The translational landscape of fission-yeast meiosis and sporulation

Caia D S Duncan & Juan Mata

Sexual development in *Schizosaccharomyces pombe* culminates in meiosis and sporulation. We used ribosome profiling to investigate the translational landscape of this process. We show that the translation efficiency of hundreds of genes is regulated in complex patterns, often correlating with changes in RNA levels. Ribosome-protected fragments show a three-nucleotide periodicity that identifies translated sequences and their reading frame. Using this property, we identified 46 new translated genes and found that 24% of noncoding RNAs are actively translated. We also detected 19 nested antisense genes, in which both DNA strands encode translated mRNAs. Finally, we identified 1,735 translated upstream open reading frames (ORFs) in leader sequences. In *S. pombe*, in contrast with *Saccharomyces cerevisiae*, sexual development is not accompanied by large increases in upstream ORF use, thus suggesting that this is an organism-specific adaptation, not a general feature of developmental processes.

Fission yeast diploid cells undergo sexual differentiation (meiosis and sporulation) upon nitrogen starvation. This process is accompanied by a complex gene expression program in which more than 50% of the genome is regulated. Microarray studies have shown that changes in RNA levels occur in successive expression waves that coincide with major biological events; genes can be grouped into starvation- and pheromone-induced genes, early genes (premeiotic S phase and recombination), middle genes (meiotic divisions and spore formation) and late genes (spore maturation). The induction of most of these groups is mediated by meiosis-specific transcription factors, although regulation of mRNA decay by RNA-binding proteins is also involved. More recently, RNA-sequencing experiments revealed that *S. pombe* cells express hundreds of meiosis-specific noncoding RNAs (ncRNAs). Although gene expression during sexual differentiation has been extensively studied, nothing is known about the contribution of translational control to this process.

Ribosome profiling can provide a genome-wide view of translation with single-nucleotide resolution. This approach is based on the isolation and sequencing of ribosome-protected mRNA fragments (RPFs), and it can be used to identify translated regions and to estimate mRNA translational efficiency. This approach has recently been applied to the meiotic program of *S. cerevisiae*, revealing extensive translational control and a meiosis-specific increase in the translation efficiency of hundreds of genes. In *S. pombe*, in contrast with *S. cerevisiae*, sexual development is not accompanied by large increases in upstream ORF use, thus suggesting that this is an organism-specific adaptation, not a general feature of developmental processes.

**RESULTS**

**Ribosome profiling in *S. pombe***

We carried out ribosome profiling of *S. pombe* diploid cells undergoing meiosis and sporulation. To achieve good synchrony, we used thermosensitive mutants of the Pat1 meiotic inhibitor. Diploid cells were blocked in G1 by nitrogen starvation, and they entered meiosis synchronously upon inactivation of the Pat1 kinase. In addition, we performed ribosome profiling in wild-type haploid cells growing vegetatively. Ribosome profiling involves the purification of RPFs, which are used to generate a library containing adaptors for Illumina sequencing. In parallel, a second library is produced from fragmented mRNAs. The first library allows the identification and quantification of translated regions, and the second is used to estimate mRNA levels. In all current protocols, the production of both libraries involves PCR-based amplification. As part of our library generation strategy, we used unique molecular identifiers (UMIs). The primer used for reverse transcription contained a random barcode sequence that uniquely tagged each cDNA fragment. Reads that contained identical sequences with the same
barcodes were very likely to have originated from the same RNA fragment, and only one of them was retained. Therefore, the reads that we used for analysis constituted a truly nonredundant data set. This approach helps avoid artifacts of PCR amplification and sampling and can be used to estimate the complexity of the original library. 83% of reads that did not map to rRNAs were unique, thus indicating that the complexity of the original libraries was high (Supplementary Table 1). Overall, we reduced 707 million reads to 250 million unique ones (Supplementary Table 1). As expected, RPFs showed a strong bias in their distribution along mRNAs: after accounting for length, 92.4% of RPFs mapped to annotated coding sequences, 6.9% were located in 5' leader sequences (also known as 5' untranslated regions (UTRs)), and only 0.7% were in 3' UTRs. By contrast, mRNA reads showed a more equal distribution (Supplementary Fig. 1b). The results were highly reproducible between independent biological repeats (Supplementary Fig. 2).

We used the densities of mRNA fragments and RPFs to estimate mRNA levels and protein synthesis rates, respectively. Both features correlated well with each other over the meiotic time course (Supplementary Table 2, average R = 0.79). Total translation rates (measured as RPF densities) are expected to be better predictors of protein abundance than mRNA levels. Consistently with this, protein levels estimated from an MS study of vegetative cells showed higher correlation with RPF density (Fig. 1a, R = 0.82) than with RNA levels (Fig. 1b, R = 0.68). Direct comparison of RPFs and RNA levels allowed us to estimate gene-specific translation efficiencies (TEs), which we calculated by dividing the normalized number of RPFs (in reads per kilobase per million reads (RPKM)) by that of mRNA fragments across coding sequences. In vegetative cells, TEs varied over a range of over 100-fold and did not correlate with mRNA half-lives (R = −0.12) or mRNA levels (average R over the time course = −0.08, Supplementary Table 2). As expected, TEs displayed a positive correlation with the mean number of bound ribosomes determined in a microarray-based polysome-profiling study (R = 0.43) (ref. 24). Poor TE in vegetative cells was associated with weakly expressed genes, including all major groups of meiotic genes. Moreover, targets of the nonsense-mediated decay (NMD) pathway were poorly translated. This phenomenon has been previously observed in S. cerevisiae, although it is unclear whether the NMD pathway directly represses translation or whether these mRNAs contain features that make their translation inefficient.

The translational program of meiosis

We used hierarchical clustering to investigate the changes in RPF levels across the meiotic time course. Translation patterns were dynamic and generally correlated with changes in RNA expression (Supplementary Fig. 3a). Changes in RNA levels were similar to those reported in a microarray study, thus providing independent validation of our data (Supplementary Fig. 3a). TEs varied over a wide range during meiosis, as in the vegetative state, and were dynamically regulated. 25.8% of all coding genes showed changes of five-fold or more in TE, and 6.7% showed changes of more than ten-fold. We looked for patterns in TE regulation by clustering the 318 genes that showed the strongest variation (Fig. 1c). Several groups of clustered genes were enriched in coexpressed genes. For example, a subset of late genes (cluster 1) was poorly translated until 7 h into meiosis, when its TE increased (Fig. 1d). This group was enriched in late-meiotic genes regulated by the Atf21 and Atf31 transcription factors. Similarly, a group of middle genes was translated efficiently only at 5 h (cluster 2), and a cluster enriched in genes expressed in response to nitrogen starvation (cluster 4) showed enhanced TE after nitrogen removal (Fig. 1d). In all three cases the peak of TE coincided with that of mRNA levels. Increases in TE coupled to those in RNA levels (potentiation) have been observed previously during responses to stress. A cluster that showed a strong drop in TE at 5 h despite high mRNA levels (cluster 3) was enriched in targets of the Meu5 RNA-binding protein (Fig. 1d). Meu5 binds to a subset of the middle genes (so-called ‘late decay’) and stabilizes their transcripts. Although the RNA levels of many of these genes peaked at 5 h, the accumulation of RPFs was delayed compared to that of mRNA fragments. This suggested that...
translation of some Meu5-target genes was repressed to delay the production of the corresponding protein. We checked this hypothesis by looking at published protein levels during a meiotic time course for Meu5 targets that did or did not show the decrease in TE. In all three cases, the peak of protein expression corresponded to that of RPF accumulation. Finally, we noticed that targets of the Mmi1 protein, an RNA-binding protein that promotes the degradation of a group of meiotic RNAs in vegetative cells, showed strong changes in TE, with high efficiency correlated with the expression of their mRNAs and a strong repression in vegetative cells, when expression of these genes is toxic for the cell (Fig. 1d). These results suggest that the regulation of TE and the control of mRNA levels are coordinated during the meiotic process. However, this relationship appears to be complex: in some cases it caused increased translation at peaks of RNA levels (potentiation), whereas in others (some Meu5 targets) it led to delays in protein production with respect to RNA accumulation.

Systematic identification of new translated regions

In a 28-nt-long RPF, nucleotide 13 will typically correspond to the first nucleotide of the codon located at the peptidyl (P) site (Fig. 2d)\(^{11}\). Consistently with this, we found that nucleotide 13 mapped to the first nucleotide within a codon in 75% of the 28-nt RPFs, whereas mRNA fragments were equally distributed across all three nucleotides (Supplementary Table 3). When the data were aggregated for all annotated coding sequences, this effect led to a 3-nt periodic pattern, which we observed at all positions (Fig. 2b). As previously noted, there was an accumulation of RPFs at the initiation codon (Fig. 2b) and at the last codon\(^{11,12}\). By contrast, mRNA fragments did not show a periodic behavior (Fig. 2b). As long as the RPF coverage is high enough, this feature allows the detection of any translated genomic region as well as the identification of its reading frame. Although translated sequences can be identified simply by the accumulation of RPFs, the use of periodicity allows the distinction between translated regions and contaminants in the RPF sample. We therefore used read periodicity to define translated regions in parts of the genome with no annotated features (’intergenic regions’), in genes annotated as ncRNAs and in 5’ leader sequences. We note that the experimental approach and subsequent analyses are strand specific, so that regions of the genome containing a feature on one strand were still analyzed on the other strand. For all genomic regions, we followed the same experimental strategy. First, we defined all possible ORFs starting with AUG as well as rarer initiation codons. Second, we screened each ORF for the presence of periodic signals from RPFs. Third, we estimated a false discovery rate (FDR) on the basis of a randomization test and used it to fine-tune the threshold values used to designate an ORF as translated.

To validate this approach, we calculated periodicity scores, which measure the fraction of codons translated in each reading frame, for all annotated S. pombe coding sequences (using both RPF and mRNA data). The periodicity score discriminated clearly between both data sets (Supplementary Fig. 4a) and confirmed the translation in the predicted frame of 4,923 out of 5,102 annotated high-confidence coding sequences (96.5%) (ref. 31). We manually examined genes that displayed periodicity in unexpected frames. This led to the discovery of several misannotated genes (Supplementary Table 4a), in which translation took place in a frame different from the annotated one or in which exon annotation was incorrect (Supplementary Fig. 4b–d). We also found a clear case of alternative intron retention, which is very rare in S. pombe (Supplementary Fig. 4d). A total of 71 coding genes in S. pombe are annotated as dubious\(^{31}\), thus indicating that the evidence for their existence is poor. Inspection of their translation profiles revealed that only 11 appeared to be clearly translated as predicted (15%), and three were translated in frames different from those annotated. In addition, in 25 cases (35%) the mRNA was well expressed, but we observed no translation, suggesting that these genes may not
be translated (Supplementary Table 4b). In the remaining genes, expression levels were not sufficient to allow the evaluation of translation. These results validate, refine and improve the annotation of the fission-yeast genome.

This analysis also revealed instances of overlapping coding sequences in the same strand (dualy decoded regions). For example, SPAC3C7.15c was translated from a long mRNA in the predicted frame during early meiosis. In late meiosis, a short transcript appeared that was completely enclosed within the long form of the mRNA but was translated in a different frame (Fig. 2c,d). We epitope tagged the predicted short form of the protein and observed a meiosis-specific polypeptide of the expected size, confirming that translation of the short ORF leads to the production of a stable polypeptide (Fig. 2e). The sensitivity of the approach was also exemplified by the fact that it could detect changes in translation caused by single-nucleotide polymorphisms in our strains compared to the reference strain. We detected translation of the C-terminal part of the Nup184 protein, which was not predicted to be translated in S. pombe, owing to an in-frame stop codon created by a single-nucleotide deletion.32 Translation of this region in the strains we used was explained by the presence of a single-nucleotide insertion that reverted the effect of the deletion in the reference sequence (Supplementary Fig. 5a,b).

We then scanned intergenic regions systematically and identified 715 translated ORFs with an estimated FDR of 10.3% (Supplementary Table 5). Most of these regions were short (Supplementary Fig. 5c) and had a tendency to use AUG as initiation codon (90.6%, Supplementary Table 6). We manually inspected the 46 translated ORFs of 45 codons or more. Examination of the corresponding mRNA data suggested that 39 of them (85%) were transcribed as independent units (Fig. 2f–i). Of the remaining six, two appeared to be extensions to annotated 5′ leaders, three were present in the 3′ UTRs of highly translated genes (possibly reflecting leaky termination of translation), and one was located downstream of the nup184 gene (discussed above). 15 translated ORFs did not overlap with annotated coding sequences (in either strand). Their sequences were generally not conserved, but one of them displayed homology to the N-terminal part of a protein present in multiple copies in Schizosaccharomyces cronyphilus and Schizosaccharomyces octoporum. Surprisingly, we found 14 cases in which the newly discovered translated ORF was antisense to an annotated coding sequence (completely overlapping in ten cases and overlapping by more than 80% in the other four) (Fig. 3a,b and Supplementary Fig. 6a,b). In addition, we found a similar situation (with complete overlap) in five annotated antisense ncRNAs. These exonic nested antisense ncRNAs (eNAGs) are extremely rare in eukaryotic cells, with a single case described in S. cerevisiae (the NAG1 gene)33. Simultaneous coding on both strands imposes strict constraints on the evolution of both proteins. The nature of these limitations depends on the relative frame of the sense and antisense coding sequences, which can adopt three different arrangements (Fig. 3c). We examined this configuration in all S. pombe eNAGs that were completely nested within the major coding sequence and found that in 11 cases arrangement 3 was preferred (Fig. 3c). In the case of S. cerevisiae NAG1, configuration 3 was also used. This organization causes very specific dependencies between the sequences encoded in both strands, because nucleotides 1 and 2 of every codon on both strands are encoded by the same DNA sequence. Because nucleotides 1 and 2 possess the most information content, the encoding of a particular amino acid on one strand can determine the nature of the amino acid on the antisense strand. For example, in arrangement 3, the presence of a CCN codon (proline) necessarily implies a GGN codon (glycine) in the antisense gene (and vice versa). Our results suggest that a specific arrangement of reading frames is evolutionarily favored when the DNA of both strands is coding.

Translation of annotated ncRNAs
The S. pombe genome contains 1,571 annotated long ncRNAs9,10,31. Visual inspection of our data revealed the presence of numerous meiosis-specific new genes. In vegetative cells and during most of meiosis, reads mapped to ncRNAs accounted for less than 3.5% of all reads. However, in midmeiosis (3 and 5 h) this number increased to 11%, suggesting that ncRNA function may be especially important during cellular differentiation. Recent work in several eukaryotes has revealed that ncRNAs are often present in ribosomal fractions, although it is unclear to what extent this association represents active translation11,14–16. We used triptet periodicity to address this question. We identified 499 translated regions in 375 genes (FDR 7.5%, Supplementary Table 7). These regions had a tendency to start with the canonical AUG (96.8%, Supplementary Table 6). Their median length was 21 codons, and 37 were longer than 45 codons (Fig. 4a). 28% of translated ncRNAs contained more than one translated ORF (Fig. 4b). Some of these translated regions were reminiscent of upstream ORFs, with a longer coding sequence preceded by several short uORFs.

Figure 3 Identification of translated nested antisense genes. (a) Nested antisense gene in SPAPYUG7.05. The heat maps show mRNA (top) and RPF levels (bottom) across the meiotic time course, starting with vegetative cells and progressing downward. The brightness of the red color is proportional to the number of reads. The scheme shows the position of UTRs (arrows) and ORFs (boxes). Data for the sense and antisense strands are displayed separately. LS, annotated ORF; AS, newly discovered antisense transcript. The arrows indicate the direction of transcription. Star on the heat map and dotted line on the scheme indicate the presence of a 5′ extension that appears at the 5-h time point and whose translation correlates inversely with that of the downstream ORF. (b) Triplet periodicity for the sense and antisense ORFs of SPAPYUG7.05. The graphs show the fraction of codons in which the majority of reads map to nucleotides 1, 2 or 3 for the S and L ORFs. (c) Possible relative configurations of reading frames in exonic nested antisense genes. For each configuration, the top strand represents translation in the 5′-to-3′ direction, and the bottom one represents the 3′-to-5′ direction. The numbers represent the position of nucleotide 1, 2 and 3 for each codon.
ORFs in ncRNAs have the potential to produce short polypeptides. To investigate whether this was the case, we epitope tagged two predicted peptides from the prl3 and prl46 ncRNA genes. The prl3 gene was well expressed in vegetative cells, whereas the prl46 gene was meiosis specific. Both contained ORFs that appeared to be highly translated (Fig. 4c–f). Tagging of the corresponding peptides allowed the detection of proteins of the predicted molecular weight and confirmed that both genes encoded expressed polypeptides (Fig. 4g,h). The majority of the translated sequences did not have homologs in other organisms, although there were some exceptions: The Prl46 protein and those encoded by SPNCRNA.557 and SPNCRNA.1597 were conserved in other *Schizosaccharomyces* species.

These results show that 24% of annotated ncRNAs are translated to produce short peptides. However, the majority of ncRNAs associate with ribosomes to translate ORFs of very few codons, thus indicating that the distinction between coding RNAs and ncRNAs is not clear-cut. The functional importance of these observations is unclear, although several non–mutually exclusive explanations are possible. First, short polypeptides encoded by ncRNAs may have biological activity. For example, peptides of 11 to 32 amino acids from the *Drosophila* tarsal-less (tal) gene regulate embryonic development, peptides shorter than 30 amino acids modulate cardiac function in *Drosophila* and the *S. pombe* Mat-Mi protein, consisting of only 42 amino acids, is a key regulator of meiosis that functions as a transcription factor. Second, translation could be used to target ncRNAs to polysomes, where they might be degraded (through NMD) or perform functions in translational control. Finally, translation of short ORFs in ncRNAs may reflect pervasive translation in which capped and polyadenylated sequences that reach the cytoplasm would be translated to some degree, even if their translation were not functionally important.

**Translation of 5′ leader sequences**

6.9% of all RPF reads mapped to 5′ leader sequences. Translated ORFs in 5′ leaders (uORFs) have the potential to regulate translation, although their effect can be neutral, positive or negative. Although translation of a uORF by a ribosome may downregulate translation by preventing it from reaching a downstream ORF, ribosome small subunits have the potential to perform scanning after termination and may recognize downstream initiation codons (reinitiation). In addition, uORFs can encode short peptides that are stably expressed. We examined all predicted ORFs in annotated 5′ leaders for periodic footprint patterns and identified 1,735 translated uORFs in 1,272 genes, with an estimated FDR of 10.0% (Supplementary Table 8). 26% of the genes contained more than a single uORF, and 7% had three or more (Fig. 5a). The latter group was enriched in genes encoding transcription factors, periodically expressed genes, middle meiotic genes and genes induced in response to nitrogen starvation. Interestingly, RNAs encoding four key regulators of meiosis contained four or more uORFs: *pat1* and *mei2* (which encode a kinase and an RNA-binding protein, respectively, that control entry into meiosis) and *atf21* (which encodes a transcription factor responsible for the induction of late-meiotic genes). uORFs had a mean length of 13.1 codons and a median of 10 codons (Fig. 5b). Some were extremely short, with 6.8% containing a single AUG codon (Fig. 5c). 5′ leader sequences, compared to other genomic regions, contained fewer ORFs starting with AUG (Supplementary Table 6), and translated uORFs used AUG as a start codon less frequently (Supplementary Table 6).

We also identified 175 uORFs that overlapped partially with the main ORF (in a different reading frame) (Supplementary Table 9). These may be particularly important as repressor elements, because they would not allow reinitiation of translation in the frame of the main ORF. Among them, there were three cases in which a strongly translated uORF overlapped substantially with a downstream ORF that was actively but poorly translated (Fig. 5d and Supplementary Fig. 6c,d). In the most extreme case (*SPCCI235.01*) the uORF was 347 codons long, starting 68 nt upstream of the ORF (Supplementary Fig. 6c). Interestingly, the amino acid sequence encoded at the 5′ end of the uORF (that does not overlap with the ORF) was conserved in other *Schizosaccharomyces* species, suggesting that both reading frames may produce functional proteins. *SPAC11D3.13* (Fig. 5d) and *SPACG69.05* (Supplementary Fig. 6d) presented a similar structure, although the length of the uORFs was shorter. *SPAC11D3.13* had an additional upstream uORF, in an arrangement reminiscent of the regulatory uORFs of the mammalian *ATF4* and *ATF5* mRNAs, which are translated only in stress situations (Fig. 5d).
In \textit{S. cerevisiae}, there is an increase in the use of uORFs during meiosis\textsuperscript{13}. We examined whether a similar phenomenon takes place in \textit{S. pombe} by comparing the ratio between the TEs of every uORF and its corresponding ORF for every time point (Supplementary Fig. 3b). Although we observed increases toward the end of meiosis, the changes were generally small, thus indicating that in \textit{S. pombe} neither nitrogen starvation nor the meiotic program caused a large rise in the use of uORFs.

To investigate global effects of uORFs on translation, we compared the TE of every uORF to that of its downstream ORF. We then examined whether changes in TE of uORFs during meiosis correlated with variations in TE of downstream coding regions. We calculated the correlation between the TE of every uORF and its downstream ORFs across all time points (examples in Supplementary Fig. 3). The majority of genes showed strong positive correlations (Fig. 5e), although a very small subset displayed a negative relationship. In SPBC1773.04, a meiosis-specific 5’ extension to the mRNA contained two uORFs whose translation was negatively correlated with that of the main ORF, suggesting that the uORFs competed for translation with and downregulated translation of the downstream ORF (Fig. 5f). Another example is displayed in Figure 3a, in this case concerning an antisense transcript. Although temporal regulation of TE through changes in the use of uORFs is common in \textit{S. cerevisiae} meiosis\textsuperscript{13}, it appears to be rare in \textit{S. pombe}.

**DISCUSSION**

Our results reveal pervasive translation of the \textit{S. pombe} transcriptome, including dually decoded regions and exonic nested antisense genes as well as frequent translation of annotated ncRNAs. Overall, we found 917 translated ORFs of 20 codons or longer (in 5’ leader sequences, annotated ncRNAs and new translated regions), results suggesting the existence of a large repertoire of small peptides with potential biological functions. We have experimentally validated the expression of some of these peptides, demonstrating that their translation results in real changes to the proteome. In addition, we have observed substantial use of TE regulation during meiosis, including homodirectional changes in RNA levels and TE (potentiation) and the use of TE changes to delay protein accumulation. The existence of a previous data set from \textit{S. cerevisiae} allowed us to address whether properties of the translational programs of cellular differentiation processes are general or organism specific. Both \textit{S. pombe} and \textit{S. cerevisiae} use extensive translational control, as demonstrated by the widespread and dynamic changes in TEs. In \textit{S. pombe}, in contrast to \textit{S. cerevisiae}, we did not detect a switch to the use of uORFs and unconventional initiation mechanisms, results suggesting that this is not a general feature of meiosis or cellular differentiation processes.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Accession codes.** All sequencing raw data have been deposited in ArrayExpress under the following accession numbers: E-MTAB-2176 (vegetative haploid cells), E-MTAB-2179 (\textit{pat1} meiosis, replicate 1), E-MTAB-2265 (\textit{pat1} meiosis, replicate 2) and E-MTAB-2470 (total RNA, vegetative cells).

\textit{Note:} Any Supplementary Information and Source Data files are available in the online version of the paper.
ACKNOWLEDGMENTS

We thank J. Bähler and R. Jackson for comments on the manuscript. This work was funded by Biotechnology and Biological Sciences Research Council (UK) research grant BB/F011531/1 (J.M.).

AUTHOR CONTRIBUTIONS

C.D.S.D. and J.M. conceived the study and designed the experiments. C.D.S.D. performed all the experiments, and J.M. did the bioinformatics analyses. C.D.S.D. and J.M. wrote the manuscript and have critically read and approved of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Yamamoto, M., Imai, I. & Watanabe, Y. In The Molecular and Cellular Biology of the Yeast Saccharomyces: Life Cycle and Cell Biology (eds. Pringle, J.R., Broach, J.R. & Jones, E.W.) 1035–1106 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA, 1997).

2. Mata, J., Lyne, R., Burns, G. & Bähler, J. The transcriptional program of meiosis and sporulation in fission yeast. Nat. Genet. 32, 143–147 (2002).

3. Horie, S. et al. The Schizosaccharomyces pombe mei4+ gene encodes a meiosis-specific transcription factor containing a forhead DNA-binding domain. Mol. Cell. Biol. 18, 2118–2129 (1998).

4. Mata, J., Wibrey, A. & Bähler, J. Transcriptional regulatory network for sexual differentiation in fission yeast. Genome Biol. 8, R217 (2007).

5. Sugimoto, A., Iino, Y., Maeda, T., Watanabe, Y. & Yamamoto, M. Schizosaccharomyces pombe ste11+ encodes a transcription factor with an HMG motif that is a critical regulator of sexual development. Genes Dev. 5, 1990–1999 (1991).

6. Amorim, M.J., Cotobal, C., Duncan, C. & Mata, J. Global coordination of transcriptional control and mRNA decay during cellular differentiation. Mol. Syst. Biol. 6, 380 (2010).

7. Harigaya, Y. et al. Selective elimination of messenger RNA prevents an incidence of untimely meiosis. Nature 442, 45–50 (2006).

8. Bitton, D.A. et al. Programmed fluctuations in sense/antisense transcript ratios drive sexual differentiation in S. pombe. Mol. Syst. Biol. 7, 559 (2011).

9. Rhind, N. et al. Comparative functional genomics of the fission yeasts. Science 322, 930–936 (2011).

10. Wilhelm, B.T. et al. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453, 1239–1243 (2008).

11. Ingolia, N.T., Ghemmaghami, S., Newman, J.R. & Weissman, J.S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324, 218–223 (2009).

12. Michel, A.M. & Baranov, P.V. Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. Wiley Interdiscip. Rev. RNA 4, 473–490 (2013).

13. Brar, G.A. et al. High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science 335, 552–557 (2012).

14. Ingolia, N.T., Lareau, L.F. & Weissman, J.S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian protein synthesis. Cell 147, 789–802 (2011).

15. Chew, G.L. et al. Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. Development 140, 2828–2834 (2013).

16. Gutman, M., Russell, P., Ingolia, N.T., Weissman, J.S. & Lander, E.S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell 154, 240–251 (2013).

17. Michel, A.M. et al. Observation of dually decoded regions of the human genome using ribosome profiling data. Genome Res. 22, 2219–2229 (2012).

18. Bazzini, A.A. et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J. 33, 981–993 (2014).

19. Iino, Y. & Yamamoto, M. Mutants of Schizosaccharomyces pombe which sporulate in the haploid state. Mol. Gen. Genet. 198, 416–421 (1985).

20. Nurse, P. Mutants of the fission yeast Schizosaccharomyces pombe which alter the shift between cell proliferation and sporulation. Mol. Gen. Genet. 198, 497–502 (1985).

21. König, J. et al. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat. Struct. Mol. Biol. 17, 909–915 (2010).

22. Mata, J. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. RNA Biol. 10, 1407–1414 (2013).

23. Marguerat, S. et al. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. Cell 151, 671–683 (2012).

24. Lackner, D.H. et al. A network of multiple regulatory layers shapes gene expression in fission yeast. Mol. Cell 26, 145–155 (2007).

25. Matía-González, A.M., Hasan, A., Moe, G.H., Mata, J. & Rodríguez-Gabriel, M.A. Functional characterization of Upf1 targets in Schizosaccharomyces pombe. RNA Biol. 10, 1057–1065 (2013).

26. Isken, O. et al. Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. Cell 133, 314–327 (2008).

27. Zhang, Z. et al. Nonsense-mediated decay targets have multiple sequence-related features that can inhibit translation. Mol. Syst. Biol. 6, 442 (2010).

28. Lackner, D.H., Schmidt, M.W., Wu, S., Wolf, D.A. & Bahler, J. Regulation of transcriptome, translation, and proteome in response to environmental stress in fission yeast. Genome Biol. 13, R25 (2012).

29. MacKay, V.L. et al. Gene expression analyzed by high-resolution state array analysis and quantitative proteomics: response of yeast to mating pheromone. Mol. Cell. Proteomics 3, 478–489 (2004).

30. Preiss, T., Baron-Benhamou, J., Ansorge, W. & Hentze, M.W. Homodirectional changes in transcriptome composition and mRNA translation induced by rapamycin and heat shock. Nat. Struct. Biol. 10, 1039–1047 (2003).

31. Wood, V. et al. PomBase: a comprehensive online resource for fission yeast. Nucleic Acids Res. 40, D695–D699 (2012).

32. Wood, V. et al. The genome sequence of Schizosaccharomyces pombe. Nature 415, 871–880 (2002).

33. Ma, J., Doob, C.J., Kysan, D.J. & Kumar, A. Unconventional genomic architecture in the budding yeast Saccharomyces cerevisiae masks the nested antisense gene NAG1. Eukaryot. Cell 7, 1289–1298 (2008).

34. Galindo, M.I., Pueyo, J.I., Fouix, S., Bishop, S.A. & Couso, J.P. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. PLoS Biol. 5, e106 (2007).

35. Kondo, T. et al. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. Nat. Cell Biol. 9, 660–665 (2007).

36. Magry, E.G. et al. Conservation regulated of cardiac calcium uptake by peptides encoded in small open reading frames. Science 341, 1116–1120 (2013).

37. Kelly, M., Burke, J., Smith, M., Klar, A. & Beach, D. Four mating-type genes control sexual differentiation in the fission yeast. EMBO J. 7, 1537–1547 (1988).

38. Hoo, H.M., Neafsey, D.E., Galagan, J. & Sachs, M.S. Evolutionary roles of upstream open reading frames in mediating gene regulation in fungi. Annu. Rev. Microbiol. 63, 385–409 (2009).

39. Jackson, R.J., Hellen, C.U. & Pestova, T.V. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. 11, 113–127 (2010).

40. Slavov, S.A. et al. Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat. Chem. Biol. 9, 59–64 (2013).

41. Shimoda, C. et al. Cloning and analysis of transcription of the mei2 gene responsible for initiation of meiosis in the fission yeast Schizosaccharomyces pombe. J. Bacteriol. 169, 93–96 (1987).
ONLINE METHODS

General methods. Standard methods and media were used for fission-yeast growth. Wild-type and pat1-driven meiosis were induced as described2. Vegetative cells were grown in rich medium at 32 °C. Proteins were tagged with TAP: Tagged proteins were detected by western blot using peroxidase-antiperoxidase–soluble complexes (Sigma P1291) diluted 1:10,000 and α-tubulin with mouse monoclonal antibodies (Sigma, clone B-5-1-2) diluted 1:20,000. For protein detection of Prt3 and Prt46, 50 µl of cells (3 × 10^8–8 × 10^10 cells/ml) were treated with 1 mM PMSF for 5 min before centrifugation at 4 °C and freezing. Cells were later thawed, washed with cold lysis buffer (20 mM Tris–HCl, pH 8.0, 140 mM KCl, 1.8 mM MgCl2 and 0.1% NP-40), resuspended in 200 µl of lysis buffer containing 1 mM PMSF and 1:100 protease-inhibitor cocktail (Sigma P8340), and lysed with a bead beater (Fastprep, MP Biomedicals) at level 6 for 1 s. Extracts were cleared by centrifugation at 7,600 × g for 3 min and used for western blotting. The short peptide overlapping SPAC3C.15c was not visible under these conditions. In order to detect it, cells were washed with cold RIPA buffer (10 mM Tris–HCl, pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, and 0.1% SDS), resuspended in 200 µl of RIPA buffer, boiled for 5 min and frozen. Samples were then processed as described above, except that RIPA buffer was used, and the extracts were not precleared by centrifugation. Original images of gels used in this study can be found in Supplementary Figure 7.

Ribosome profiling, library preparation and sequencing. 3 × 10^6 exponentially growing wild-type haploid cells, or between 3 × 10^6 and 12 × 10^6 meiotic cells, were incubated for 5 min with 100 µg/ml cycloheximide, pelleted at 4 °C and frozen in liquid N2. Each culture was split for footprint isolation and mRNA fragmentation. We generally followed a published protocol11 with the modifications stated below. Cells were resuspended in 100 µl of lysis buffer (20 mM Tris–HCl, pH 8.0, 140 mM KCl, 5 mM MgCl2, and 1% Triton X-100) with 1 g of chilled glass beads (Biospec) and lysed with a Fastprep 24 bead beater at level 6 for 13 s. The extract was diluted with 400 µl of lysis buffer and cleared by centrifugation in two steps at 4 °C at 16,000 g (5 min followed by 15 min). For footprint isolation, 600 A260 units of wild-type vegetative-cell extract was digested with 750 U of RNase A (Life Technologies) for 30 min, or 800 A260 units of pat1 diploid extract was treated for 10 min with 1,500 U RNase I. Reactions were quenched with 600 U of SUPERaseI (Life Technologies). Digested extracts in 500 µl were loaded onto a 14-ml linear 10–50% (w/v) sucrose gradient prepared with a Gradient Master–min-intron-length 29–max-intron-length 819–zpacker 0–splice-mismatches 0. RNase I (Life Technologies) for 30 min, or 800 A260 units of pat1 diploid extract was treated for 10 min with 1,500 U RNase I. Reactions were quenched with 600 U of SUPERaseI (Life Technologies). Digested extracts in 500 µl were loaded onto a 14-ml linear 10–50% (w/v) sucrose gradient prepared with a Gradient Master (Biocep) and separated by centrifugation for 160 min at 35,000 r.p.m. in a SW 40Ti rotor (Beckman). The gradients were then fractionated by upward displacement with 55% (w/v) sucrose, and fractions containing monosomes were selected for further processing. RNAs were then purified by phenol extraction, passed through a YM-100 column (Millipore), and run on 15% TBE-urea gels (Life Technologies). Fragments of around 28 nt were extracted from the gel. For the preparation of mRNA fragments, total RNA was purified by phenol extraction, and polyadenylated RNA was purified from 150 µg of total RNA (vegetative cells) or 300–450 µg (pat1 cells) with oligo(dT)25 magnetic beads (Life Technologies) according to the manufacturer’s instructions. Purified mRNA was fragmented by mixing of 20 µl of mRNA with 20 µl of 2x alkaline fragmentation buffer (2 mM EDTA and 100 mM Na2CO3). This was followed by incubation for 15 min at 95 °C. Samples were run on 15% TBE-urea gels (Life Technologies), and fragments of around 28 nt were extracted from the gel. From this point, both mRNA- and ribosomal-footprint samples were processed identically. RNA samples were purified with Purelink RNA microcolumns (Life Technologies) as described by the manufacturer, except that the samples were initially passed through the column in the presence of 70% ethanol (to favor binding of small RNAs). Samples were then treated with polynucleotide kinase (PNK, Fermentas) as described11. Nonredundant sequences available from GenDB (http://old.genedb.org/), now PomBase (http://www.pombase.org/), were written in Perl (http://www.perl.org/), and all downstream statistical analysis was performed with custom scripts written in Perl (http://www.perl.org/), and all downstream statistical analysis was performed with custom scripts written in Perl (http://www.perl.org/), and all downstream statistical analysis was performed with custom scripts written in Perl (http://www.perl.org/), and all downstream statistical analysis was performed with custom scripts written in Perl (http://www.perl.org/). For all analyses, S. pombe annotations and sequences available from GeneDB (http://old.genedb.org/), now PomBase (http://www.pombase.org/), or 9 May 2011 were used31.

The RT primers (Supplementary Table 10) included a 4-nt barcode that allows multiplexing, which was used to allocate reads from different samples to separate files. The RT primers also included a 5-nt random sequence that served as a unique molecular identifier (UMI). Reads that contain the same UMI followed by an identical sequence are highly likely to have arisen from the same RNA molecule, and only one of them is retained. This step creates a nonredundant data set, thus avoiding sampling biases and PCR amplification artifacts12,23. Nonredundant reads were then processed to remove A residues at their 3’ ends. Reads were mapped to the S. pombe rDNA genome32 with TopHat2 (ref. 42) and the following parameters:–min-intron-length 29–max-intron-length 819–splice-matches 0–splice-mismatches 0–max-multihits 1. Unmapped reads were recovered and aligned to the full S. pombe genome32 with TopHat2 and the following settings:–min-intron-length 29–max-intron-length 819–splice-matches 0–splice-mismatches 0. A GFF file containing annotation of the S. pombe genome32 was provided as a source of exon-exon junction data for TopHat. Aligned data were visualized with the Integrated Genome Viewer43. RPF libraries are expected to contain a higher fraction of contaminating rRNA than mRNA libraries, because the latter are oligo(dT) selected. Consistently with this, 81.7% of reads from RPF-derived and 26.5% from mRNA-derived libraries mapped to the rDNA genome (median from all experiments, Supplementary Table 1). UMIs were used to calculate the fraction of unique reads for those reads mapping to rRNA and for the remaining set. rRNA reads in the RPF samples typically showed low complexity (median unique 10.1%, Supplementary Table 1). This may be because the majority of these reads originate from a small pool of sequences (presumably owing to the sequence preferences of Rnase I). Because the UMI length is 5 nt, there are 1,024 different UMIs. If the number of fragments derived from the same sequence largely exceeds this figure, the likelihood that independent fragments with the same sequence share the same UMI increases and artificially decreases the observed number of unique sequences. This is unlikely to be a problem for RNA reads in the poly(A)-purified sample (median unique 73%, Supplementary Table 1) or for non-rRNA reads (median unique 83.4%, Supplementary Table 1), which are evenly distributed along the RNAs.

For clustering of RPFs (Supplementary Fig. 3a), the 1,719 genes showing the strongest changes in gene expression were selected. All the data were normalized to expression levels in vegetative cells of the corresponding time course, and RPF values were used for clustering. For clustering of TE5s (Fig. 1c), the 418 genes that showed the largest variations in TE across the time course were chosen. Clustering was performed with Cluster 3.0 (refs. 44,45), filtering out data with more than 20% values missing, log-transforming the data, determining Pearson correlation and creating an average-linked tree. Clusters were visualized with Treeview46.

To calculate overall periodicity and to calibrate reads, we counted the number of reads in which position 13 of a read matched the first, second, or third position of every codon for all annotated coding sequences of the S. pombe genome (Supplementary Fig. 4a). The choice of nucleotide 13 is arbitrary; similar results would be obtained with a different nucleotide. The advantage of nucleotide 13 is that for a ribosome starting translation at the initiation codon, it corresponds to the position of the AUG on the mRNA. This was done for mRNA and RPFs and analyzed separately for reads of lengths between 25 and 32 nt (Supplementary Fig. 8b). No bias was observed in any position within a codon for the mRNA fragments. By contrast, enrichments in position 1 were observed for RPFs of 28, 29 and 30 nt (Supplementary Fig. 8b). The fact that fragments longer than 28 nt
show periodicity is unexpected and suggests that some feature of the ribosome allows more-precise cutting of RNase I at the 5′ than at the 3′ of the protected fragments. Reads between 28 and 30 nt were selected for the discovery of translated features (such as uORFs and ncRNAs), and all reads were used to calculate translational efficiencies.

For the identification of new translated regions, annotated coding sequences, noncoding RNAs, uORFs and intergenic regions were analyzed separately. Open reading frames were defined as follows. First, selected regions were scanned in each frame until an AUG was encountered. This was defined as the start of the ORF, which was elongated until a stop codon was found. Second, the process was repeated as above for UUG, except when an ORF overlapped with a previously defined one in the same frame it was discarded. The process was then performed as above for CUG and GUG. A total of 384,032 ORFs were defined in intergenic regions, 50,574 ORFs in ncRNAs, and 31,996 in 5′ leader sequences. Of these, 49.2%, 49.2% and 27.2% started with an AUG codon in intergenic regions, ncRNAs, and 5′ leaders, respectively (Supplementary Table 6). This indicates that AUG ORFs are depleted in 5′ leader sequences compared to other parts of the genome. The resulting ORFs were analyzed for read periodicity by quantification of the fraction of codons showing enrichment of reads in the first nucleotide. For every codon, the enriched nucleotide (if existing) was defined as that having at least 60% of all the reads that mapped to the codon. In this way, all codons have equal contributions regardless of the total number of reads that map to each of them, thus avoiding biases created by a small number of codons with a very high number of reads. The fraction of codons enriched in nucleotide 1 within an ORF was defined as the ORF periodicity score. An ORF was defined as translated when its periodicity score was ≥0.6. To avoid noise from weakly expressed genes, a total number of ten reads was required for a region to be considered as translated. False discovery rates (FDRs) were calculated by randomizing the position of the reads within each codon as follows: if a codon contains a1, a2, a3 reads in positions 1, 2 and 3, respectively, the numbers a1, a2, and a3 were randomly assigned to a position within a codon (to create 3! = 6 possibilities per codon). This was followed by the analysis of periodicity as described above. A P value was also calculated for each feature by assuming a binomial distribution for the fraction of codons enriched at position 1. Under this very conservative assumption, the smallest possible P value for an ORF of a single codon is 0.33. However, given that our threshold for calling a feature as translated requires at least ten reads and that more than 60% of the reads map to the first nucleotide, the probability of passing the threshold for such an ORF is lower than 0.0035 (on the basis of the less conservative assumption that reads are randomly distributed within the codon).

42. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).
43. Thorvaldsdóttir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192 (2013).
44. de Hoon, M.J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. Bioinformatics 20, 1453–1454 (2004).
45. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95, 14863–14868 (1998).
46. Saldanha, A.J. Java Treeview: extensible visualization of microarray data. Bioinformatics 20, 3246–3248 (2004).