nature structural & molecular biology



# Exploring functional relationships between components of the gene expression machinery

Todd Burckin[1,6], Roland Nagel[1,2,6], Yael Mandel-Gutfreund[1,2,5], Lily Shiue[3], Tyson A Clark[1,5], Jean-Leon Chong[4], Tien-Hsien Chang[4], Sharon Squazzo[1], Grant Hartzog[1,3] & Manuel Ares Jr[1–3]

**Eukaryotic gene expression requires the coordinated activity of many macromolecular machines including transcription factors and RNA polymerase, the spliceosome, mRNA export factors, the nuclear pore, the ribosome and decay machineries. Yeast carrying mutations in genes encoding components of these machineries were examined using microarrays to measure changes in both pre-mRNA and mRNA levels. We used these measurements as a quantitative phenotype to ask how steps in the gene expression pathway are functionally connected. A multiclass support vector machine was trained to recognize the gene expression phenotypes caused by these mutations. In several cases, unexpected phenotype assignments by the computer revealed functional roles for specific factors at multiple steps in the gene expression pathway. The ability to resolve gene expression pathway phenotypes provides insight into how the major machineries of gene expression communicate with each other.**

Our understanding of gene expression is based on a combination of genetic and biochemical methods to assign roles to specific factors in separate steps of the gene expression pathway, such as transcription or splicing. As these separate pictures have emerged, physical and genetic interactions between factors involved in different steps of the gene expression pathway have suggested that the activities of the gene expression pathway are linked[1,2]. Even so, knowing that two proteins are associated is only the first step toward understanding whether and how they might work together. So far, coupled *in vitro* systems that faithfully and efficiently carry out more than one major step in the gene expression pathway are not available, making it difficult to examine directly the mechanisms by which the integrated function of these steps is achieved. Thus the functional integration of the steps of gene expression remains to be determined.

Mutations that give rise to similar phenotypes usually represent genes required for shared function. The consequences of the loss of complex functions often produce idiosyncratic effects that geneticists can recognize as distinct phenotypes. According to this reasoning, a mutation of a general component of the transcription apparatus might produce a genome-wide gene expression phenotype similar to that produced by the loss of other general transcription components, yet different than that produced by the loss of function of a general splicing factor. Furthermore, if a component were functionally required for both transcription and splicing, then its gene expression phenotype might have characteristics of both classes. Thus it seems possible that by analyzing gene expression phenotypes in detail, factors required for more than one process might be recognized.

To explore this question, we used microarrays capable of detecting both mRNAs and pre-mRNAs to analyze the phenotype caused by mutations or perturbations affecting the basal gene expression machinery in *Saccharomyces cerevisiae*. We used these data as a complex, quantitative phenotype to ask how the different steps in the gene expression pathway are connected. This analysis included hierarchical clustering of data, as well as use of multiclass support vector machines (SVMs), a supervised learning technique in which a computer is trained to distinguish between classes of data based on prior knowledge. In several cases, the SVM assigned a mutation to an unexpected functional class, suggesting unappreciated roles for that factor in gene expression. We further tested two of these predictions, providing new evidence that transcription elongation factors may influence mRNA export from the nucleus, and that translation factor Ded1p associates with the spliceosome and influences splicing.

## RESULTS

We assessed the impact of 80 perturbations in the *S. cerevisiae* gene expression pathway on the steady-state levels of mRNAs and pre-mRNAs using splicing-sensitive microarrays[3]. Perturbations included mutations affecting components of the transcription, capping, splicing, 3′-end formation, nuclear decay, nuclear export, translation and cytoplasmic decay machineries, as well as heat shock, and treatment with the drugs rapamycin or cycloheximide. Expression measurements were normalized to a set of ~1,200 intronless genes, a strategy that both normalizes transcript levels and preserves splicing-specific information about gene expression. Array results were validated by selected quan-

[1]Department of Molecular, Cell & Developmental Biology, [2]Center for Molecular Biology of RNA and [3]Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA. [4]Department of Molecular Genetics, The Ohio State University, 484 West 12th Avenue, Columbus, Ohio 43210, USA. [5]Present addresses: Faculty of Biology, Technion, Haifa 32000, Israel (Y.M.-G.) and Affymetrix, 3380 Central Expressway, Santa Clara, California 95051, USA (T.A.C.). [6]These authors contributed equally to this work. Correspondence should be addressed to M.A. (ares@biology.ucsc.edu).

Published online 30 January 2005; doi:10.1038/nsmb891

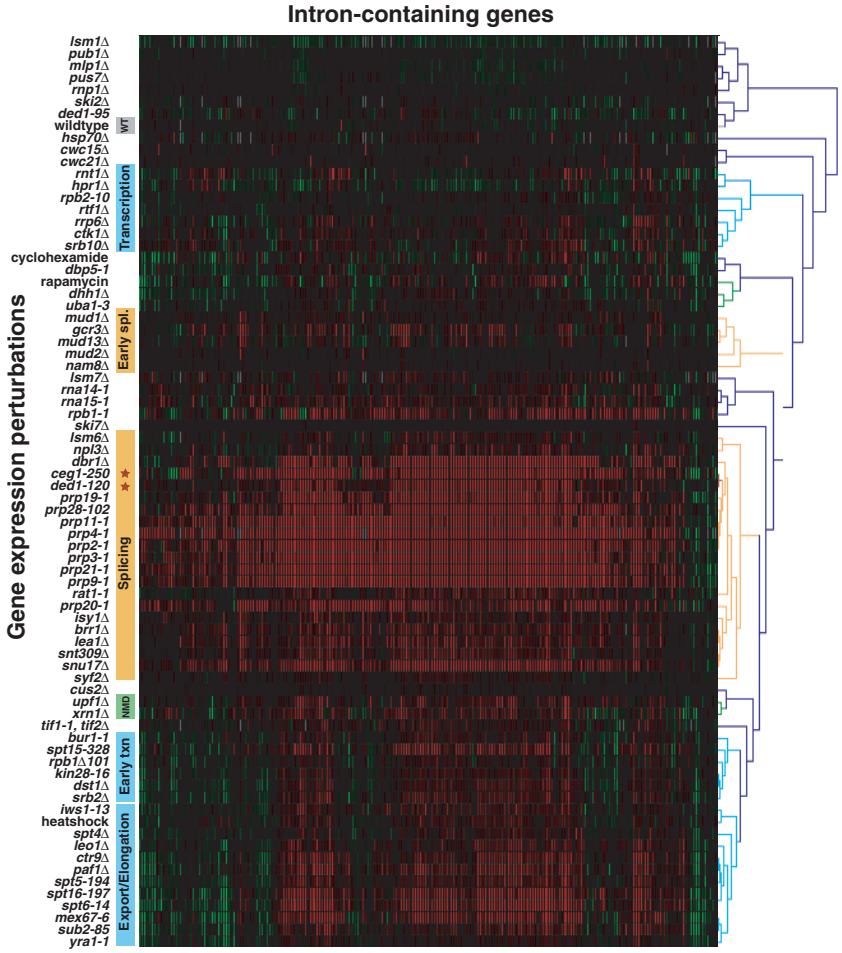


**Figure 1** Hierarchical clustering of intron accumulation indexes. Genesis (release 1.2.2)[60] was used to create a hierarchical cluster using an uncentered Pearson correlation metric and average linkage algorithm from the IA values of 76 experiments, using the data in **Supplementary Table 3** online. Horizontal columns represent cellular perturbations and the ~240 intron-containing genes are clustered in rows along the vertical axis. Scale is as follows: saturated red, 1.5-fold increase in IA value; saturated green, 1.5-fold decrease; black, no change; gray, absent data. Functionally related cellular perturbations are color-coded.

titative PCR experiments (**Supplementary Fig. 1** and **Supplementary Tables 1** and **2** online). This effort captured gene-specific changes in mRNA level as well as pre-mRNA level caused by the perturbations. Measurements of different transcripts from the same gene can thus be analyzed separately or combined to create other metrics such as the intron accumulation (IA) index, which normalizes the change in an intron-specific signal to the change in signal for the total RNA derived from the gene as reported by an exon probe[3]. We clustered the IA indexes (**Fig. 1** and **Supplementary Table 3** online) to identify relationships in the impact of these mutations on expression of the ~240 intron-containing genes of yeast.

A notable number of transcription mutants caused an increase in the IA for many intron-containing genes (**Fig. 1**). To explore this, we analyzed the temperature-sensitive *rpb1-1* mutant, which has been widely applied to study RNA decay[4–6]. An increase in the IA index for many intron-containing genes is observed after heat shift of the *rpb1-1* mutant (**Fig. 1**), consistent with a direct role for RNA polymerase II (Pol II) in splicing[7,8]. However, inspection of the log ratios for the intron features (rather than the IA index) indicates that, for *rpb1-1*, the increase in the IA is primarily due to the slower decrease in intron-containing sequences as compared with total transcripts. In contrast, when splicing is inhibited the IA index increases owing to increased intron-containing sequences (**Supplementary Fig. 2** online). Thus the increase in relative intron levels in *rpb1-1* cells is probably a consequence of more rapid decay of mRNA than intron sequences in the mutant after the temperature shift, rather than a direct role for the RNA polymerase large subunit in splicing. Although the distribution of intron sequences in the forms of spliced-out intron lariats, intermediate decay products, or unspliced pre-mRNA is uncertain and probably different in different mutants, the qualitative and quantitative aspects of these changes seem to characterize distinct transcriptional defects (see below).

## Mutants cluster largely by known function

The *prp4-1* heat shift experiment provides a strong and specific block to splicing to which other more subtle splicing phenotypes can be compared[3]. The *prp4* heat shift experiment clusters tightly with *prp2-1*, *prp3-1*, *prp9-1*, *prp11-1*, *prp19-1*, *prp21-1* (heat shift), and *prp28-102* (cold shift, 16 °C), each of which provide strong blocks to splicing (**Fig. 1**). Other splicing factor mutations that cause more subtle blocks to splicing cluster separately, possibly because these factors contribute less strongly or differentially to the splicing of subsets of introns[3]. Contained within the strong splicing cluster are two unexpected mutants, *ceg1*, encoding a catalytic subunit of the mRNA capping enzyme[9,10], and *ded1*, previously shown to have a role in translation[11]. In contrast to the *rpb1-1* mutation, the increase in IA observed for members of this cluster was primarily due to an accumulation of pre-mRNA in the mutant (data not shown; see **Supplementary Figs. 2** and **3** online).

The tight cluster most closely related to the strong splicing mutants contains components implicated in both transcription and mRNA export. Mutations in genes for two components of the TREX complex[12], *yra1-1* and *sub2-85*, as well as *mex67-6*, a mutation in a gene encoding an export factor that competes with Sub2p for binding to Yra1p[12–14], produce very similar relative intron accumulation phenotypes (**Fig. 1**). In contrast, mutation of *HPR1*, a gene for another TREX component, produced a distinct phenotype[12,15] (**Fig. 1**). Recent studies in *Drosophila melanogaster* have also found that perturbation of the fly homologs of Sub2p and Mex67p lead to similar expression profiles whereas that of fly Hpr1p is distinct[16,17]. Notably, mutations in several chromatin-specific transcription elongation factor genes[18], *spt4*Δ, *spt5-194*, *spt16-197*, *spt6-14*, and those encoding components of the Paf1 complex[19–21], *leo1*Δ, *ctr9*Δ and *paf1*Δ, also cluster with these export mutations (**Fig. 1**). Previous work has demonstrated an extensive set of physical and genetic interactions between these elongation factors[19–22]. In addition, the Paf1 and TREX complexes may share a subunit and show similar patterns of gene occupancy in chromatin immunoprecipitation experiments[23,24]. The similarity in IA indexes for these chromatin-specific transcription elongation mutants as compared with those

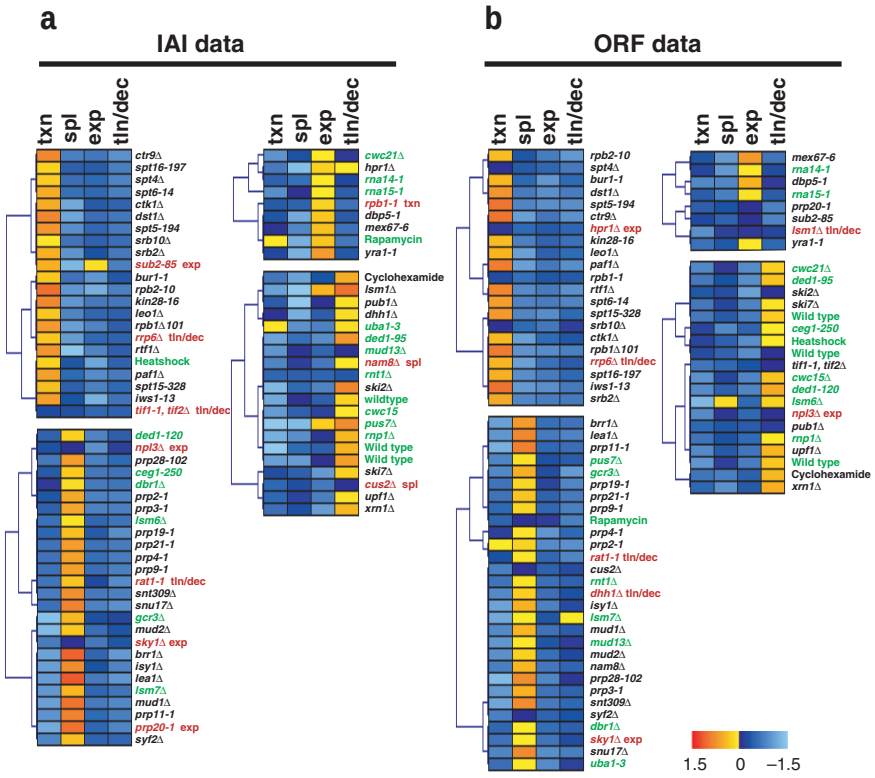


**Figure 2** Multiclass SVM analysis of gene expression data. (**a**) IA indexes (IAI) for intron-containing genes. (**b**) Expression ratios for ORFs. Training labels (transcription, txn; splicing, spl; export, exp; or translation and decay, tln/dec) were assigned based on prior knowledge of the function of the mutant gene tested in the experiment. Each experiment received a prediction score from the SVM for each of the four labels. Score values represent distance from the separating hyperplane, with positive values in yellow and negative values in shades of blue. Color-coding of the experimental sample names is as follows: black, strongest SVM prediction is consistent with training label; red, strongest SVM prediction is different from training label (shown); green, SVM prediction for experiments not included in training.



of the mRNA export mutants suggests that loss of any of these components leads to a similar gene expression phenotype, and thus that they function together (see below).

Mutations affecting a second, distinct set of transcription factors cluster separately from the TREX/Paf1/elongation factor group (**Fig. 1**). This group includes *rpb1Δ101*, a deletion of all but 10 of the 26 heptad repeats of the Pol II C-terminal domain (CTD) , and mutations in *KIN28* and *BUR1*, genes for two Pol II CTD kinases implicated in initiation and elongation, respectively[25]. Also in this group are mutations in the TBP-encoding gene *SPT15* and the *SRB2* component of the mediator complex[26], both of which are required for initiation, as well as mutations affecting transcription elongation factors TFIIS (*dst1Δ*)[26] and Iws1p/Spn1p[21,22,27]. This indicates that the transcriptional defects caused by loss of function for these proteins are distinct from other components in the transcription apparatus, and suggests that these proteins share similar major functional roles.

A third distinct cluster of transcription factors includes mutations in *SRB10* and *CTK1*, which encode CTD kinases implicated in repression of initiation (Srb10p)[25], elongation and 3′-end formation (Ctk1p)[25,28,29]. Also clustering with these two kinases are *RTF1*, a component of the Paf1 complex[19–21], and *HPR1*, a member of the TREX-associated THO elongation complex[12,15]. Notably, deletions of *RTF1* lead to mutant phenotypes and genetic interactions that are somewhat different than those associated with loss of other members of the Paf1 complex, and Hpr1p may also associate with the Paf1 complex[19,20,24,30]. Thus the gene expression phenotype represented by the IA index seems to distinguish different activities in transcription and export, identifying functionally distinct subunits within the same complex (**Fig. 1**).

**Table 1 Performance of the multiclass SVM**

Intron accumulation index

| | SVM prediction | | | |
|---|---|---|---|---|
| Input label | txn | spl | tln/dec | exp |
| txn | **18** | 0 | 0 | 1 |
| spl | 0 | **16** | 2 | 0 |
| tln/dec | 2 | 1 | **8** | 0 |
| exp | 1 | 3 | 0 | **4** |

Intronless ORF expression

| | SVM prediction | | | |
|---|---|---|---|---|
| Input label | txn | spl | tln/dec | exp |
| txn | **19** | 0 | 0 | 0 |
| spl | 0 | **18** | 0 | 0 |
| tln/dec | 0 | 2 | **8** | 1 |
| exp | 1 | 1 | 1 | **5** |

Transcription, txn; splicing, spl; export, exp; translation and decay, tln/dec. Boldface type indicates the numbers of mutants whose SVM-predicted phenotype matched their input label.

## Formal analysis of complex phenotypes

Hierarchical clustering depends on binary decisions that may obscure important similarities between mutants that share overlapping phenotypes owing to secondary functional relationships. To complement the hierarchical clustering described above (**Fig. 1**), we used SVMs, a supervised learning method that uses prior knowledge of gene function for training (**Fig. 2**). After training, an SVM can identify key features that distinguish labeled classes of data, identify samples that may have been mischaracterized by the input 'knowledge,' as well as classify and label new data. SVMs have been applied successfully to the related problems of deciding which genes are part of a coregulated set[31], and distinguishing superficially related tumor types[32,33]. Recently SVMs have been applied to classify phenotypes associated with broad categories of gene function[34]. Here we have applied this method to discriminate gene expression phenotypes in order to identify proteins that play roles in multiple processes.

Mutations thought to affect primarily a single step in the gene expression pathway were assigned to one of four different functional groups: transcription (txn), splicing (spl), mRNA export (exp), and translation and RNA decay (tln/dec), using our own evaluation of the literature, as well as the 'gene ontology'[35] (GO, http://www.geneontology.org). Because classical SVMs can discriminate between only two classes at a

**Table 2 Array analysis and genetic tests for seven genes implicated in transcription and export**

| IA index cluster | Gene | Biochemical complex | Initial label | SVM call IA | SVM call ORF | Genetic test |
|---|---|---|---|---|---|---|
| | *SUB2* | TREX/THO | exp | txn | exp | *sub2-85* weakly suppressed by *spt4-3* |
| | *MEX67* | | exp | exp | exp | *mex67-6* suppressed by *spt4-3* |
| | *YRA1* | TREX/THO | exp | exp | exp | *yra1-1* inviable with *spt4*Δ, *paf1*Δ, *spt6-140*, *spt16-197* |
| | *PAF1* | Paf1 | txn | txn | txn | *paf1*Δ inviable with *yra1-1*, *hpr1*Δ, *spt4*Δ[a] |
| | *SPT16* | FACT | txn | txn | txn | *spt16-197* inviable with *yra1-1* |
| | *SPT6* | | txn | txn | txn | *spt6-140* inviable with *yra1-1*, *hpr1*Δ |
| | *SPT4* | Spt4–Spt5 | txn | txn | txn | *spt4-3* suppresses *sub2-85*, *mex67-6*, *spt4*Δ is inviable with *yra1-1*, *hpr1*Δ[b], *paf1*Δ[a] |
| | *HPR1* | TREX/THO | exp | exp | txn | *hpr1*Δ inviable with *spt4*Δ[b], *paf1*Δ, *spt6-140* |

Transcription, txn; splicing, spl; export, exp; translation and decay, tln/dec.

[a]Synthetic lethality of *paf1 spt4* double mutant previously reported by Squazzo *et al.*[2]. [b]Also reported by Rondon *et al.*[1], who found that *spt4 hpr1* and *spt4 tho2* double mutants were inviable.

time, we used a 'one versus all' (OVA) approach that allows discrimination of more than two classes of data (multiclass SVMs[36]). We separately trained the four SVM classifiers, each time defining the separating hyperplane that best separates one class of gene mutations from all the others (such as 'txn' versus 'not txn'). For cross-validation, each mutation was individually withheld from the training set, tested against each of the four SVM classifiers and assigned a label based upon the SVM classifier that gave the highest score. The scores (distances to the separating hyperplane in each of the four classifiers) provide a quantitative estimate of how well each label applies to the mutation or perturbation in question, allowing it to be assigned with more than one label (**Fig. 2**; see also **Supplementary Table 4** online). Finally, we tested several mutations whose impacts might be unclear. These mutations were not included in our training set, but were tested against each of the four SVM classifiers, given scores, and ranked (**Fig. 2**).

To evaluate the multiclass SVM we noted when the highest score a mutant received from among the four classifiers matched the functional group to which it was originally assigned (**Table 1**). This analysis shows that transcription and splicing phenotypes could be accurately predicted from either IA or open reading frame (ORF) data; however, it was more difficult for the SVM to learn the translation and decay phenotype accurately from either set of data. This is probably due to the small number of examples tested and the fact that we likely included nonoverlapping translation and decay functions within the same group for training. Thus this label must be considered with caution, as the controls and mutations with very little evident changes in gene expression also tend to be labeled by the SVM as translation and decay (**Fig. 2**). In contrast, although the export classifier seems to perform more poorly, the kinds of classification discrepancies observed are more consistent with multiple secondary roles for these factors in splicing and transcription, rather than a weak phenotype (see below).

Comparison of the two sources of data, the IA indexes for the ~240 intron-containing genes and the log ratios for ~1,200 intronless genes, reveals that both sets of genes are affected in characteristic ways by the mutations. One notable observation from the SVM analysis is that a specific splicing phenotype can be recognized in the expression profiles of genes lacking introns (**Fig. 2b**). It seems likely that the secondary consequences of splicing inhibition feed back on the expression of intronless genes in a characteristic way.

### Dual function of Lsm6p and Lsm7p

To provide the SVM with the best training, our assignment of labels included only cases in which there was strong evidence for the indicated function. We withheld *lsm6*Δ and *lsm7*Δ from the training set because Lsm6p and Lsm7p are associated with both the U6 snRNP[37] and with the cytoplasmic decay machinery[38]. In contrast, we included *lsm1*Δ in the training set because although Lsm1p is part of the cytoplasmic decay machinery in complex with Lsm6p and Lsm7p, it is not associated with U6 (refs. 38,39). When tested, SVMs trained using the intron-containing gene data clearly identified Lsm6p and Lsm7p as splicing factors, whereas those trained on the intronless gene data labeled them as both decay factors and splicing factors (**Fig. 2**). In addition, Lsm1p was appropriately not recognized as a splicing factor. Thus, the gene expression data must carry signatures of both the common and distinct functions of the Lsm proteins. In other cases, however, multiple functions have been suggested but remain unproven. For some of the perturbations tested here, the SVM assigned a functional label that was different from the label we assigned for training. In some cases the SVM trained with IA index seemed to sense a different function than the SVM trained using the intronless ORF data (**Fig. 2**), as might be expected because the intronless class does not require splicing for its expression. Below we discuss some of these cases and provide new experimental evidence for function at multiple steps in gene expression.

### Ceg1p in transcription and splicing

Previous work has indicated that the capping enzyme Ceg1p contributes to but is not essential for splicing of at least several genes[9,10]. We found that the overall impact of the loss of Ceg1p function *in vivo* is nearly identical to that of bona fide splicing factors (**Fig. 1**; **Supplementary Fig. 3** online). What could account for this apparent strong activating influence of Ceg1p on splicing? The increase in IA index is not primarily due to loss of mRNAs, because intron levels increased in this mutant (**Supplementary Fig. 3** online). In addition, Ceg1's role in splicing must go beyond providing the cap as a binding site for the nuclear cap–binding complex of Mud13p and Gcr3p, previously shown to contribute to splicing[40,41], because loss of these factors produces a much milder defect (**Fig. 1**). It is possible that recognition of the 7-methylguanosine cap (or a product of its removal) may be a prerequisite to efficient decay, and that the triphosphate cap produced in the absence of Ceg1p activity does not suffice for normal decay. Possibilities for normal Ceg1p functions include acting directly as a splicing factor, stimulating pre-mRNA decay, or communicating completion of a transcription step that stimulates splicing indirectly. Whatever the mechanism, our data demonstrate that Ceg1p plays a broad and important role in splicing, similar to that of several essential splicing factors.

### Rrp6p in transcription and decay

The *rrp6*Δ mutation, which eliminates a nuclear subunit of the exosome complex of 3′→5′ exonucleases, has a phenotype recognized by the SVM as similar to transcription mutants, rather than decay mutants (**Fig. 2**),

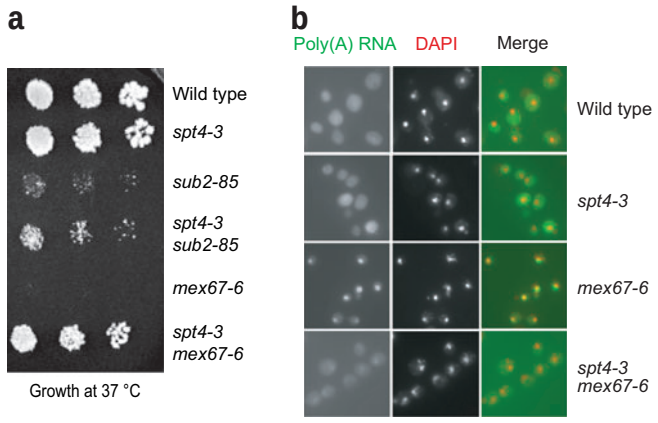


**Figure 3** Suppression of *mex67-6* mRNA export defect by an *spt4*-null allele. (**a**) Growth of strains with different combinations of mutations. (**b**) Suppression of the mRNA export defect of a *mex67-6* mutation by the null allele *spt4-3*. Log-phase cultures of cells with the indicated genotypes were shifted to 39° and then analyzed by *in situ* hybridization with a fluorescently labeled oligo-dT probe, and were also counterstained with DAPI.

consistent with recently observed association of Rrp6 with transcription elongation factors in flies[42]. The clustering experiment agrees with the SVM, and places *rrp6*Δ with *srb10*Δ and *ctk1*Δ (**Fig. 1**). It is unclear exactly how the loss of Rrp6p produces a transcription phenotype or how this protein might contribute to transcription.

### Sub2p in transcription and mRNA export

The *sub2-85* mutation produces a phenotype strongly characteristic of both a transcription factor and an export factor when classified using the IA index data (**Fig. 2a**). Using the intronless gene data, the *sub2-85* phenotype is scored most highly as an export factor, with transcription as the second assignment (**Fig. 2b**). Although Sub2p and its human homolog UAP56 have been implicated in splicing[43–45] and export[13,46–48], it is also a component of the TREX complex and may therefore act in transcription elongation[15,48,49]. The possibility that Sub2p might be a transcription factor is not obvious from its hierarchical clustering with the other export factors Yra1p and Mex67p. However clustering links these export factors with transcription elongation factors, and genetic and molecular evidence (described below) indicates that the gene expression phenotypes exhibited by these mutants suggest function *in vivo* in both transcription and export.

Prompted by the array data analysis, we tested for synthetic lethal interactions between mutations in genes encoding the transcription elongation factors Spt4p, Spt6p, Paf1p and Spt16p, and the subunits of the TREX/THO complex Yra1p and Hpr1p (**Table 2**). Yeast carrying the temperature-sensitive *yra1-1* mutation died at permissive temperature if they also carried any of the mutations *spt4*Δ, *spt6-140*, *spt16-197* or *paf1*Δ. In addition, the *hpr1*Δ mutation was synthetic lethal or showed extremely poor growth in combination with *spt4*Δ, *spt6-140* or *paf1*Δ. These genetic interactions indicate that the export factors of the TREX/THO complex function closely with the Spt4–Spt5 complex and elongation factors of the Paf1 complex. More revealing mechanistically is the suppression of *sub2-85* and *mex67-6* by the null allele *spt4-3* (**Fig. 3a**).

To test whether the mechanism of suppression extends to the process of mRNA export, we used fluorescent *in situ* hybridization to localize polyadenylated transcripts in cells carrying different combinations of *SPT4* and *MEX67* alleles (**Fig. 3b**). In the *spt4-3* cells, polyadenylated RNA was exported in a fashion similar to wild type as indicated by diffuse cytoplasmic fluorescence and the absence of strong nuclear fluorescence. Cells carrying *mex67-6* did not efficiently export polyadenylated RNA at 39 °C (**Fig. 3b**). RNA export in the *mex67-6* strain was restored by *spt4-3*, demonstrating that suppression by *spt4* loss of function includes restoration of efficient nuclear export. Rescue of a block to mRNA export function by loss of elongation factor function suggests at least two mechanisms for the coordinated action of Spt4–Spt5 with the TREX/THO complex. On one hand, loss of Spt4p could slow down elongation rates so that they are compatible with a compromised export machinery. Alternatively, Spt4p could antagonize inappropriate deposition of export complexes on nascent transcripts. Thus loss of Spt4p would rescue the compromised export machinery by allowing export complexes to assemble where they ordinarily would not.

### Ded1p in translation and splicing

Originally implicated in splicing on the basis of its isolation as a suppressor of the *prp8-1* mutation[50], Ded1p was found to be a translation initiation factor[11]. Recently, Ded1p was found to be associated with spliceosomal complexes by mass spectrometry in yeast[51] and human[52] preparations. In addition to the cold-sensitive (cs) *ded1-120* allele, we tested another cs allele, *ded1-199*, and a leaky heat-sensitive (ts) allele, *ded1-95*. Both cs alleles caused a similar strong intron accumulation indicative of splicing inhibition, whereas the *ded1-95* allele had a negligible splicing defect (**Fig. 1**). The SVM classifies *ded1-120* as a splicing factor when using IA indexes, yet it classifies this allele as a translation and decay factor using ORF data. The *ded1-95* allele is classified under translation and decay using either data set. The array data independently predict that Ded1p has roles in both translation and splicing (**Figs. 1** and **2**), consistent with genetic and biochemical observations that initially seemed to conflict.

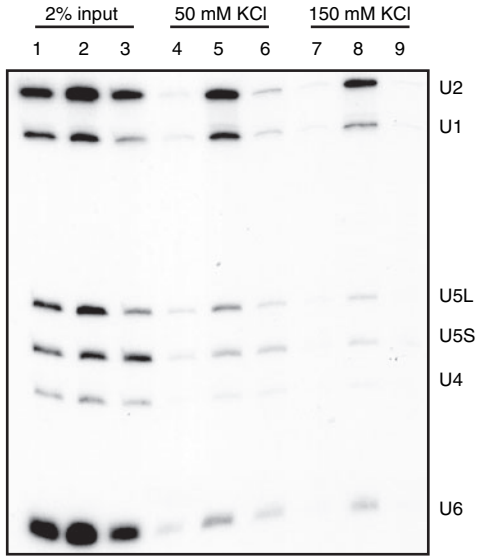


**Figure 4** Ded1p associates with snRNAs *in vitro*. Northern blot analysis of RNA that coprecipitated with protein A–tagged Ded1p was probed with $^{32}$P-labeled PCR products specific for U1, U2, U4, U5 and U6 snRNAs. Lanes 1–3, RNA from input material; lanes 4–6, immunoprecipitates washed in 50 mM KCl; lanes 7–9, immunoprecipitates washed in 150 mM KCl; lanes 1, 4 and 7, extract derived from cells lacking protein A; lanes 2, 5 and 8, extract derived from cells expressing protein A–Ded1p; and lanes 3, 6 and 9, extract derived from cells expressing protein A alone.

To begin investigating the molecular basis of the *ded1-120* phenotype, we tested for the physical association of Ded1p with the splicing apparatus, using extracts from wild-type and protein A–tagged yeast strains (**Fig. 4**). Ded1p is specifically associated with snRNPs in yeast splicing extracts. Based on this experiment, the mass spectrometry result[51], the original genetic findings[50], and additional results indicating that Ded1p contributes to splicing *in vitro* (J.-L.C. and T.-H.C., unpublished data), we suggest that Ded1p function is required both for proper translation in the cytoplasm and for the correct function of the splicing apparatus in the nucleus. Future experiments will be necessary to determine whether and how these two functions might help coordinate and regulate the connections between splicing and ribosome function in yeast, for which there is abundant circumstantial evidence[53].

## DISCUSSION

We have shown that microarray data derived from both intron-containing and intronless genes contain phenotypic information that can identify loss of distinct functions in the gene expression pathway. These data contain information about distinct functions within the transcriptional machinery as shown by hierarchical clustering, and can be more broadly analyzed using SVMs to formally define a phenotype, as well as to identify factors that function in multiple steps of gene expression. For example, this has allowed detection of effects of splicing inhibition on genes that lack introns. As intronless genes make up the vast majority of genes in this organism, *S. cerevisiae* offers the opportunity to trace the regulation the cell may impose to coordinate splicing with other processes. We also find that subtle but relevant differences in function can be detected by separately analyzing splicing data as well as simple expression data from intronless genes. In the absence of coupled *in vitro* assays, this approach allows one to identify proteins coupling distinct steps in the gene expression pathway.

Previously, SVMs have been used to identify genes that exhibit coregulation in many experiments[31]. Here we have used SVMs to identify perturbations that lead to similar expression phenotypes, and infer that these mutations cause loss of function in the same pathway. These experiments provide a context in which to examine the phenotype of any new yeast mutation, allowing evaluation of its contribution to the gene expression pathway. The type of analysis demonstrated here and elsewhere[34] can be applied to other functional categories as well. In effect, the genome is used as a robust reporter of the complex set of direct effects as well as the cascade of indirect consequences of a mutation, and can be used to identify gene functions. As with data from other large-scale approaches, focused molecular tests will be required to distinguish direct from indirect effects. A principled means of characterizing global gene expression phenotypes, such as the SVM approach used here, should substantially complement large-scale physical association studies[54–56] and genome-wide genetic interaction screens[57].

These results provide many specific hypotheses available for testing by molecular and genetic methods. Initial tests (**Figs. 3** and **4**) have supported connections suggested by the phenotype analysis and deserve more detailed investigation. Other findings, such as the classification of Rrp6p as a transcription factor whose loss is similar to that of two CTD kinases, the possible direct roles of the capping enzyme Ceg1p, the Ran GEF Prp20p, or the 5′→3′ nuclear exonuclease Rat1p in splicing, have not yet been tested. Although not every relationship in the data may be validated with available tools in terms of direct functional connections, models of the coordination and coupling of the machineries of the gene expression pathway must eventually account for indirect effects as well. Whereas the development of *in vitro* assays coupling steps in gene expression remains an important goal, *in vivo* analyses like those presented here will remain necessary for understanding the integration of these processes at the cellular and molecular levels.

## METHODS

**Array construction.** Arrays were produced as described[3]. Details of the array designs used are available at the Gene Expression Omnibus (GEO) repository at the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/geo/) (entry GSE1684). Briefly, 5′-amino-linked oligonucleotides were prepared at a concentration of ~10 μg ml$^{-1}$ in 150 mM $NaPO_4$ containing 6.25% (w/v) $Na_2SO_4$ and 0.0005% (w/v) SDS, and robotically spotted onto CodeLink slides. The oligonucleotides used included those described[3] and the Operon whole yeast genome 70-mer set.

**Strains and genetic analysis.** A list of yeast strains and their genotypes as well as experimental conditions for microarray experiments conducted in this work is presented in **Supplementary Table 5** online. The table describes the experimental design for each mutant or treatment, as well as the samples used. To analyze interactions between *yra1-1* and transcription elongation factors, a strain carrying a *yra1* deletion and a wild-type copy of *YRA1* on a *URA3* plasmid was crossed to strains carrying *spt4-3*, *spt6-140*, *spt16-197* and *paf1*Δ mutations. Double-mutant segregants were transformed with *YRA1 TRP1* and *yra1-1 TRP1* plasmids and multiple Trp$^+$ transformants were purified and replica-plated to 5FOA media. In each case, the double mutant was not viable on 5FOA, indicating synthetic lethality between *yra1-1* and the mutation in question.

**RNA isolation, labeling and hybridization.** RNA preparation from yeast cells, probe synthesis and labeling with Cy3 and Cy5 dyes and hybridization were carried out as described[3]. Briefly, RNAs extracted from wild-type and mutant or treated cells were used to make cDNA labeled with Cy3- or Cy5-dUTP. The cDNA target samples were hybridized to oligonucleotide arrays under LifterSlips at 62 °C for 16–24 h in 4× SSC containing 0.1 mg ml$^{-1}$ poly(A) and 0.1% (w/v) SDS. Hybridizations were carried out in duplicate with Cy dyes reversed on the second array (a 'reciprocal label' or 'dye swap' experiment).

**Image processing, normalization, and clustering.** Arrays were scanned with Axon Instruments 4000 series scanners. Preliminary data filtering was done with GenePix Pro software (Axon Instruments). Data were normalized and further processed using R and Bioconductor[58]. Specifically, composite normalization of $\log_2$ ratios was done as described[59], using a set of 1,219 intronless ORFs. Pairs of reciprocally labeled (dye-swapped) normalized data sets were averaged, and IA indexes were calculated as described[3]. Hierarchical clustering was carried out using the Genesis package[60].

**Quantitative PCR analysis.** cDNA synthesis for quantitative PCR (QPCR) was carried out as described for fluorescently labeled target synthesis[3] except that equal concentrations of all four deoxyribonucleotides and no Cy dyes were used. Mock reactions without reverse transcriptase were used to determine genomic DNA contamination. Amplifications were conducted in a Bio-Rad iCycler using iQ SYBR Green Supermix (Bio-Rad) and 200 μM primer according to the manufacturer's instructions, using the indicated primers (**Supplementary Table 2** online). Representative transcripts were assayed in triplicate. These included intron-containing and intronless ribosomal protein and nonribosomal protein genes, as well as structural, noncoding RNAs (**Supplementary Table 1** online). To compare the QPCR with array values we normalized QPCR values to the *PBN1* mRNA. *PBN1* was chosen as a suitable reference gene, as the array data indicated that its expression was unchanged in the seven mutants used in the comparison.

**SVM analysis.** All SVM experiments described in the paper were carried out with Gist version 2.0.8 (ref. 61). A linear kernel was applied for all SVM classifiers. For the multiclass SVM analysis we applied the OVA approach[36]. To evaluate the SVM performances of each classifier we plotted a ROC (receiver operating characteristic) curve for each classifier and calculated the ROC value. A ROC curve describes the relationship between the false positive rate (FPR, 1-specificity), and the true positive rate (TPR, sensitivity). FPR is the fraction of negatives incorrectly predicted as positive (FP / all negatives) whereas the TPR is the fraction of positives correctly predicted (TP / all positives). The ROC value (between 0 and 1) is the fractional area under the ROC curve (**Supplementary Fig. 4** online). For transcription (ROC values of 0.97 for IA, 0.98 for ORF data) and splicing (0.96 IA, 0.98 ORF) in particular, and to a lesser extent for translation and decay (0.87 IA, 0.86 ORF) and export (0.76 IA, 0.88 ORF), labels were assigned

correctly. This indicates that mutations in these groups share characteristic gene expression signatures representing their loss of function. SVM scores are found in **Supplementary Table 4** online.

***In situ* analysis of poly(A) RNA.** This method was adapted from a Singer lab protocol found at http://singerlab.aecom.yu.edu/protocols. Yeast cells were grown in 45 ml YPD cultures at 30 °C to early log phase ($1–2 \times 10^7$ cells $ml^{-1}$) and split cells into equal 22.5 ml cultures. One culture was fixed by the addition of formaldehyde to a final concentration of 4% (v/v), with shaking occasionally for 40 min at room temperature. Cells were gently washed 3× with 1 ml buffer B at 4 °C (1.2 M sorbitol, 0.1 M potassium phosphate, pH 7.5), pelleted by centrifugation each time for 2 min at 2,300*g* and 4 °C, and finally resuspended in 1 ml buffer B. The second culture was heat-shifted by pelleting and resuspending cells in prewarmed 39 °C media, and incubating at 39 °C for 1 h. Cells were then fixed and washed as described above. Next, cells were spheroplasted in 500 µl buffer B containing 10 mM VRC, 28 mM 2-mercaptoethanol, 0.06 mg $ml^{-1}$ PMSF and 40 mg $ml^{-1}$ µg $ml^{-1}$ zymolyase at 37 °C for 15 min, washed with 1 ml ice-cold buffer B and resuspended in 500 µl buffer B. Cell suspension (20 µl) was added to a polylysine-coated slide well, allowed to adhere for 30 min at 4 °C, and removed. Wells were then covered with 70% (v/v) ethanol and incubated for ~6 h at –20 °C. Cells were rehydrated with 20 µl 2× SSC, 50% (v/v) formamide for 5 min at room temperature. Cy3 was coupled to a 60-nucleotide oligo-dT polymer with an amino-substituted dT at every tenth position. This oligonucleotide (5 µg) was diluted to 35 µl with 0.1 M $Na_2CO_3$, pH 9.0, and combined with Cy3 mono NHS ester (Amersham) according to the manufacturer's instructions, and incubated overnight in the dark. The labeled oligonucleotide was recovered on a G50 column, and adjusted to 45 ng $µl^{-1}$. Cy3-oligo $dT_{60}$ probe was allowed to hybridize to slides overnight at 37 °C in a 20 µl mixture containing 10% (w/v) dextran sulfate, 2 mM VRC, 0.02% (w/v) RNAse-free BSA, 25 µg *E. coli* tRNA, 12.5 µg single-stranded DNA, 2× SSC, 50% (v/v) formamide and 50 ng probe (CY3-oligo $dT_{60}$). Slides were washed twice with 2× SSC, 15% (v/v) formamide, 0.1% (w/v) SDS for 30 min at 37 °C, and once with PBS. DAPI was diluted 1:500 in PBS, 3% (w/v) BSA, 0.05% (v/v) Tween, 0.04% (w/v) sodium azide, added to cells for 1 min, and removed. Slides were then washed three times with PBS, 0.05% (v/v) Tween and once with PBS. Finally, 3 µl mounting media was added to each well and a coverslip sealed on with nail polish. Images were captured with a Zeiss Axioskop II fluorescence microscope.

**Ded1-protein A immunoprecipitation**. Splicing extracts were prepared and immunoprecipitation reactions were conducted as described[62]. Briefly, splicing extracts prepared from tagged (YTC212 *MATa ded1:TRP1 ura3-52 lys2-801 ade2-101 trp1Δ1 his3Δ200 leu2Δ1* (pDED1033:DED1-protein A)) and untagged (YTC75 *MATa ded1:TRP1 ura3-52 lys2-801 ade2-101 trp1Δ1 his3Δ200 leu2Δ1* (pDED1033:DED1)) strains were incubated with IgG-Sepharose beads for 1 h at 4 °C and washed seven times with 1 ml of buffer D (20 mM HEPES, pH 7.9, 0.2 mM EDTA, 20% (v/v) glycerol, 0.5 mM DTT, 0.05% (v/v) NP-40 containing either 50 mM or 150 mM KCl, as indicated). Splicing extract incubated without IgG-Sepharose beads served as an input control. RNA present in immunoprecipitation reaction was isolated by the addition 250 µl RNA extraction buffer (50 mM NaOAc, 1 mM EDTA, 0.1% (w/v) SDS), 300 µl phenol-chloroform and incubation at 65 °C for 10 min. Extracted RNA was ethanol-precipitated, separated on a 6% (w/v) acrylamide denaturing PAGE gel and transferred to nylon membrane. Northern hybridization was conducted using standard techniques. Randomly labeled probes specific to U1, U2, U4, U5 and U6 small nuclear RNAs (snRNAs) were prepared from equal-length DNA fragments obtained by PCR amplification.

**Deposit of array data at GEO.** Array data described in this manuscript is available through the Gene Expression Omnibus (GEO) repository at the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/geo/) (accession code GSE1684).

*Note: Supplementary information is available on the Nature Structural & Molecular Biology website.*

ACKNOWLEDGMENTS

We thank B. Noble for help with SVM analysis, S. Ruby, P. Silver, E. Hurt and L. Hicke for yeast strains, and A. Caroll and J. DeRisi for advice and assistance with microarrays. This work was primarily funded by grants to M.A. from the US National Institutes of Health (NIH) (GM040478), the Packard Foundation, and the W.M. Keck Foundation (to the University of California Santa Cruz RNA Center). The Howard Hughes Medical Institute Professors program supported Y.M.-G. Grants to G.H. from the NIH (GM060479), and the University of California Cancer Research Coordinating Committee supported this work as well, and T.-H.C. was supported by the NIH (GM48752) and the US National Science Foundation (MCB-9982726).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 26 July; accepted 6 December 2004
Published online at http://www.nature.com/nsmb/

1. Reed, R. Coupling transcription, splicing and mRNA export. *Curr. Opin. Cell. Biol.* **15**, 326–331 (2003).
2. Dimaano, C. & Ullman, K.S. Nucleocytoplasmic transport: integrating mRNA production and turnover with export through the nuclear pore. *Mol. Cell. Biol.* **24**, 3069–3076 (2004).
3. Clark, T.A., Sugnet, C.W. & Ares, M. Jr. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* **296**, 907–910 (2002).
4. Wang, Y. *et al.* Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. USA* **99**, 5860–5865 (2002).
5. Holstege, F.C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
6. Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M.D. & Hughes, T.R. Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals post-transcriptional control of ribosome biogenesis factors. *Mol. Cell. Biol.* **24**, 5534–5547 (2004).
7. Hirose, Y., Tacke, R. & Manley, J.L. Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev.* **13**, 1234–1239 (1999).
8. Zeng, C. & Berget, S.M. Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. *Mol. Cell. Biol.* **20**, 8290–8301 (2000).
9. Schwer, B. & Shuman, S. Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing *in vivo*. *RNA* **2**, 574–583 (1996).
10. Fresco, L.D. & Buratowski, S. Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but is not required for, mRNA splicing. *RNA* **2**, 584–596 (1996).
11. Chuang, R.Y., Weaver, P.L., Liu, Z. & Chang, T.H. Requirement of the DEAD-box protein ded1p for messenger RNA translation. *Science* **275**, 1468–1471 (1997).
12. Stutz, F. & Izaurralde, E. The interplay of nuclear mRNP assembly, mRNA surveillance and export. *Trends. Cell Biol.* **13**, 319–327 (2003).
13. Strasser, K. & Hurt, E. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* **413**, 648–652 (2001).
14. Segref, A. *et al.* Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)$^+$ RNA and nuclear pores. *EMBO J.* **16**, 3256–3271 (1997).
15. Strasser, K. *et al.* TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**, 304–308 (2002).
16. Herold, A., Teixeira, L. & Izaurralde, E. Genome-wide analysis of nuclear mRNA export pathways in *Drosophila*. *EMBO J.* **22**, 2472–2483 (2003).
17. Rehwinkel, J. *et al.* Genome-wide analysis of mRNAs regulated by the THO complex in *Drosophila melanogaster*. *Nat. Struct. Mol. Biol.* **11**, 558–566 (2004).
18. Hartzog, G.A., Speer, J.L. & Lindstrom, D.L. Transcript elongation on a nucleoprotein template. *Biochim. Biophys. Acta* **1577**, 276–286 (2002).
19. Mueller, C.L. & Jaehning, J.A. Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Mol. Cell. Biol.* **22**, 1971–1980 (2002).
20. Squazzo, S.L. *et al.* The Paf1 complex physically and functionally associates with transcription elongation factors *in vivo*. *EMBO J.* **21**, 1764–1774 (2002).
21. Krogan, N.J. *et al.* RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol. Cell. Biol.* **22**, 6979–6992 (2002).
22. Lindstrom, D.L. *et al.* Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell. Biol.* **23**, 1368–1378 (2003).
23. Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F. & Buratowski, S. Transitions in RNA polymerase II elongation complexes at the 3′ ends of genes. *EMBO J.* **23**, 354–364 (2004).
24. Chang, M. *et al.* A complex containing RNA polymerase II, Paf1p, Cdc73p, Hpr1p, and Ccr4p plays a role in protein kinase C signaling. *Mol. Cell. Biol.* **19**, 1056–1067 (1999).
25. Kobor, M.S. & Greenblatt, J. Regulation of transcription elongation by phosphorylation. *Biochim. Biophys. Acta* **1577**, 261–275 (2002).
26. Lee, T.I. & Young, R.A. Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**, 77–137 (2000).
27. Fischbeck, J.A., Kraemer, S.M. & Stargell, L.A. SPN1, a conserved gene identified by suppression of a postrecruitment-defective yeast TATA-binding protein mutant. *Genetics* **162**, 1605–1616 (2002).
28. Skaar, D.A. & Greenleaf, A.L. The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3′ cleavage/polyadenylation through the processing component Pti1p. *Mol. Cell* **10**, 1429–1439 (2002).
29. Ahn, S.H., Kim, M. & Buratowski, S. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. *Mol. Cell* **13**, 67–76 (2004).

30. Betz, J.L. *et al.* Phenotypic analysis of Paf1/RNA polymerase II complex mutations reveals connections to cell cycle regulation, protein synthesis, and lipid and nucleic acid metabolism. *Mol. Genet. Genomics* **268**, 272–285 (2002).
31. Brown, M.P. *et al.* Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. USA* **97**, 262–967 (2000).
32. Lee, Y. & Lee, C.K. Classification of multiple cancer types by multicategory support vector machines using gene expression data. *Bioinformatics* **19**, 1132–1139 (2003).
33. Furey, T.S. *et al.* Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* **16**, 906–914 (2000).
34. Mnaimneh, S. *et al.* Exploration of essential gene functions via titratable promoter alleles. *Cell* **118**, 31–44 (2004).
35. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
36. Ramaswamy, S. *et al.* Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. USA* **98**, 15149–15154 (2001).
37. Pannone, B.K., Kim, S.D., Noe, D.A. & Wolin, S.L. Multiple functional interactions between components of the Lsm2–Lsm8 complex, U6 snRNA, and the yeast La protein. *Genetics* **158**, 187–196 (2001).
38. Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M. & Seraphin, B. A Sm-like protein complex that participates in mRNA degradation. *EMBO J.* **19**, 1661–1671 (2000).
39. Tharun, S. *et al.* Yeast Sm-like proteins function in mRNA decapping and decay. *Nature* **404**, 515–518 (2000).
40. Izaurralde, E. *et al.* A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**, 657–668 (1994).
41. Lewis, J.D., Gorlich, D. & Mattaj, I.W. A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing. *Nucleic Acids Res.* **24**, 3332–3336 (1996).
42. Andrulis, E.D. *et al.* The RNA processing exosome is linked to elongating RNA polymerase II in *Drosophila*. *Nature* **420**, 837–841 (2002).
43. Libri, D., Graziani, N., Saguez, C. & Boulay, J. Multiple roles for the yeast SUB2/yUAP56 gene in splicing. *Genes Dev.* **15**, 36–41 (2001).
44. Kistler, A.L. & Guthrie, C. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. *Genes Dev.* **15**, 42–49 (2001).
45. Fleckner, J., Zhang, M., Valcarcel, J. & Green, M.R. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.* **11**, 1864–1872 (1997).
46. Jensen, T.H., Boulay, J., Rosbash, M. & Libri, D. The DECD box putative ATPase Sub2p is an early mRNA export factor. *Curr. Biol.* **11**, 1711–1715 (2001).
47. Gatfield, D. *et al.* The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in *Drosophila*. *Curr. Biol.* **11**, 1716–1721 (2001).
48. Zenklusen, D., Vinciguerra, P., Wyss, J.C. & Stutz, F. Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Mol. Cell. Biol.* **22**, 8241–8253 (2002).
49. Rondon, A.G., Jimeno, S., Garcia-Rubio, M. & Aguilera, A. Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. *J. Biol. Chem.* **278**, 39037–39043 (2003).
50. Jamieson, D.J., Rahe, B., Pringle, J. & Beggs, J.D. A suppressor of a yeast splicing mutation (prp8-1) encodes a putative ATP-dependent RNA helicase. *Nature* **349**, 715–717 (1991).
51. Stevens, S.W. *et al.* Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol. Cell* **9**, 31–44 (2002).
52. Zhou, Z., Licklider, L.J., Gygi, S.P. & Reed, R. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182–185 (2002).
53. Ares, M., Jr., Grate, L. & Pauling, M.H. A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* **5**, 1138–1139 (1999).
54. Shevchenko, A., Schaft, D., Roguev, A., Pijnappel, W.W. & Stewart, A.F. Deciphering protein complexes and protein interaction networks by tandem affinity purification and mass spectrometry: analytical perspective. *Mol. Cell. Proteomics* **1**, 204–212 (2002).
55. Legrain, P. & Selig, L. Genome-wide protein interaction maps using two-hybrid systems. *FEBS Lett.* **480**, 32–36 (2000).
56. Zhong, J., Zhang, H., Stanyon, C.A., Tromp, G. & Finley, R.L. Jr. A strategy for constructing large protein interaction maps using the yeast two-hybrid system: regulated expression arrays and two-phase mating. *Genome Res.* **13**, 2691–2699 (2003).
57. Tong, A.H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
58. Dudoit, S., Gentleman, R.C. & Quackenbush, J. Open source software for the analysis of microarray data. *Biotechniques* (suppl.), 45–51 (2003).
59. Yang, Y.H. *et al.* Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**, e15 (2002).
60. Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics* **18**, 207–208 (2002).
61. Pavlidis, P., Wapinski, I. & Noble, W.S. Support vector machine classification on the web. *Bioinformatics* **20**, 586–587 (2004).
62. Cheng, S.C., Newman, A.N., Lin, R.J., McFarland, G.D. & Abelson, J.N. Preparation and fractionation of yeast splicing extract. *Methods Enzymol.* **181**, 89–96 (1990).