its structure. In fact, to our knowledge, none of the examples of architecture changes highlighted here have been previously reported in the literature. Further work on a gene-by-gene basis will be required to determine the full extent and purpose of transcript architecture changes.

Recent RNA-seq experiments identified widespread changes in 3' end length in yeast (Yoon and Brem 2010). Interestingly, their examination of dithiothreitol (DTT) stress revealed numerous alternative polyadenylation events that result in truncation within the coding region. For our analysis of meiosis, numerous shortened UTRs appear, yet few invade the coding region. The differences may reflect distinctive strategies that the cell applies to 3' end regulation during stress versus development.

Splicing, or internal changes to transcript architecture, has been so intensively studied in yeast that the discovery of new intronic expression patterns reported here is noteworthy. Alternative splicing, commonly used in higher eukaryotes, confers the advantage of added “modularity” in transcript architecture. Although alternative splicing is used sparingly in budding yeast (Ares et al. 1999; Davis et al. 2000; Juneau et al. 2009; Mishra et al. 2011), the modularity afforded by exon/intron arrangements is also used in this simpler eukaryote as exemplified by the *MRK1* locus, which encodes a glycogen synthase kinase 3 β (GSK3 β) paralog. The *MRK1* transcripts can differentially express a correctly spliced form, a form that only contains the 5'-terminal exon or a form that lacks the 5'-terminal exon. The second (3'-terminal) exon of *MRK1* has the potential to be translated in the same reading frame as the spliced mRNA and would encode an apparently intact, highly conserved kinase domain, but lacking the additional N-terminal amino acids encoded in the first exon. This modularity may provide added opportunities for the cell to regulate the expression, localization, or activity of these critical enzymes. The meiosis-specific mRNA isoforms are expressed at levels as high or higher than the annotated mature isoform, which suggests that they may represent functional transcripts. These two expressed segments highlight the exquisite interplay that must be required between the splicing apparatus, the transcriptional machinery, and the signaling cascades required to precisely time such changes in architecture.

The mechanisms driving architecture changes are unknown and must be resolved on a gene-by-gene basis. The simplest and most likely mechanism behind most 5' UTR architecture changes would surely be a change in transcription itself. For example, a shorter 5' UTR could result from enhancement of a weak downstream promoter site by

a newly activated transcription factor. Likewise, changes in the 3' UTR could be driven by cotranscriptional choice of the 3' cleavage site or by a downstream post-transcriptional processing step. The discovery of the novel transcript architecture changes reported here will likely lead to the discovery of novel mechanisms that regulate the changes themselves.

The vast unannotated transcriptome

We report not only the existence of new unannotated transcripts but also that their architectural patterns change. Not only is the unannotated genome bursting with transcription, but the “dark matter” of the genome can undergo dynamic alterations in transcript architecture over development.

Not only do unannotated transcripts exhibit dynamic expression patterns in terms of both architecture and overall expression, but the regulation of this expression appears to play a critically important role in the progression of meiosis. A recent tiling array study discovered an unexpected role for *RRP6* and unannotated RNA stability during yeast meiosis (Lardenois et al. 2011). Rrp6 is post-translationally down-regulated during meiosis, which correlates with a coincident increase in unannotated RNA expression. Cells lacking Rrp6 are unable to initiate replication and MI.

Although unannotated, novel transcripts are not necessarily noncoding. Early annotation efforts required a minimum of 100 codons between start and stop codons to be annotated as an ORF. Despite the tremendous value of these early annotation efforts, the existence of small ORFs and microORFs went virtually unnoticed. A prime example is the *yPR036w-a* locus, which was not annotated as an ORF until 2001. At only 204 nt, this ORF escaped detection during early annotation efforts and was not included in any ORF array studies or tagging/deletion library collections. Our results not only confirm its existence, but also reveal a startling and abrupt extension of the 5' UTR that occurs during mid-meiosis. The reason for the extension is unknown, but it is intriguing that the cell would purposefully regulate the length of the 5' UTR for a small ORF with a mysterious function.

A total of 78% (3064 out of 3910) of the unannotated expressed segments reported here lie antisense to a known coding gene. Appreciation is growing for the dramatic role natural antisense RNAs can play in gene regulation (Lapidot and Pilpel 2006; Faghihi and Wahlestedt 2009). In most organisms, and in the few examples known in yeast, antisense RNAs play a negative regulatory role in controlling the expression of their sense partners (Hongay et al. 2006; Berretta et al. 2008; Camblong et al. 2009; Faghihi and Wahlestedt 2009). The antisense RNAs are most likely not due to reverse transcription error during array probe synthesis, as all probe synthesis was performed in the

presence of Actinomycin D to limit priming errors (Perocchi et al. 2007).

The unannotated transcripts described here are not likely to be CUTs. These are typically identified by increased accumulation in a nuclear-exosome-deficient (*rrp6*) mutant background (Wyers et al. 2005; Neil et al. 2009). All studies here were performed in a wild-type laboratory strain of SK1, and all transcripts reported were detected above background. Finally, cDNA was prepared for microarray analysis using anchored oligo(dT) primers. Therefore, all signals from the array must be due to RNAs containing a poly(A) tail or an extended poly(A) tract. Accordingly, the unannotated messages reported here are likely RNAPII products processed by the traditional 3' end formation machinery.

Closing notes

This work has exposed a highly unexpected degree of plasticity in the transcriptome of budding yeast. On a genome-wide level, it is now clear that transcript architecture alterations and dynamic ncRNA expression are common occurrences during meiosis and sporulation. Moreover, on a transcript-specific level, each dynamic change in architecture or expression will undoubtedly reveal a fascinating story of gene regulation in its own right. We predict that many other stress conditions or developmental cascades will also induce novel examples of transcriptome regulation.

MATERIALS AND METHODS

Strains and sporulation protocols

The yeast SK1 strain DKB98 (MAT a/ α , *ho::LYS2*⁺, *lys2*⁺, *ura3*⁺, *leu2::hisG*⁺, *his4X::LEU2/his4B::LEU2*, *arg4-Nsp1/arg4-BglII*) was graciously provided by Doug Bishop and used for sporulation and diploid analysis. The haploid sister strain DKB83 (MAT a, *lys2*, *ho::LYS2*, *ura3*, *leu2::hisG*, *his4X::LEU2*, *arg4-Nsp1*) was used for haploid analysis.

To sporulate, the Bishop Lab protocol was used (<http://bishoplabs.d.uchicago.edu/protocols/SporulationProtocol.pdf>). In brief, diploids were streaked from freezer stock. Single colonies were used to inoculate rich YPDA media. A portion of the colony was patched to an SPM plate to confirm the ability to sporulate. After overnight growth at 30°C, the saturated culture was diluted into SPS presporulation media. These cultures were confirmed to reach log phase in SPS and were grown overnight, diluted to stay in log phase. Cultures were never permitted to exceed 24 h total growth in SPS. Cultures were harvested at an OD₆₀₀ of ~0.8. Cultures were not harvested if below 0.5 or above 1.0. Cells were harvested by centrifugation, washed once with sterile water, and resuspended in SPM media. The 0-h timepoint was immediately taken and the remainder of the culture was grown at 30°C with rapid shaking. For continuity of analysis, time points were taken every 2 h for 12 h from the same flask. A final time point was

taken at 24 h to confirm complete sporulation. At each 0–12-h time point, samples were collected by centrifugation and frozen for RNA analysis. In addition, 0.5 mL of culture was mixed with an equal volume of ethanol for DAPI counting.

RNA preparation

Frozen cell pellets were resuspended in extract buffer and lysed in the presence of acid phenol: chloroform: isoamyl alcohol (125:5:1) (Ambion) and 0.5-mm Zirconium/Silicon beads. The cells were bead-beaten for four cycles of 1 min beating with a 1-min rest on ice between each cycle. The samples were then heated for 10 min at 65°C with vortexing every 2 min. TE buffer was added and the samples were centrifuged in the cold for 10 min. The aqueous phase was extracted once more with acid phenol:chloroform and again with chloroform. The purified RNA was precipitated, resuspended in water, and stored at –80°C.

Array preparation

cDNA synthesis, dye coupling, array hybridization, and array scanning were performed by Nimblegen using their standard protocols. Note that cDNA synthesis was performed in the presence of actinomycin D to reduce the occurrence of RT artifacts (Perocchi et al. 2007).

Tiling array design

SK1 sequence was graciously provided by Ed Louis and Richard Durbin as part of the “*Saccharomyces* Resequencing Project” carried out at the Sanger Institute. Approximately 60-mer, T_M -balanced probes were designed for both strands of the genome. The probes are tiled across the genome with a 10-nt offset on the same strand and a 5-nt offset on opposite strands. Areas that escaped coverage include the rDNA island and other repeat-rich regions.

Array analysis method

Array data was normalized to the 95th percentile. The background threshold was calculated separately for each array by creating a histogram of the signal from all probes on the array. The histogram displays a bimodal distribution on all arrays, except the genomic DNA control arrays. The lower distribution was assumed to arise from nonhybridizing probes that represent background signal. Two standard deviations above the mean of the lower distribution were taken as the cut-off value for reliable signal intensity.

To determine the boundaries of the expressed segments, all 27 microarrays were collectively considered. Any probes that displayed signal above background and that overlapped with another probe in any array were considered as part of the same expressed segment. Data for all expressed segments are available online at <http://groups.molbiosci.northwestern.edu/sontheimer/sk1meiosis.php>. A lower segment length boundary of 50 nt was applied to remove array noise.

To determine whether multiple isoforms can exist at the same location but at different times, non-negative matrix factorization (NNMF) was used (Brunet et al. 2004). In brief, NNMF calculates an information score of the signal in the expressed segment.

Principal Component Analysis is then applied to the error matrix to determine whether the data fit the model or whether other signals exist. This process is reiterated until no structure remains in the error matrix. More details are available in the Supplemental Material.

RACE

RACE analysis was performed using the SMARTer RACE kit (Clontech). In brief, RNA was verified to be intact by gel electrophoresis and pure by Nanodrop spectrophotometer readings. cDNA was prepared using an anchored oligo(dT) primer, the SMARTer IIa oligo, and SMARTscribe reverse transcriptase. Amplifications were carried out with a gene-specific primer and a primer complementary to the SMARTer IIa oligo. RACE reactions were gel purified and TOPO-TA cloned (Invitrogen). Individual clones were then sequenced. Alignment was carried out using BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and CLUSTAL (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>).

SK1 annotation

Annotations are based on a derivative of the S288c strain and were downloaded from the *Saccharomyces* Genome Database on November 2009. A lift-over package was used to assign annotations on the SK1 sequence provided by the Sanger Institute (detailed implementation methods are available at <http://groups.molbiosci.northwestern.edu/sontheimer/sk1meiosis.php>). PhastCONS analysis was implemented for the SK1 sequence (detailed information is available at <http://groups.molbiosci.northwestern.edu/sontheimer/sk1meiosis.php>) (Siepel et al. 2005). To display the sequence, tiling array, annotations, and conservation analysis, a new server was created based on the UCSC genome browser (Kent et al. 2002; Kuhn et al. 2009). The server can be accessed at <http://groups.molbiosci.northwestern.edu/sontheimer/sk1meiosis.php>.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank the Keck Biophysics Facility (Northwestern University) for use of instrumentation and the Center for Genetic Medicine (Northwestern University) for sequencing analysis. We are also indebted to Ed Louis and Richard Durbin for access to SK1 sequence data. We thank Doug Bishop (University of Chicago) for strains, assistance with sporulation, and for helpful comments. K.S.K.G. was supported by a postdoctoral fellowship from the American Cancer Society. This work was funded by the Chicago Biomedical Consortium with support from The Searle Funds at the Chicago Community Trust.

Received September 23, 2011; accepted January 27, 2012.

REFERENCES

Alexander RD, Innocente SA, Barrass JD, Beggs JD. 2010. Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell* 40: 582–593.

- Ares M Jr, Grate L, Pauling MH. 1999. A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* **5**: 1138–1139.
- Ben-Ari G, Zenvirth D, Sherman A, David L, Klutstein M, Lavi U, Hillel J, Simchen G. 2006. Four linked genes participate in controlling sporulation efficiency in budding yeast. *PLoS Genet* **2**: e195. doi: 10.1371/journal.pgen.0020195.
- Benanti JA, Matyskiela ME, Morgan DO, Toczyski DP. 2009. Functionally distinct isoforms of Cdk1 are differentially regulated by APC/C-mediated proteolysis. *Mol Cell* **33**: 581–590.
- Berretta J, Pinskaya M, Morillon A. 2008. A cryptic unstable transcript mediates transcriptional *trans*-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev* **22**: 615–626.
- Brody Y, Neufeld N, Bieberstein N, Causse SZ, Bohnlein EM, Neugebauer KM, Darzacq X, Shav-Tal Y. 2011. The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol* **9**: e1000573. doi: 10.1371/journal.pbio.1000573.
- Brunet JP, Tamayo P, Golub TR, Mesirov JP. 2004. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci* **101**: 4164–4169.
- Camblong J, Iglesias N, Fickentscher C, Dieppois G, Stutz F. 2007. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* **131**: 706–717.
- Camblong J, Beyrouthy N, Guffanti E, Schlaepfer G, Steinmetz LM, Stutz F. 2009. *Trans*-acting antisense RNAs mediate transcriptional gene cosuppression in *S. cerevisiae*. *Genes Dev* **23**: 1534–1545.
- Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I. 1998. The transcriptional program of sporulation in budding yeast. *Science* **282**: 699–705.
- Clancy MJ. 1998. Meiosis: step-by-step through sporulation. *Curr Biol* **8**: R461–R463.
- David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM. 2006. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci* **103**: 5320–5325.
- Davis CA, Grate L, Spingola M, Ares M Jr. 2000. Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res* **28**: 1700–1706.
- Engel SR, Balakrishnan R, Binkley G, Christie KR, Costanzo MC, Dwight SS, Fisk DG, Hirschman JE, Hitz BC, Hong EL, et al. 2010. *Saccharomyces* Genome Database provides mutant phenotype data. *Nucleic Acids Res* **38**: D433–D436.
- Faghihi MA, Wahlestedt C. 2009. Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol* **10**: 637–643.
- Grzechnik P, Kufel J. 2008. Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol Cell* **32**: 247–258.
- Hinnebusch AG. 1993. Gene-specific translational control of the yeast *GCN4* gene by phosphorylation of eukaryotic initiation factor 2. *Mol Microbiol* **10**: 215–223.
- Hongay CF, Grisafi PL, Galitski T, Fink GR. 2006. Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell* **127**: 735–745.
- Jacquier A. 2009. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat Rev Genet* **10**: 833–844.
- Juneau K, Nislow C, Davis RW. 2009. Alternative splicing of *PTC7* in *Saccharomyces cerevisiae* determines protein localization. *Genetics* **183**: 185–194.
- Keller W, Minvielle-Sebastian L. 1997. A comparison of mammalian and yeast pre-mRNA 3'-end processing. *Curr Opin Cell Biol* **9**: 329–336.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996–1006.
- Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci* **100**: 11484–11489.
- Kuhn RM, Karolchik D, Zweig AS, Wang T, Smith KE, Rosenbloom KR, Rhead B, Raney BJ, Pohl A, Pheasant M, et al. 2009. The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res* **37**: D755–D761.
- LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D. 2005. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**: 713–724.
- Lapidot M, Pilpel Y. 2006. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep* **7**: 1216–1222.
- Lardenois A, Liu Y, Walther T, Chalmel F, Evraud B, Granovskaia M, Chu A, Davis RW, Steinmetz LM, Primig M. 2011. Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. *Proc Natl Acad Sci* **108**: 1058–1063.
- Lemay JF, D'Amours A, Lemieux C, Lackner DH, St-Sauveur VG, Bahler J, Bachand F. 2010. The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nucleolar RNAs. *Mol Cell* **37**: 34–45.
- MacDiarmid CW, Gardner RC. 1998. Overexpression of the *Saccharomyces cerevisiae* magnesium transport system confers resistance to aluminum ion. *J Biol Chem* **273**: 1727–1732.
- Mayr C, Bartel DP. 2009. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**: 673–684.
- Millevoi S, Vagner S. 2010. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* **38**: 2757–2774.
- Mishra SK, Ammon T, Popowicz GM, Krajewski M, Nagel RJ, Ares M Jr, Holak TA, Jentsch S. 2011. Role of the ubiquitin-like protein Hub1 in splice-site usage and alternative splicing. *Nature* **474**: 173–178.
- Mitchell AP. 1994. Control of meiotic gene expression in *Saccharomyces cerevisiae*. *Microbiol Rev* **58**: 56–70.
- Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A. 2009. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**: 1038–1042.
- Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**: 457–463.
- Perochi F, Xu Z, Clauder-Munster S, Steinmetz LM. 2007. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res* **35**: e128. doi: 10.1093/nar/gkm683.
- Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**: 1851–1854.
- Primig M, Williams RM, Winzeler EA, Tevzadze GG, Conway AR, Hwang SY, Davis RW, Esposito RE. 2000. The core meiotic transcriptome in budding yeasts. *Nat Genet* **26**: 415–423.
- Rhead B, Karolchik D, Kuhn RM, Hinrichs AS, Zweig AS, Fujita PA, Diekhans M, Smith KE, Rosenbloom KR, Raney BJ, et al. 2010. The UCSC Genome Browser database: update 2010. *Nucleic Acids Res* **38**: D613–D619.
- Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. 2008. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**: 1643–1647.
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**: 1034–1050.
- Takagaki Y, Manley JL. 1998. Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Mol Cell* **2**: 761–771.
- Takagaki Y, Seipelt RL, Peterson ML, Manley JL. 1996. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* **87**: 941–952.

- Tardiff DF, Lacadie SA, Rosbash M. 2006. A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly post-transcriptional. *Mol Cell* **24**: 917–929.
- Vershon AK, Pierce M. 2000. Transcriptional regulation of meiosis in yeast. *Curr Opin Cell Biol* **12**: 334–339.
- Vilela C, McCarthy JE. 2003. Regulation of fungal gene expression via short open reading frames in the mRNA 5' untranslated region. *Mol Microbiol* **49**: 859–867.
- Wu K, Wu P, Aris JP. 2001. Nucleolar protein Nop12p participates in synthesis of 25S rRNA in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **29**: 2938–2949.
- Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, Regnault B, Devaux F, Namane A, Seraphin B, et al. 2005. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**: 725–737.
- Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM. 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**: 1033–1037.
- Yoon OK, Brem RB. 2010. Noncanonical transcript forms in yeast and their regulation during environmental stress. *RNA* **16**: 1256–1267.
- Zhang Y, Li J, Kong L, Gao G, Liu QR, Wei L. 2007. NATsDB: Natural Antisense Transcripts DataBase. *Nucleic Acids Res* **35**: D156–D161.