

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Splicing is regulated in response to environmental changes in yeast

Permalink

<https://escholarship.org/uc/item/5n39v100>

Author

Whitworth, Gregg Brooks

Publication Date

2008-07-28

Peer reviewed|Thesis/dissertation

Splicing is regulated in response to environmental changes in yeast

by

Gregg B. Whitworth

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Dedication and Acknowledgements

First and foremost, the dedication of this thesis goes to my advisor, Christine Guthrie. Christine has been a teacher, mentor, and guru to me in both my scientific work and life. Throughout my time in Christine's lab she challenged and supported me, affording me both the space and tools to develop as a scientist. The elegance and beauty with which Christine approaches science, and her dedication to the humanity inherent in the endeavor, represent ways of being in science to which I will always strive. I am deeply honored to have been her student, and eagerly look forward to continuing to learn from her as much in the coming years as in those that have passed.

Wendy Gilbert first introduced me to the key question which has motivated all of my work as a graduate student: how are post-transcriptional processes tweaked by changes in the environment? Since our first meeting, I have stood in awe of both the creativity and rigor of Wendy's approach to science. Wendy has shown me what science can be at its best, and I am eager to continue to follow along as she moves into the next phase of her work.

Without Jeffrey Pleiss, none of the work in this thesis would have been possible. Jeff's stunning technical work has had a direct or indirect hand in generating the vast majority of data which appears on the following pages. It was wonderfully fortunate that the synergy between our intellectual interests in the lab built the framework for an incredibly close collaboration. Jeff has been a wonderful mentor to me throughout my graduate work, and it has been a great pleasure to get to work so closely with him.

Perhaps more than anyone, however, Megan Bergkessel has shaped the intellectual focus of the most meaningful facets of this work, and of my own thinking. Her love of synthesizing wide ranging findings, and comfortably adventuring into novel and challenging intellectual terrain, are unmatched and utterly contagious. Everything is, in fact, connected – and I look forward to every opportunity that presents itself to continue explore the details with Megan.

I also owe an incredible debt of thanks to everyone I have had the pleasure of working with, and learning from, in scientific family that is the Guthrie lab. Graduate students – Kent Duncan, Karen Kim, Maki Inada, Tamara Brenner (I eventually remembered to publish and go home), and Alex Polcik – and post-docs – Anne de Bruyn-Kups, Stephen Rader, Pascal Preker, Tommaso Villa, Alan Kutach, Mette Lund, Tracy Kress, Gwen Wilmes, Quinn Mitrovich, and Corina Maeder – have all been wonderful mentors, collaborators, and friends.

My committee members, Sandy Johnson, Erin O’Shea, and Jonathan Weissman have each generously offered their time and many valuable insights over the last many years. I have been a fortunate beneficiary of their dedication to graduate education and scientific progress.

Finally, I would like to acknowledge two other wonderful scientific mentors who led me down the road to my graduate work. Paulette Levantine offered me my first glimpse of the astounding beauty of the molecular basis of life. Her approach teaching, thinking about, and exploring the world of science profoundly changed the course of my life. It was in Leslie Gregg-Jolly’s classroom and research lab that I truly fell in love with

science. She mentored my transition from a consumer of scientific information, to an understanding that I possessed the agency to contribute to the cannon of scientific knowledge myself – and she created a space in which doing so was simply thrilling. I am eagerly looking forward to the opportunity to turn once more towards her mentorship.

Published Materials

Chapter 1

The text of this chapter is a reprint of the material as it appears in PLoS Biology (citation below), and represents work performed in collaboration with Jeffery A. Pleiss and Megan Bergkessel.

Pleiss, J.A., **G.B. Whitworth**, M. Bergkessel and C. Guthrie. (2007) Transcript Specificity in Yeast Pre-mRNA Splicing Revealed by Mutations in Core Spliceosomal Components. *PLoS Biology* **5**: e90

Chapter 2

The text of this chapter is a reprint of the material as it appears in Molecular Cell (citation below), and represents work performed in collaboration with Jeffery A. Pleiss and Megan Bergkessel.

Pleiss, J.A., **G.B. Whitworth**, M. Bergkessel and C. Guthrie. (2007) Rapid, Transcript-Specific Changes in Splicing in Response to Environmental Stress. *Molecular Cell* **27**: 928-937.

Unpublished Materials

Appendix 1

This appendix represents work performed in collaboration with Megan Bergkessel. Additionally, the following individuals contributed microarray data included in the analysis: Anne de Bruyn Kops, Wendy Gilbert, Haralambos Hadjivassiliou, Alan Kutach, Tracy Kress, Mette Lund, Jeffery A. Pleiss, and Tomasso Villa.

Abstract

Eukaryotic mRNAs must undergo a diverse set of complex processing events before the messages that they carry can be competent for translation. Each of these stages in the life of an mRNA represent potential opportunities for regulating the fate of a message. In this work we have explored the relationship between changing cellular needs, induced by alterations in environmental conditions, and the process of splicing. Across a wide range of different environmental contexts we see evidence of rapid, transcript-specific changes in splicing efficiency. The splicing of transcripts can be both up-regulated and down-regulated in response to environmental cues. In some cases large numbers of transcripts are affected, in others only defined subsets. In total, our data reveal that splicing is a remarkably dynamic and flexible mechanism for post-transcriptional gene regulation in yeast.

Table of Contents

Introduction: Life in the Eukaryotic mRNA World	1
Chapter 1: Transcript Specificity in Yeast pre-mRNA Splicing Revealed by Mutations in Core Spliceosomal Components	18
Chapter 2: Rapid, Transcript-Specific Changes in Splicing in Response to Environmental Stress	77
Appendix 1: Exploring the Scope and Dynamics of Splicing Regulation	110
Appendix 2: Rapid Changes in the Behavior of Nuclear mRNA Export Factors Following Environmental Stress	130
Appendix 3: Screen for Environmental Stress-Induced Growth Phenotypes Caused by Mutations in mRNA Processing Factors	152
Appendix 4: Microarray Data Analysis Toolkit	162

List of Tables

Table 1. Relative copy numbers of total and pre-mRNA determined by quantitative RT-PCR	65
Table 2. Oligonucleotide sequences used in QPCR experiments.....	66
Table 3. Evidence for differential expression in the <i>prp2-1</i> mutant.	67
Table 4. Evidence for differential expression in the <i>prp8-1</i> mutant.	68
Table 5. Evidence for differential expression in the <i>prp5-1</i> mutant.	69
Table 6. Evidence for differential expression in the <i>prp8-101</i> mutant.	70

List of Figures

Figure 1. Properties of yeast introns.	46
Figure 2. Monitoring genome-wide changes in pre-mRNA splicing.	47
Figure 3. Comparing the cellular and molecular phenotypes of three different spliceosomal mutants.	48
Figure 4. Two different alleles of Prp8 produce very different molecular phenotypes....	50
Figure 5. A global view of defects in pre-mRNA splicing.	51
Figure 6. Transcript-specific view of defects in mRNA splicing.	52
Figure 7. Analysis of splice site sequences and transcript behavior in comparisons between <i>prp2-1</i> , <i>prp8-1</i> and <i>prp5-1</i>	53
Figure 8. Validation of microarrays using quantitative RT-PCR.	54
Figure 9. The extent of precursor accumulation is independent of initial precursor levels.	55
Figure 10. Splicing mutants examined.	56
Figure 11. Individual splicing profiles of each mutant examined.	57
Figure 12. High-throughput sample collection and preparation.	58
Figure 13. Design scheme for mature mRNA probes.	60
Figure 14. Microarray design.	61
Figure 15. Global quality assessment for an individual microarray.	62
Figure 16. Quality assessment of within-array and between-array replicate agreement.	63
Figure 17. Splicing-specific microarrays.	95
Figure 18. Regulation of pre-mRNA splicing in response to amino acid starvation.	96

Figure 19. Quantitative RT-PCR validates the rapid, transcript-specific downregulation of splicing in response to amino acid starvation.	97
Figure 20. The splicing response to amino acid starvation does not require the activity of Gcn2.	98
Figure 21. Regulation of pre-mRNA splicing in response to Ethanol Toxicity.	99
Figure 22. Comparison of splicing responses to amino acid starvation and ethanol toxicity.	100
Figure 23. Accumulation of pre-mRNAs as measured using different primers.	101
Figure 24. Analysis of splice site sequences and transcript behavior following amino	102
Figure 25. A variety of stresses induce defined effects on a small subset of transcripts.	124
Figure 26. U3 splicing efficiency is anti-correlated with RPG expression levels.	126
Figure 27. Comparison of amino acid starvation induced by 3-AT addition, and environmental amino acid withdrawal.	127
Figure 28. Analysis of similarity in splicing behavior of all intron-containing transcripts in response to mRNA processing mutations and environmental stress.	128
Figure 29. Changes in Npl3 localization and poly(A) RNA association following salt stress.	143
Figure 30. Changes in mRNA association with Npl3 versus Nab2 following salt stress.	144
Figure 31. Changes in the association of ribosomal protein gene transcripts with Npl3 and Nab2 following salt stress.	146
Figure 32. Changes in Npl3 localization following 300mM LiCl treatment.	148
Figure 33. Changes in Sub2 localization following 300mM LiCl treatment.	149
Figure 34. Changes in Yra1 localization following 300mM LiCl treatment.	150

Figure 35. Changes in poly(A) RNA localization following treatment with NaCl, LiCl and heat shock.....	151
Figure 36. Stress phenotypes of deletions of non-essential splicing factors.....	157
Figure 37. Stress phenotypes of deletions of Sky1 and closely related kinases.....	158
Figure 38. Allele specific stress phenotypes of Sub2 mutants.....	159
Figure 39. Stress phenotypes of deletions of nuclear pore components.....	160
Figure 40. Stress phenotypes of factors implicated in mRNA transport and 3' end processing.....	161
Figure 41. Schematic representation of the flow of information through our microarray data analysis procedure.....	165
Figure 42. Dot plot of replicate spots.....	166
Figure 43. Variation of M values for spot replicates on a single array.....	167
Figure 44. Variation of M values for spot replicates between arrays.....	168
Figure 45. Correlation plot for dye-flipped replicates.....	169
Figure 46. Feature type comparison for a given transcript.....	170
Figure 47. Overview of the Ida splicing microarray data analysis software suite.....	171

Introduction: Life in the Eukaryotic RNA World

Scientific Context

The life of a eukaryotic mRNA is astoundingly complex. From the moment a nascent transcript emerges from the elongating RNA polymerase to the its final moments of existence, eukaryotic mRNAs are constantly engaged with, and acted upon, by a sophisticated set of molecular machines. If all goes well, transcripts are capped at their 5' ends, their introns are removed, their 3'ends are cleaved and a poly(A) tail added, they are handed off to the nuclear export machinery and make their way to the cytoplasm, and are finally remodeled and transitioned to a state of active translation. Each of these stages in an mRNA's life are carried out by sophisticated macromolecular complexes of proteins, and in some cases functional RNAs. The activities of these machines are energy dependent and amazingly complex; yet, each of these stages in an mRNA's life are necessary for gene expression and are universally conserved across the eukaryotic domain.

So, why all the bother? Our bacterial cousins largely couple translation directly to the process of transcription, both physically and temporally, reducing the role of mRNA to its namesake: a simple and direct messenger. Meanwhile the eukaryotic cell has taken a very different path. It has vastly expanded the role of the mRNA by virtue of the sheer number of manipulations which it must undergo in order for the genetic information it carries to see fruition in new protein production. In fact, it is likely that these mRNA processing steps which gate gene expression are as intrinsically eukaryotic as the nucleus itself.

A priori, the explanation I find most compelling is that fundamentally, in the flux through an information carrying pathway, complexity is useful insofar as it affords the opportunity for fine tuned manipulation of the system. To put this in more biological terms, the model is: complexity emerges because it provides an opportunity for regulation. This model is not meant as a explanation for how the happy accidents of increased complexity initially arise in particular biological systems (for example the origins of introns; Izquierdo and Valcarcel, 2006.), but instead it is meant to offer one compelling reason for why increasingly complex systems become codified in biology over the course of evolutionary time. This model is at the heart of the question which has motivated this thesis: is complexity in the various stages of eukaryotic gene expression maintained over the course of time because it affords cells the opportunity to regulate gene expression?

Splicing as a model for post-transcriptional regulation

Many decades of careful work have lead us to a fairly coherent view of how transcription is regulated to effect changes in gene expression in response to changing cellular needs. Although many interesting questions remain, the fundamental rules which govern the regulation of the initial synthesis of an mRNA are now well understood. For the other stages in a eukaryotic mRNA's life, however, much less is known. Only now are we beginning to appreciate the extent to which eukaryotic gene expression *can* be regulated post-transcriptionally. How and why regulation may occur at different steps in mRNA processing remain open and exciting questions.

Armed with an interest in trying to identify contexts in which the post-transcriptional mRNA processing machinery might be leveraged by the cell to modulate flux in gene expression, the most important decision I had to make in the course of my work was to pick a likely, tractable, and timely model process. There were several reasons why splicing rose to the top as the clear winner.

Alternative Splicing

First, there were compelling examples of specific kinds of regulated splicing. Here, the most prolific literature comes from studies of alternative splicing in metazoans (Gravely, 2001). In higher eukaryotes, where the majority of transcripts contain multiple introns, modulating spliceosomal recognition of particular splice sites can lead to inclusion or exclusion of specific exons. These alternative exons allow single genetic loci to effectively encode many different protein isoforms. In an extreme example, the potential information complexity of a single locus can exceed the number of genes in the entire genome (Schmucker et al., 2000). It is very likely that the development of alternative splicing has been a major factor in allowing the metazoan branch to experience a phenomenal increase in tissue diversity while maintaining relatively moderate increases in total genome size.

Alternative splicing is regulated by at least two large families of proteins, which can both promote and inhibit splice site recognition and subsequent splicing efficiency of their substrates in various contexts (Gravely, 2000; Matlin et al., 2005). Enhancer/repressor sequences recognized by these factors have been found in both exonic and intronic sequences. In a small number of well studied examples, developmental and tissue-specific regulation of the expression or activity of splicing enhancers and repressors has been shown to allow for condition-specific regulation of alternative splicing decisions.

However, although alternative splicing exists outside of the metazoan branch, the capacity for alternative splicing it is far less prevalent across the rest of the eukaryotic domain. In *Saccharomyces cerevisiae*, for example, the vast majority of spliced transcripts contain a single intron. Whereas largely degenerate splice site sequences in higher-eukaryotes promote flexibility in the interaction between introns and the spliceosome, which can in turn be enhanced or repressed by the activity of alternative splicing regulators, splice sites in yeast adhere to an impressively strict consensus.

Nonetheless, yeast, like all eukaryotes, have introns and have conserved the full suite of the nuclear splicing machinery. The presence of introns in yeast and across all eukaryotic genomes in contexts where there is little or no capacity for alternative splicing, suggests a role for splicing that is ancestral to the specific case of alternative exon regulation.

Identity of intron-containing genes in yeast

The second draw to studying splicing in *Saccharomyces cerevisiae* came from a consideration of the identities of intron-containing transcripts. Unlike most metazoans, a relative minority of yeast genes have introns: only ~260 of the ~6,000 genes in the genome. This 6% of the genome, however, represents a remarkable collection of factors. First, intron-containing genes are generally highly expressed, representing about 30% of the total transcript population under normal conditions (Ares et al., 1999). Second, the vast majority of intron-containing genes can be directly linked to processes which are integral to setting the physiological pace of the cell.

The largest group of intron-containing genes are the ribosomal protein genes (RPGs). A full 40% of introns fall in RPGs, and approximately 60% of RPGs contain introns. Other statistically over represented categories of intron containing genes include factors involved in protein targeting and vesicle fusion, structure and dynamics of the actin and tubulin cytoskeletons, and factors involved in post-translational modifications of proteins including ubiquitination and ubiquitin-like modifications.

A priori, the set of intron-containing genes in yeast appears to be composed of exactly the kinds of factors one would expect to be under sophisticated and dynamic regulatory control.

A well understood mechanism

Another significant advantage to choosing splicing as a model for post-transcriptional control of gene expression is our impressively detailed understanding of the mechanisms that drive the process (Staley and Guthrie, 1998). Although many compelling questions remain, we have a sophisticated understanding of the mechanistic interplay between the 5 snRNPs that comprise the spliceosome and its intron-containing substrates. Furthermore, splicing has been sufficiently well studied across a variety of different organisms, that we have a clear appreciation of how tightly conserved the mechanisms which underlie splicing are across the eukaryotic domain.

The work which appears in chapter 1 heavily leverages both the rich intellectual history behind our understanding of the mechanisms of splicing, and the vast array of mutations in core spliceosomal components available in our experimental toolkit.

The right tool for the job

The last piece of the puzzle was finding the right assay. If the goal of a project is to ask “if”, “when”, and “how” splicing is regulated in yeast, then the ideal experimental approach is an assay that lets one observe changes to the splicing efficiency of every intron-containing transcript in cells engaged in adaptation to changing environmental conditions. I was incredibly fortunate that, at the start of my graduate work, a new experimental platform matching exactly those criteria was being developed in the Guthrie lab by Jeffrey Pleiss. The splicing-specific microarrays (Figure 2) which we developed over the course of the following years have allowed us to simultaneously measure differences in the levels of total mRNA, pre-mRNA, and mature mRNA species of all intron-containing transcripts between two experimental samples. By comparing RNA collected from yeast grown under normal conditions, to RNA collected from cells at various stages of adaption to a new environmental contexts, we have been able to compile a global picture of how splicing activity is able to respond to changes in the environment.

In Appendix 4, I discuss a related line of tool development that was a necessary component of this project: refining our strategies for microarray data analysis. This effort led us to develop new visualization and data exploration techniques, to allow us to interact with the complex data produced by the splicing-specific microarrays in a meaningful way. To support these techniques we have implemented a host of novel software solutions.

The capacity for splicing regulation

Fundamentally, in order for a step in gene expression to be a meaningful target for biological regulation it must exhibit one key feature: its activity must be able to be modulated in a transcript-specific manner. In the realm of splicing in yeast there was cause for skepticism on this front. When we began this work, the prevailing notion was that, in yeast, “generic” spliceosomes interacted uniformly with “generic” intron-containing substrates. This idea was fueled by the strong consensus sequences at yeast splice sites, the apparent efficiency with which many yeast transcripts are spliced, and the absence of any known *cis* or *trans* enhancers of splicing activity. For this later point it is important to note that although yeast has at least three clear homologs of mammalian SR proteins, Npl3, Gbp2, and Hrb1, the activities of each have been traditionally ascribed to other steps in gene expression. Only very recently has a role in splicing been discovered for one of these factors, Npl3 (Tracy Kress, work in progress).

The work described in chapter 1 addressed this question. In our investigation of the splicing profiles associated with a wide array of different mutations in core spliceosomal factors, we discovered a remarkable degree of complexity in the interaction between the spliceosome and the gamut of its substrates. Although the majority of mutations we examined in this study had been identified because they (1) inhibited growth when cells were shifted to a non-permissive temperature, and (2) inhibited the splicing of a reporter construct *in vitro*, no single mutation we examined showed a block to the splicing of all expressed intron-containing genes. The splicing effects of different mutants on specific subsets of transcripts ranged from subtly non-overlapping to profoundly distinct. Figure 6 illustrates the most important conclusion from this study, by highlighting numerous examples of non-overlapping sensitivities of specific transcripts to particular

spliceosomal mutations: that transcript specificity is built into the activity of the core spliceosome.

Although studying the activity of the spliceosome through the lens of conditional alleles in splicing factors shifted to their non-permissive temperatures represents a fairly artificial context, the finding that the yeast spliceosome was *capable* of exhibiting transcript-specificity in its activity opened the door to the possibility that splicing *could be* effectively regulated in yeast.

Splicing regulation in action

Chapter 2 describes the first experiments which ultimately demonstrated that splicing in yeast *is* regulated.

In this study, we showed that two different environmental conditions, amino acid starvation induced by the drug 3-aminotriazole (3-AT) and exposure to toxic concentrations of ethanol, both induced rapid and transcript-specific changes to splicing in yeast. In response to amino acid starvation we observed an inhibition in the splicing efficiency of a large number of ribosomal protein genes. By contrast, ethanol toxicity induced an inhibition in the splicing efficiency of an unrelated set of transcripts, and increased the splicing efficiency of a third subset. Together these data demonstrated that the splicing activity of specific subsets of transcripts could be rapidly down-regulated or up-regulated in response to changing environmental conditions.

In more recent work described in Appendix 1, these themes are explored across a wider range of environmental conditions. Indeed, under nearly all of the environmental contexts we have assayed we see evidence for transcript-specific changes in splicing efficiency. Interestingly, many stresses appear to induce splicing changes in a small number of defined transcripts, more akin to the ethanol toxicity response than the 3-AT induced amino acid starvation response.

In total, what emerges from our collected assays of the splicing profiles of yeast as they are adapting to changing environmental conditions is a dazzling image of splicing as a remarkably dynamic and flexible mechanism for post-transcriptional regulation.

Why regulate Splicing?

But why regulate gene expression at the level of splicing? What sorts of advantages might splicing regulation confer over the regulation of transcription or other steps in gene expression? In this section I will briefly explore a couple of my favorite models.

Response kinetics

Partially by virtue of our experimental design, all of the examples we have found of splicing regulation have occurred extremely rapidly after shifting cells into new environments. Most are manifest within minutes of the initial shift, and all occur with faster kinetics than the recognized transcriptional responses to those new environments. In the original amino acid starvation dataset, for example, the splicing response can be seen as early as 2 minutes after addition of 3-AT, and the response reaches the peak of its magnitude just as the first signs of activation of the major transcriptional control circuit

appear (in this case, the general control response, which induces transcription of amino acid biosynthesis factors).

In many ways it makes sense that the modulation of splicing can effect changes in the population of cellular mRNAs more rapidly than transcriptional regulation. Transcriptional states are effectively “meta-stable”, for example, often involving complex chromatin re-arrangements and transitions to opened or closed promoter states. In our current view of splicing, individual intron splicing events are one-off occurrences, whether they occur co-transcriptionally or on a free nuclear population of transcripts.

Noise

Another advantage that regulating gene expression at the level of splicing might offer over transcription is that splicing may be less sensitive to cell-to-cell variability, or noise. Due to the inherently stochastic nature of the biochemical interaction between small populations of DNA binding transcription factors and single or double copies of a given gene target in the genomic DNA, activation and repression of transcription of a particular gene in an individual cell is usually “noisy” (Raser and O’Shea, 2005). Experimentally, if the activation state of a given gene in individual cells is observed immediately following a change to a new transcriptional program, a diversity of actual activation states can be observed across a population of cells. Of course different transcriptional circuits exhibit varying levels of noise; if a regulatory mechanism floods the nucleus with an over abundance of a particular transcription factor than the resulting gene activation or repression will be considerably less noisy than a circuit which relies on a much smaller population of effector molecules.

Although noise has not yet been studied in other steps of eukaryotic mRNA processing, the enzymes which carry out these steps are operating on relatively large populations of substrates (as compared to one or two genomic copies of a target promoter region). We would therefore expect post-transcriptional regulation to be far less noisy than transcription. We would expect this to be particularly true at the very onset of fast responses or for responses that need to subtly adjust mature transcript levels. Indeed, we see many examples of both extremely rapid and relatively subtle changes to splicing in our environmental stress response data set. In the case of ribosomal protein gene expression, we see splicing regulation at play when a very fast and relatively subtle change in mature mRNA levels is needed (the response to 3-AT induced amino acid starvation). Although testing this hypothesis requires development of a good in vivo system for testing the splicing state of at least one reporter construct on a cell-by-cell basis, it would be very interesting to quantitatively characterize noise in the splicing responses we have observed as compared to transcriptional noise.

More is better

If we look at the behavior of a single class of transcripts across our entire dataset what we find is that a variety of different types of regulatory control are exerted on them over diverse sets of environmental conditions. In the case of the RPGs, for example, while one stress condition, the 3-AT induced amino acid starvation, transiently inhibits the splicing of nearly all of members of the set, some stresses inhibit or promote the splicing of particular subsets (osmotic stress, cation toxicity), and others dramatically alter their total levels (ostensibly due to a modulation in decay rate; Appendix 1). It seems likely that no one mechanism of regulation would be perfectly suited to every kind of modulation in gene expression, nor would we expect it to be the case that any one mechanism is used to provide a consistent kind of regulation across different contexts.

The unifying principle of gene regulation in the eukaryotic domain may simply be that more is better: the more regulatory circuits that can be applied to the behavior of any given transcript the more combinatorial complexity can be achieved to fine tune control of its expression levels.

How is the regulation of splicing achieved?

The work presented in this thesis is almost exclusively phenomenological. We have provided many answers to the questions of “if” and “when” splicing is regulated in response to the environment in yeast, but we have not even scratched the surface of “how?”

I will briefly discuss three types of answers to this question which I think are compelling grounds for future work.

Accessory factors

Although there is not yet evidence that the yeast SR protein homologues function to modulate splicing of particular yeast introns in response to the environment, as they do in higher eukaryotes to effect alternative splicing decisions, it is quite possible that other accessory factor systems exists which tweak splicing of yeast introns in response to different environmental conditions. Several key properties we might predict for factors that fill this role are: (1) they are not essential for constitutive splicing, (2) they demonstrate physical or genetic interactions with the spliceosome, and (3) loss of function mutations cause condition-specific growth phenotypes. In Appendix 3 we identify a number of factors which meet all of these criteria. In the continuing search for

the mechanisms which lie upstream of the condition-specific splicing phenotypes we have uncovered in this work, these factors represent excellent first candidates to test.

Targets in the core spliceosome

The results in Chapter 1 demonstrate that by artificially tweaking the activities of core components of the spliceosome we observe transcript-specificity in the resulting blocks to splicing. Remarkably this phenomenon appears to be conserved in higher eukaryotes (Park et al., 2004). This raises the provocative possibility that environmental signals could modulate splicing activity by directly affecting the activity of core constituents of the spliceosome. A likely mechanism for modulation would be post translational protein modifications. In fact, there is ample evidence for different kinds of modifications of splicing factors, including ubiquitination and ubiquitin-like modifications (incidentally, processes controlled by many intron-containing genes), but the functional consequences of these types of modifications remain unclear. Given the rapidly increasing sophistication of mass-spectrometric techniques and the ability to identify post-translational protein modifications in complex samples, it is becoming possible to investigate whether or not particular types of modifications to the spliceosome correlate with specific environmental contexts.

Signals in and around the message

Another important question that needs to be answered is what are the *cis* signals in, or associated with, the message itself that allow for transcript-specific regulation of splicing. So far our bioinformatic pursuit of this question has come up relatively empty handed. We have looked for correlations between clusters of environmentally regulated splicing behaviors and transcript splice site sequences, intron size, intron position, pyrimidine tract composition and length, transcript length, and transcript AU-richness, and have

found no statistically convincing patterns. However as we continue to build our library of splicing phenotypes and refine the methodologies we use to group transcripts which exhibit similar splicing behaviors I am optimistic that these types of analysis will become increasingly numerically robust.

A fascinating set of experiments performed by Megan Bergkessel has recently suggested that, at least in the case of the 3-AT induced amino acid starvation response, the determinant of transcript specificity may not reside in the transcript but in its promoter context. Megan has been able to show that placing a promoter associated with an RPG whose splicing is inhibited in response to 3-AT upstream of a transcript whose splicing is unchanged after this treatment is sufficient to confer 3-AT induced splicing regulation on the transcript made from this latter locus. This finding raises the interesting possibility the promoter context can shape the fate of a transcript during post-transcriptional processing, perhaps by altering the dynamics of co-transcriptional splicing.

Regulation of other steps in gene expression

Our focus on splicing as a model for post-transcriptional regulation of gene expression proved fruitful. But there are many other important eukaryotic mRNA processing events to consider. Many of the possible advantages discussed above to regulating gene expression at the level of splicing, as compared to transcription, can be extended to 3'end processing, mRNA transport, mRNA decay, and translational activity, among others. Appendix 2 explores an early look at the ways in which the nuclear mRNA export machinery responds to changing environmental conditions. Here we find an

interesting set of changes in the subcellular localization of several key components of the mRNA export machinery that immediately follow salt stress. It is my hope that these data may serve as the preliminary steps towards investigating the interplay between environmental cues and mRNA export, akin to of our work investigating environmental modulation of splicing.

Over the course of our time working on this project, many other new and interesting examples have emerged of environmental regulation in splicing and other post-transcriptional mRNA processing steps, including 3'end cleavage site choice, mRNA decay mechanics, and translational activity. Indeed, it seems as though we may be entering the golden age of appreciating the dynamic, complex, post-transcriptional life of eukaryotic mRNAs.

References

Ares, M., Grate, L., and Pauling, M.H. (1999). A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* **5**: 1138-1139.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. *RNA* **6**: 1197-1211.

Graveley, B.R. (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends in Genetics* **17**: 100-107.

Izquierdo, J.M. and Valcarcel, J. (2006) A simple principle to explain the evolution of pre-mRNA splicing. *Genes & Dev.* **20**: 1679-1684.

Matlin, A.J., Clark, F., and Smith, C.W. (2005). Understanding alternative splicing: towards a cellular code. *Nature Reviews Molecular Cellular Biology* **6**: 386-398.

Park, J.W., Parisky, K., Celotto, A.M., Reenan, R.A., and Graveley, B.R. (2004) Identification of alternative splicing regulators by RNA interference in *Drosophila*. *Proc Natl Acad Sci* **101**: 15974-15979.

Raser, J.M. and O'Shea, E.K. (2005). Noise in Gene Expression: Origins, Consequences, and Control. *Science* **23**

Schumucker, D., Clemens, J., Shu, H., Worby, C., Xiao, J., Muda, M., Dixon, J. , and Zipursky, S. (2000). *Drosophila* DSCAM is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **92**: 315-326.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs and things. *Cell* **92**: 315-326.

Chapter 1: Transcript Specificity in Yeast Pre-mRNA Splicing
Revealed by Mutations in Core Spliceosomal Components.

Summary

The spliceosome is a large RNA-protein machine responsible for removing the noncoding (intron) sequences that interrupt eukaryotic genes. Nearly everything known about the behavior of this machine has been based on the analysis of only a handful of genes, despite the fact that individual introns vary greatly in both size and sequence. Here we have utilized a microarray-based platform that allows us to simultaneously examine the behavior of all intron-containing genes in the budding yeast *S. cerevisiae*. By systematically examining the effects of individual mutants in the spliceosome on the splicing of all substrates, we have uncovered a surprisingly complex relationship between the spliceosome and its full complement of substrates. Contrary to the idea that the spliceosome engages in “generic” interactions with all intron-containing substrates in the cell, our results show that the identity of the transcript can differentially affect splicing efficiency when the machine is subtly perturbed. We propose that the wild-type spliceosome can also distinguish among its many substrates as external conditions warrant to function as a specific regulator of gene expression.

Abstract

Appropriate expression of most eukaryotic genes requires the removal of introns from their pre-messenger RNAs (pre-mRNAs), a process catalyzed by the spliceosome. In higher eukaryotes a large family of auxiliary factors known as SR proteins can improve the splicing efficiency of transcripts containing suboptimal splice sites by interacting with distinct sequences present in those pre-mRNAs. The yeast *Saccharomyces cerevisiae* lacks functional equivalents of most of these factors; thus, it has been unclear whether the spliceosome could effectively distinguish among transcripts. To address this question, we have used a microarray-based approach to examine the effects of mutations in 18 highly conserved core components of the spliceosomal machinery. The kinetic profiles reveal clear differences in the splicing defects of particular pre-mRNA substrates. Most notably, the behaviors of ribosomal protein gene transcripts are generally distinct from other intron-containing transcripts in response to several spliceosomal mutations. However, dramatically different behaviors can be seen for some pairs of transcripts encoding ribosomal protein gene paralogs, suggesting that the spliceosome can readily distinguish between otherwise highly similar pre-mRNAs. The ability of the spliceosome to distinguish among its different substrates may therefore offer an important opportunity for yeast to regulate gene expression in a transcript-dependent fashion. Given the high level of conservation of core spliceosomal components across eukaryotes, we expect that these results will significantly impact our understanding of how regulated splicing is controlled in higher eukaryotes as well.

Introduction

The coding regions of most eukaryotic genes are interrupted by introns, which must be removed for proper gene expression. The removal of introns requires single-nucleotide precision in order to faithfully convert genomic information into functional protein. The process of intron removal is performed by the spliceosome, a large ribonucleoprotein that catalyzes two sequential transesterification reactions (1,2). The spliceosome itself is highly conserved across all eukaryotes, consisting of five small nuclear RNAs (snRNAs) and well over 100 proteins (3,4). A combination of genetic and biochemical experiments have revealed conformational rearrangements in both RNA and protein components that are required for the spliceosome to accurately process pre-messenger RNA (pre-mRNA) transcripts (5). About 50 of these proteins can be considered core components of the spliceosome in that their activity is required for cell viability in budding yeast, whereas the remainder of the factors can be deleted with little or no effect on cell growth under standard laboratory conditions.

To date, the vast majority of what is known about the mechanism of pre-mRNA splicing has been deduced from experiments using, at most, a handful of transcripts. Yet it remains unknown how well these transcripts represent the behavior of the entire complement of spliceosomal substrates. In higher eukaryotes, where genes are often interrupted by multiple introns, it is known that the spliceosome can utilize specific sequences present in individual transcripts to regulate both quantitative and qualitative aspects of gene expression (6,7). Two large groups of proteins, the SR and hnRNP families, are known to modulate splicing activity, allowing the spliceosome to generate multiple distinct proteins from a single genomic locus, significantly increasing proteomic diversity. In particular, the sequence-specific RNA-binding members of the SR family

improve the processing efficiency of introns containing suboptimal signals at their 5' or 3' splice sites. Binding of the SR proteins to enhancer sequences, which may be located in exons or introns, facilitates recruitment of the core machinery to the suboptimal splice sites.

By comparison to higher eukaryotes, splicing in the yeast *Saccharomyces cerevisiae* appears much simpler in a number of ways. Whereas more than 95% of human genes are interrupted by an intron (8), only about 250 yeast genes, or less than 5% of all genes, contain an intron (9). Moreover, the vast majority of the intron-containing genes in yeast contain only a single intron, differing significantly from humans, where the average gene is interrupted by eight introns (8). As seen in Figure 1A, yeast introns tend to conform to very strong consensus sequences at the branch point and at both the 5' and 3' splice sites. Perhaps accordingly, only three SR-like proteins have been identified in yeast, none of which is known to regulate splicing activity. On the other hand, while relatively few yeast genes are interrupted by introns, those that are tend to be highly expressed and tightly regulated genes. It has been previously shown, for example, that intron-containing genes account for nearly one-third of total cellular transcription (10). Notably, several functional categories are overrepresented among intron-containing genes. The most dramatic of these corresponds to transcripts that encode ribosomal protein genes (RPGs): 102 of the 139 RPGs in yeast contain introns. Interestingly, all yeast introns tend to be small relative to those found in humans; the largest yeast intron contains only 1,002 nucleotides, whereas many human introns stretch to well over 10,000 nucleotides. Nevertheless, there is a bimodal distribution of intron lengths that has a strong correlation with gene function; introns found in genes encoding ribosomal proteins tend to be longer, while those in nonribosomal proteins are generally shorter (Figure 1B).

Given the relative simplicity and similarity of intron-containing transcripts in yeast and the limited role of the SR proteins, it has been a reasonable expectation that the yeast spliceosome would have a restricted capacity to distinguish among its different intron-containing substrates. In this case, the behavior of a few individual transcripts could be extrapolated to accurately describe the behavior of most spliceosomal substrates. The advent of splicing-specific microarrays, which facilitate a simultaneous examination of all intron-containing transcripts (11), allows us to directly ask whether the spliceosome interacts similarly with all pre-mRNAs, or whether transcript-specific differences can be identified. Here we have used this approach to examine the splicing responses of all ~250 intron-containing transcripts resulting from inactivation of 18 different core components of the spliceosome and five factors involved in up- or downstream steps in pre-mRNA processing. Far from being homogeneous in their effects on the different intron-containing transcripts, these mutants reveal a remarkable ability of the core components to differentiate among transcripts. Interestingly, most ribosomal proteins are encoded at two distinct genomic loci in yeast. Whereas the coding sequences are nearly identical at these two locations, the sequences of the introns that interrupt them are generally quite dissimilar. For some of these paralogs, dramatic differences can be seen in response to the different spliceosomal mutations. The capacity of the spliceosome to differentiate among substrates suggests that splicing offers an important opportunity to regulate gene expression in a transcript-specific manner.

Results

Monitoring Genome-Wide Changes in Splicing

To examine genome-wide changes in splicing in yeast, we have designed microarrays similar to those previously described (11) that allow us to distinguish between the spliced and unspliced isoforms of transcripts derived from intron-containing genes. As seen in Figure 2A, three different oligos are used to target each intron-containing gene (see also Materials and Methods). One oligo targets a portion of the last exon and is used to detect changes in total transcript level (probe T). A second oligo targets a region of the intron, allowing us to detect changes in pre-mRNA levels (probe P), while the third oligo targets the junction of the two exons, allowing for a determination of changes in levels of mature mRNA (probe M).

Because there are multiple, independent probes describing the behavior of each intron-containing gene, analysis of a splicing-specific microarray is inherently more complicated than that of a standard expression array. In order to assess the splicing behavior of any given transcript it is important to simultaneously examine the changes in signal on all three feature types. In analyzing data from splicing-specific microarrays, others have generated splicing indices for each gene by compressing the data derived from the different feature types. For example, splicing changes can be represented as a precursor/mature (PM) index by dividing the ratio of the movement of the pre-mRNA feature by the ratio of the mature mRNA feature (11,12). Although this manipulation simplifies the data, allowing them to be more easily examined in the context of traditional microarray analyses, for reasons described below we find that a more detailed

understanding of individual transcript behavior is achieved by considering the independent movement of all three feature types simultaneously.

Figure 2B and Figure 2C show an example of such an analysis derived from a comparison of RNA from a wild-type strain and a strain harboring the temperature-sensitive *prp2-1* mutation after both were shifted to the nonpermissive temperature of 37 °C. The spliceosomal Prp2 protein is a member of the DEAD/H box family of RNA-dependent ATPases and plays an essential role in catalyzing an as yet unidentified structural rearrangement required for activation of the spliceosome prior to the first chemical step (13). As such, the primary result of a defect in Prp2 activity is expected to be a broad decrease in pre-mRNA splicing. The splicing profile resulting from this experiment, where each horizontal line describes the behavior of a single gene, is displayed for both a single time point (Figure 2B) and across the time course of inactivation (Figure 2C). A simple prediction for the behavior of a given transcript in an experiment where the activity of a core spliceosomal component is impaired is that the precursor species would accumulate with a concomitant loss of the mature species. Indeed, for many of the transcripts, including those shown with a red bar in Figure 2C, this is precisely the pattern that is observed: precursor species are seen to accumulate to nearly ten times the normal level, and up to 90% of the mature mRNA is lost. For most of these genes, the behavior of the total mRNA probe closely follows that of the mature species, suggesting that the majority of the transcripts for these genes are present in their spliced forms. However, different profiles are apparent for many other genes. For example, many show robust accumulation of precursor species with very little decrease in mature mRNA, including those shown with a green bar in Figure 2C.

The different behaviors displayed by these two groups of transcripts presumably reveal important differences in the complex interplay of cellular machineries that define the steady-state levels of both the precursor and mature forms of each of these transcripts. For example, the rate of precursor accumulation observed for a given transcript depends upon, among other things, the transcription rate of that gene, the degradation rate of the transcript, the rate at which the transcript is normally spliced, and the extent to which its splicing is inhibited. Likewise, the decrease in mature mRNA is a function of both the extent of splicing inhibition and the decay rate of the mature form of that transcript. For those genes marked by the red bar in Figure 2C, the dramatic decrease in levels of mature mRNA may be explained by either a strong block to the production of mature species, a rapid rate of decay of the mature species, or some combination of both. By comparison, a less complete block to splicing or a slower rate of mRNA decay may explain the failure to detect significant decreases in mRNA for those genes marked with the green bar. Importantly, the process of creating splicing indices to describe the behavior of these transcripts in many cases eliminates the ability to differentiate between these scenarios. For example, genes which show both a modest increase in precursor levels and a modest decrease in mature mRNA levels generate nearly identical PM index values as do genes that show a strong precursor accumulation but little decrease in mature levels.

By simultaneously examining each of the feature types as described above, the global splicing profile resulting from the *prp2-1* mutation shows that precursors rapidly accumulate for about 80% of the intron-containing genes. While the detailed behaviors of the different transcripts are varied, this result appears to satisfy the simple prediction that defects in this factor would result in defective splicing of all actively transcribed genes. At least two different scenarios may explain the failure to detect precursor

accumulation for the remaining 20% of the transcripts. It is almost certainly true that some of the intron-containing genes are transcriptionally quiescent during logarithmic growth conditions. Naturally, no precursor accumulation would be expected for such genes. Alternatively, these transcripts may be insensitive to the particular defect introduced by the *prp2-1* mutation.

Transcript Specificity of Spliceosomal Mutations

In order to further distinguish between transcript behaviors, we examined the RNA from strains containing mutations in two additional core spliceosomal components: *PRP5* and *PRP8*. Like *PRP2*, *PRP5* encodes an essential member of the DEAD/H box family of RNA-dependent ATPases. By comparison with Prp2, the activity of Prp5 is required to promote a molecular rearrangement earlier in the splicing pathway, catalyzing the stable addition of the U2 small nuclear ribonucleoprotein to the branch point sequence (14). As with mutations in Prp2, loss-of-function mutations in Prp5 result in a block to splicing prior to completion of the first chemical step. Notably, Prp5 and Prp2 are only transiently associated with the spliceosome, whereas Prp8 is a stable component of the U5 small nuclear ribonucleoprotein and is thought to form the physical core of the spliceosome (15,16). As with Prp5 and Prp2, most mutations in Prp8 block spliceosomal activity prior to the first chemical step; however, mutants have been isolated that can perform the first but not the second transesterification reaction (17,18).

As seen in Figure 3A, strains containing the loss-of-function mutations *prp2-1*, *prp8-1*, or *prp5-1* all exhibit conditional growth phenotypes, showing nearly wild-type growth at 25 °C but an inability to support growth at 37 °C. A comparison of growth at intermediate temperatures allows these mutations to be ordered according to the strength of the

defect they impart: *prp2-1* is the most severe defect as it is unable to support growth above 25 °C; *prp8-1* is slightly less severe as it can support weak growth at 30 °C; and *prp5-1* is the weakest as it shows nearly wild-type growth even at 33 °C. Figure 3B shows the time-resolved splicing profiles derived from shifting each of these strains from 25 °C to 37 °C and comparing them to a similarly treated wild-type strain. Importantly, the ordering of the genes is consistent for all three profiles, meaning that the behavior of a single intron-containing gene can be seen by following a single horizontal line across all three profiles.

As expected, inactivation of each of these factors results in inhibition of splicing of a large number of intron-containing transcripts. Interestingly, while the overlapping set of transcripts that are affected by inactivation of all three factors is quite large, a closer examination suggested important differences in the particular transcripts whose splicing was affected and prompted us to more carefully compare the profiles of these mutants. Microarray experiments were performed that directly compared the pairwise behaviors of these mutant strains (Figure 3C), revealing three distinct classes of transcript behavior. The first class is composed of transcripts for which splicing is equally affected by each of these mutations. Approximately 100 transcripts behave in this fashion, exemplified by those indicated with a red bar in Figure 3B and C and highlighted in Figure 3E. For these transcripts, the comparison of mutant versus wild-type behavior seen in Figure 3B shows a strong splicing defect for all three mutations. Further, the direct comparisons of the mutants in Figure 3C show that each of the mutants causes a nearly identical level of precursor accumulation and mature mRNA reduction. Interestingly, found within this class is the transcript encoding Act1, which is widely used as a reporter for both in vivo and in vitro splicing assays. Figure 8 shows the results of quantitative RT-PCR experiments validating a subset of these findings.

The second class is also composed of transcripts that exhibit a canonical splicing defect in response to inactivation of each of the single mutants, but for whom the magnitude of the defect is different for each of the mutants. About 100 transcripts behave in this fashion, exemplified by those indicated with a green bar in Figure 3B and C and highlighted in Figure 3F. As with the first class, the experiments shown in Figure 3B demonstrate that the splicing of each of these transcripts is affected by each of the mutations; however, the direct comparisons of the mutants in Figure 3C show a larger precursor accumulation in the *prp2-1* strain than in either the *prp8-1* or *prp5-1* strains, and a larger accumulation in the *prp8-1* strain than in the *prp5-1* strain. Interestingly, the molecular splicing defects observed for this class of transcripts correlate with the conditional growth defects: the *prp2-1* mutation produces the strongest phenotype, followed by the *prp8-1* mutation, and finally the *prp5-1* mutation. Notably, as indicated in Figure 3D, the vast majority of the transcripts in this class encode RPGs, and most RPGs belong to this class.

The final class is composed of a smaller number of transcripts, those that exhibit splicing inhibition in response to one or two of the mutant strains but not all three. Examples of several transcripts belonging to this class are shown in Figure 3G. Interestingly, the splicing of the Rps30a transcript is dramatically affected by the *prp2-1* mutation, but shows little defect in response to either the *prp8-1* or *prp5-1* mutations. Conversely, the splicing of the Rpl19b transcript is affected by both the *prp8-1* and *prp5-1* mutations, but is largely unaffected by the *prp2-1* mutation. Likewise, the Hnt1 transcript shows little defect in response to the *prp8-1* mutation, while splicing of the Cox4 transcript is strongly affected by the *prp5-1* mutation but is only mildly affected in the *prp2-1* strain, the inverse of the ordering seen for the second class of transcripts described above.

Given the predominance of RPG transcripts in the second class of transcripts, we sought to identify properties of these transcripts that might distinguish them from the others. As an initial attempt to identify such differences, we asked whether RPG transcripts showed a difference in splicing efficiency relative to other transcripts under standard laboratory growth conditions in wild-type cells. Quantitative RT-PCR was performed using primers specific to both intron and exon regions of 12 different intron-containing genes. Using genomic DNA to generate standard curves, relative copy numbers of both the precursor and total mRNA species were then calculated and compared. As seen in Table 1, the RPG transcripts tend to show exceptionally high levels of splicing efficiency, particularly relative to non-RPG transcripts. The relative copy numbers for the total mRNA features is in good agreement with previously published mRNA abundance values (19). Interestingly, while the mature RPG transcripts are quite abundant, there is in fact less detectable precursor species for most RPGs than for non-RPGs. Nevertheless, as seen in Figure 9, the ability to detect large increases in precursor levels on our microarrays is independent of the initial level of precursor mRNA.

The differences that we observed between mutations in different spliceosomal factors led us to ask whether distinct mutations in a single factor might also cause such transcript-specific responses. Shown in Figure 4A and B are the time-resolved splicing profiles obtained from shifting strains harboring either the *prp8-1* mutation or the temperature-sensitive *prp8-101* mutation, respectively, to the nonpermissive temperature and compared to similarly treated wild-type strains. Strikingly, the splicing of most intron-containing transcripts is almost completely unaffected by the *prp8-101* mutation. Instead, this mutation appears to affect the splicing of only a small number of transcripts. As it had previously been shown that the *prp8-101* mutation results in a

defect in catalyzing the second transesterification reaction (18), we examined a defect in another factor necessary for the second step, the helicase Prp16 (20). Figure 4C shows the time-resolved splicing profile derived from shifting a strain harboring the cold-sensitive *prp16-302* mutation to the nonpermissive temperature of 16 °C. The broad splicing defect seen for this mutant suggests that the differences between the *prp8-1* and *prp8-101* splicing profiles are not simply the result of comparing mutants defective for the first or second chemical steps, but rather truly reflect allele-specific differences in substrate specificity.

A Systematic Examination of Splicing Factors

The transcript-specific phenotypes that we observed with this small subset of conditional mutants compelled us to examine the effects of a much wider variety of spliceosomal defects. To facilitate these experiments, a set of high-throughput methods was developed that allowed us to automate many of the steps in a microarray experiment (see Materials and Methods). Using these methods, we examined the effects of 18 different mutations in core spliceosomal components and five additional mutations in up- or downstream processes in mRNA processing. Figure 10 shows the different mutants that were examined using these methods, the step in pre-mRNA processing at which they are presumed to be defective, and their conditional growth phenotype. Two different views of the data resulting from these experiments are presented in Figure 5 and Figure 11. By clustering the data across all of the experiments, Figure 5 tends to highlight those pre-mRNAs for which splicing is negatively affected by most of the mutants. By comparison, the clusters derived from each individual mutant shown in Figure 11 allow a more detailed analysis of the behavior of the pre-mRNAs in response to each individual mutation. As expected, many of the spliceosomal mutations result in splicing defects

over a broad range of transcripts. Interestingly, the splicing profile resulting from inactivation of the *brr5-1* mutant is quite similar to that seen for many other canonical splicing mutants. While this mutant was originally isolated as a splicing mutant (21), it has been subsequently shown that Brr5 functions during 3' end processing and is the yeast homolog of mammalian CPSF-73 (22,23). By comparison, the *rna14-64* mutant, which is also defective in 3' end formation (38), produces a noticeably different phenotype.

Importantly, a careful examination of the response of individual transcripts to this panel of mutations makes clear the fact that all transcripts are not equally affected by mutations in core components. To exemplify this point, Figure 6 shows the behavior of a select group of transcripts in response to all of the mutations examined. For some transcripts, such as the U3 small nucleolar RNA, splicing is rapidly blocked by nearly all spliceosomal mutations, while remaining largely unaffected by mutations in other mRNA processing factors. Interestingly, while the *prp19-1* mutation does affect the splicing of many pre-mRNAs, it has little effect on the splicing of this transcript. A comparison of the two U3 paralogs, *SNR17A* and *SNR17B*, whose mature products are nearly identical but whose introns share little sequence homology, also shows some differences in their responses to some of the splicing mutants. For example, the U3a transcript shows a stronger defect in response to the *prp5-1* mutation than the U3b transcript.

For some transcripts, the subset of mutations that affect their splicing is much different. A dramatic example of this is provided by the *RPS30* paralogs. Whereas splicing of the Rps30b transcript is affected by most of the splicing mutants examined, the Rps30a transcript is not. The splicing of the Rps30a transcript shows a strong defect in response to the *prp2-1* and *prp16-302* mutations, a somewhat weaker response to the *prp28-1*,

brr2-1, *brr1-1*, and *brr5-1* mutations, and very little defect in most of the other mutants. The *RPL19* paralogs also display divergent behaviors: whereas the Rpl19a transcript shows a similar defect in response to each of the mutants, the Rpl19b transcript shows strong defects in response to the *prp8-1*, *prp5-1*, and *prp4-1* mutants, but less severe defects in response to the other mutants. Differences can also be seen between non-RPG transcripts by comparing the behavior of the Act1 and Tef4 transcripts, for example.

Discussion

Here we describe the results of genome-wide experiments designed to examine the *in vivo* responses of all intron-containing transcripts in *S. cerevisiae* to mutations in core spliceosomal factors. For largely technical reasons, most experiments designed to study the efficiency of pre-mRNA splicing to date have focused on a relatively small number of transcripts, leaving open the question of how much variety there is among different spliceosomal substrates. The experiments presented here overcome this limitation by utilizing splicing-specific microarrays that allow for the simultaneous examination of all spliceosomal substrates. Splicing-specific microarrays are particularly amenable to studies in budding yeast, where there are a limited number of intron-containing genes and there is very little alternative splicing. By using this approach to examine the time-resolved effects of inactivating 23 different factors involved in pre-mRNA processing, the work presented here reveals a complex relationship between the activity of the core spliceosome and the full complement of transcripts with which it must interact.

RPG Transcripts versus Non-RPG Transcripts

A large collection of data such as this offers a unique opportunity to both compare and contrast the behaviors of all 250 intron-containing transcripts according to their individual responses to the different spliceosomal mutations. From this perspective, the detailed comparison of mutations in *PRP2*, *PRP8*, and *PRP5* shown in Figure 3 provides important information about the different spliceosomal substrates, and suggests that all transcripts are not equally affected by particular spliceosomal mutations. Each of these three spliceosomal proteins plays an important role in catalyzing events necessary for the first chemical step in splicing. As such, mutations in these factors that reduce their activity at this step can be expected to negatively affect the splicing efficiency of all actively transcribed intron-containing genes. Indeed, while the majority of pre-mRNAs show such a defect in response to the *prp2-1*, *prp8-1*, and *prp5-1* mutations, the strengths of the splicing defects imparted by these mutations are variable. Whereas the splicing of many transcripts encoding factors other than ribosomal proteins is equally inhibited by mutations in any of these three factors, the strength of the defect observed in the splicing of most ribosomal protein gene transcripts is dependent upon the particular defect in the spliceosome. For these mutations, the strength of the molecular splicing defect correlates with the temperature sensitivity that the mutation imparts upon the strain: *prp2-1* causes the strongest defect and *prp5-1* causes the weakest. While the fraction of new transcript from any single gene whose splicing is inhibited cannot be determined from these experiments alone, these results nevertheless demonstrate that a greater fraction of most RPG transcripts is blocked by the *prp2-1* mutation than by either the *prp8-1* or *prp5-1* mutations. This difference is unlikely to simply reflect differences in the severity or onset of the protein mutations because nearly identical fractions of newly transcribed Act1, for example, are blocked by all three of the spliceosomal mutations.

Rather, it suggests that the RPG-encoding transcripts are fundamentally different from the other transcripts in their susceptibility to these mutations.

We have not yet been able to identify any single feature of intron-containing RPGs that explains this altered susceptibility. While it is true that RPG introns tend to be longer than non-RPG introns, and that RPGs tend to be among the more highly transcribed genes in the genome, neither of these features alone seems sufficient to explain the different responses. The Act1 transcript, for example, is interrupted by an intron that is similar in length (308 nucleotides) to an average RPG intron (mean length of 402 nucleotides), and is transcribed at a similar rate as many RPGs (24), yet its behavior is distinct from that of most of the RPGs.

In an effort to identify properties that might differentiate RPG transcripts from other intron-containing transcripts, we used quantitative RT-PCR to determine the amount of precursor and mature mRNA species present in wild-type cells for a variety of different intron-containing genes. While all 12 of the transcripts that we examined were efficiently spliced (>80%), the RPG transcripts showed remarkably high partitioning toward mature species. It is noteworthy that, in spite of the high levels of total transcript present for most RPGs, the amount of precursor species present is remarkably low relative to the other intron-containing transcripts. For example, while there is more than ten times as much total Rps21b transcript in a cell as there is total Nmd2, 50 times less Rps21b pre-mRNA can be detected as compared to Nmd2 pre-mRNA.

It is tempting to speculate that the different splicing efficiencies and the different susceptibilities to spliceosomal mutations displayed by the RPGs may be mechanistically related. For example, while much recent progress has been made in

identifying protein components of the spliceosome (4), it remains unknown whether the entire set of components is required for splicing of all transcripts, or whether different transcripts will utilize the activities of different subsets of factors. Given that the promoter elements driving transcription of RPGs are often different from non-RPGs, it is easy to imagine that the complement of proteins present during their splicing could also be different. The low levels of precursor species detected under steady-state conditions for the RPG transcripts suggest that the time between transcription and splicing is shorter for these transcripts than for non-RPG transcripts. In this light, it will be important to determine whether unique spliceosomal complexes are present during the splicing of RPG transcripts that allow for enhanced catalytic rates of either the first or second chemical steps of splicing. Alternatively, the rate of cotranscriptional loading of spliceosomal components onto RPG transcripts could be enhanced. In this case, genome-wide chromatin immunoprecipitation experiments might be expected to show higher levels of spliceosomal association with the RPGs than the non-RPGs. In either case, the presence or absence of unique, auxiliary spliceosomal components on the RPG transcripts might also change the susceptibility of these transcripts to defects in the core spliceosomal components.

All RPGs Are Not Equal

Importantly, not all RPG transcripts behave similarly in these experiments. Whether considering the simple comparisons of the *prp2-1*, *prp8-1*, and *prp5-1* mutations, or the entire panel of factors as shown in Figure 5 and Figure 6, several RPG transcripts show significantly different behaviors from the others. Compelling examples of these differences can be seen among certain pairs of RPG paralogs. While many paralogs display similar defects in response to the mutations, some pairs show dramatically

different responses to these mutants (Figure 6). The *RPS30* paralogs offer the most dramatic example of these differences. Whereas splicing of the Rps30b transcript is adversely affected by almost all of the spliceosomal mutations that we examined, splicing of the Rps30a transcript is only affected by a small subset of these mutants. The particular features unique to each of these transcripts that underlie this differentiation remain unclear. Presumably the ability of the Rps30a transcript, for example, to be efficiently spliced in the *prp8-1* mutant does not indicate that Prp8 function is unnecessary for the splicing of this transcript. Rather, it suggests that the unique set of interactions that this transcript forms with the spliceosome are less sensitive to the defect imparted by this particular mutation in Prp8. Importantly, because the Rps30a and Rps30b transcripts behave so differently in response to so many mutations, they should prove to be extremely useful as tools for more traditional biochemical experiments designed to understand the function of individual spliceosomal factors.

Deciphering Protein Function from Genome-Wide Experiments

Most of the mutants that we have examined here were isolated from conditional lethal screens. As such, this subset of mutations may be biased towards those with strong growth defects and hence broad transcript specificity. Perhaps because of this, we have found it difficult to associate a particular mechanistic role with any of these factors based simply on the subset of pre-mRNAs whose splicing is affected in these experiments. Rather, our findings seem to indicate that the signature of introns that are affected by a particular spliceosomal mutation is unique to that mutation. Nevertheless, it is tempting to imagine isolating other mutations in these core components that affect the splicing of a smaller subset of transcripts. Such mutations may be more difficult to isolate, as they may be aphenotypic for growth under standard laboratory conditions. Nevertheless, a

genome-wide splicing analysis of such a mutant may provide important insights into the activity of the protein. Indeed, for the *prp8-101* mutation, which appears to represent such a mutant, knowing the identity of transcripts that are either strongly affected or unaffected provides important information for designing more traditional biochemical experiments in the future to examine the mechanism of its activity in the spliceosome. Likewise, the genome-wide splicing profiles derived from similar mutations in other factors might prove more revealing about the mechanistic role of those factors.

Implications

Interestingly, the capacity of core spliceosomal components to elicit transcript-specific changes in splicing activity does not appear to be limited to yeast. Recent RNAi-mediated depletion experiments targeting core spliceosomal components in *Drosophila melanogaster* show differential effects on splicing activity at alternative splice sites (25). Given the level of conservation of the core spliceosomal components across eukaryotes, it seems highly likely that they could also play important roles in augmenting the regulatory roles of the SR and hnRNP families of proteins in higher eukaryotes. The surprising finding that the macular degenerative disease retinitis pigmentosa is associated with mutations in core components of the spliceosome may in fact reflect such a role (26–28). As seen here for the *prp8-101* mutation, the mutations associated with retinitis pigmentosa may cause defects in the splicing of a distinct group of transcripts, presumably uniquely associated with retinal function.

We have recently shown that the yeast spliceosome can rapidly and specifically alter the splicing efficiency of distinct subsets of transcripts in response to at least two unrelated environmental stresses (J.A.P., G.B.W., M.B. and C.G., unpublished data). Notably, in

response to amino acid starvation, the splicing of virtually all RPG transcripts is specifically downregulated. Given that *S. cerevisiae* lacks the large SR and hnRNP protein families responsible for modulating transcript-specific splicing activity in higher eukaryotes, the observations presented in the current work suggest that the core spliceosomal components themselves could be targets of environmental regulation. As seen with stable genetic modifications here, post-translational modifications of these components might similarly facilitate transcript-specific regulation of pre-mRNA splicing. Such modifications would allow the cell to both rapidly and specifically regulate the production of translatable mRNA for particular transcripts in a reversible manner. Whether or not the complex relationships between transcripts and core spliceosomal components revealed by these mutant studies represent features of a system that has evolved to be the target of biological regulation is a provocative question that remains to be fully addressed.

The work presented here does however suggest that the relationship between the spliceosome and the full complement of transcripts with which it must interact is much more complex than previously believed. Important questions for the future concern both the mechanism by which the spliceosome distinguishes among these substrates and the biological rationale for this specificity. The simple hypothesis that all mutations in splicing factors that inhibit the growth of a yeast cell would also inhibit the splicing of all (expressed) intron-containing transcripts is inconsistent with our observations. The splicing efficiency of a given intron can not yet be ascribed to obvious *cis* features of the intron, nor does the biochemical activity presumed to be affected by a mutation appear to dictate which introns will be affected. Instead, in much the same way that transcription is now known to be regulated by the combinatorial control of its holoenzyme, it appears

that the efficiency of pre-mRNA splicing may be dependent upon the complexes present in the spliceosome as well as the particular transcript upon which it assembles.

Materials and Methods

Sample collection and preparation

Unless otherwise indicated, cultures were grown according to standard techniques in rich medium supplemented with 2% glucose at 25 °C. Both mutant strains and the corresponding wild-type strain were grown in parallel, 100 ml cultures until their optical densities were between $A_{600} = 0.5$ and $A_{600} = 0.7$. An initial 15 ml sample was collected at 25 °C prior to initiation of the time course. Cells were collected by filtration using Millipore HAWP0025 filters (<http://www.millipore.com>). The filters were immediately frozen in N₂(l). The culture flasks were then transferred into water baths at either 16 °C or 37 °C, with additional 15 ml aliquots removed and collected at the appropriate times. Total cellular RNA was isolated as previously described (29) with a few exceptions. Tubes containing PhaseLock gel (Eppendorf, <http://www.eppendorf.com>) were used during phase separation steps. Also, RNA samples were precipitated using isopropanol.

For each microarray sample, cDNA was prepared from 25 µg of total RNA in a 50 µl reaction mixture containing 50 mM TrisHCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 0.3 mM TTP, 0.2 mM aa-dUTP, 12.5 µg dN₉ primer, and 5 ng murine Moloney leukemia virus (M-MLV) RT. Primers were annealed to the RNA by heating to 65 °C for 5 min in the presence of buffer and salt alone. Reactions were allowed to incubate at 42 °C for at least 2 h. Remaining RNA was

then hydrolyzed by incubation for 15 min at 65 °C in the presence of 0.1 M NaOH and 10 mM EDTA. This was neutralized with HCl, then purified using a Zymo25 DNA purification column according to the manufacturer's protocol (<http://www.zymoresearch.com>). Typical yields ranged from 10 to 15 µg of cDNA from 25 µg of starting material. The purified cDNA was then conjugated to the appropriate fluorescent dye in a 10 µl reaction containing 50 mM sodium bicarbonate [pH 9.0], 50% DMSO, and □20 µg of NHS-derivatized fluorophore. Reactions were incubated in the dark at 60 °C for 60 min, after which time the cDNA was repurified using a Zymo25 DNA purification column.

High-throughput sample collection and preparation

Cultures were grown in sets of four mutant strains along with their corresponding wild-type strains in rich medium in flasks at 25 °C until their optical densities were between $A_{600} = 0.5$ and $A_{600} = 0.7$. After the strains were synchronized, the cultures were transferred into the wells of a 96-well growth plate at 25 °C. A reverse time course was then initiated using a multichannel pipettor at the appropriate times to transfer the cultures into a second 96-well plate submerged in a 37 °C water bath (Figure 12). The entire time course was then collected by centrifugation of the plate at 5,000g for 5 min. The cell pellets were then frozen in N₂(l). With the help of a Biomek FX liquid handling system (Beckman Coulter, <http://www.beckmancoulter.com>), total cellular RNA was then isolated using largely standard procedures. Typical yields from a single 1.8 ml well ranged from 15 µg to 25 µg of total RNA. For each well, cDNA synthesis was performed in a 50 µl reaction as described above. After cDNA synthesis, the remaining RNA was hydrolyzed by addition of 25 µl of 0.3 M NaOH/0.03 M EDTA and heating to 60 °C for 15 min. This solution was neutralized by addition of 25 µL of 0.3 M HCl. To purify the cDNA from this solution, 0.5 ml of cDNA binding buffer (5 M GdnHCl, 30% isopropanol, 60 mM

potassium acetate, 90 mM acetic acid) was added to each well. This material was passed over a 96-well glass-fiber filtration plate (Nalgene Nunc, <http://www.nalgenenunc.com>). The filters were then washed once with 0.5 ml of Wash I (10 mM TrisHCl [pH 8.0], 80% EtOH), and once with 0.5 ml of Wash II [80% EtOH]. Purified cDNA was then eluted in two steps of 100 μ l water. The appropriate dyes were then coupled to the cDNA as described above. Fluorescently labeled cDNA was then repurified using 96-well glass-fiber filtration plates as described above.

Quantitative PCR

Total cellular RNA for quantitative RT-PCR experiments was prepared as described above. Prior to conversion to cDNA, the RNA was treated with DNase I (Fermentas, <http://www.fermentas.com>) according to the manufacturer's protocol. The cDNA was then produced from 2 μ g of total RNA in a 20 μ l reaction mixture containing 50 mM TrisHCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 0.5 mM each dNTP, 25 nM each gene-specific primer, 2.5 μ g dN₉ primer, and 5 ng M-MLV RT. Quantitative PCR was then performed using an Opticon from MJ Research (<http://www.mjresearch.com>). Primer sets used in this study are shown in Table 2. Relative copy numbers shown in Table 1 are the result of triplicate measurements from a single biological sample. Standard deviations were only slightly higher when comparing biological replicates. A series of 10-fold dilutions of genomic DNA covering a total range of 10⁶ molecules was used to generate the standard curves. Genomic DNA was purified using a ZR Fungal DNA kit (Zymo Research) according to the manufacturer's protocol. The copy numbers presented in the table have been normalized to the least abundant species detected in these experiments, the precursor isoform of the RPS21b transcript. Importantly, these copy numbers do not represent cellular copy numbers.

Microarray design and construction

In an initial set of experiments, we examined a variety of probe designs for their ability to specifically and efficiently identify the mature mRNA species derived from intron-containing genes. Whereas other splicing-specific microarrays have utilized fixed-length probes that are geometrically centered about the exon–exon junction, our best results were obtained when the probe was thermodynamically balanced between the two sides of the junction. Accordingly, oligonucleotides were designed for every exon–exon junction such that the hybridization energies were equivalent on either side of the junction (Figure 13). Because of the different GC contents in the transcripts, the junction oligos range in length from 28 nucleotides to 40 nucleotides.

Total mRNA feature probes (expression oligos) and premature mRNA feature probes (intron oligos) 32 nucleotides in length were designed using a previously described algorithm (30). Intron and junction oligos were designed to target 256 intron features in 246 different transcripts. Expression oligos were designed to target the last exon of these 246 intron-containing transcripts, as well as 689 non–intron-containing transcripts and 99 noncoding RNAs (including snRNAs and tRNAs).

Probes were printed on poly-lysine-coated glass slides using standard techniques. Each probe was print on the array three times by each of two different printing pins (for a total of six replicate probe spots for each feature on each array; Figure 14). Prior to use, arrays were preprocessed by 60 s of incubation at 25 °C in 3× SSC, 0.2% SDS. Arrays were then blocked using standard procedures. Mutant and wild-type samples were then

competitively hybridized against one another using the Agilent Hybridization Buffer (Agilent, <http://www.agilent.com>) at 60 °C for a period ranging from 12 to 16 h.

Image analysis, data preprocessing, and higher order analysis

Microarray images were acquired using an Axon Instruments GenePix 4000B scanner, reading at wavelengths of 635nm and 532nm (Axon Instruments, <http://www.axon.com>). Image analysis was performed using Axon Instruments GenePix Pro version 4.0 (number 2500–137, Rev E, 2001). Ratio values derived from the median pixel intensities for the 635 nm and 532 nm images of each spot were used to represent probe behaviors in data preprocessing analysis. A standardized qualitative assessment of array quality was performed using the Bioconductor arrayQuality package (Paquet AC, Yang YH, “arrayQuality: Assessing array quality on spotted arrays,” version 1.4.0) (31,32). Arrays were manually removed from further analysis if they showed an above average spot intensity to ratio bias or a strong spatial ratio bias (Figure 15). Spot ratio data were log₂ transformed and normalized within each array using global loess regression implemented in the Bioconductor marray package (33). All experiments were performed as dye-flipped pairs using biological replicates. Both within-array and between-array data replication were analyzed for data quality (Figure 16) using the Bioconductor limma package (34). Linear analyses of the data were performed using correlations of within-array and between-array spot replicates to assess evidence for differential expression of each RNA species across all of the experimental factors (35). Results from these linear analyses pertaining to conclusions highlighted in the text are shown in Table 2 and Table 3.

For downstream hierarchical clustering, feature ratio values were calculated by averaging the median within-array spot replicate measures between replicate arrays. Where appropriate, replicate variation was used to weight feature measurements. Hierarchical clustering was performed using the C Clustering Library version 1.32 (36). Data were clustered using average linkage and Pearson correlation as the distance measure.

All microarray data are available at the Gene Expression Omnibus (GEO) repository at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/geo>).

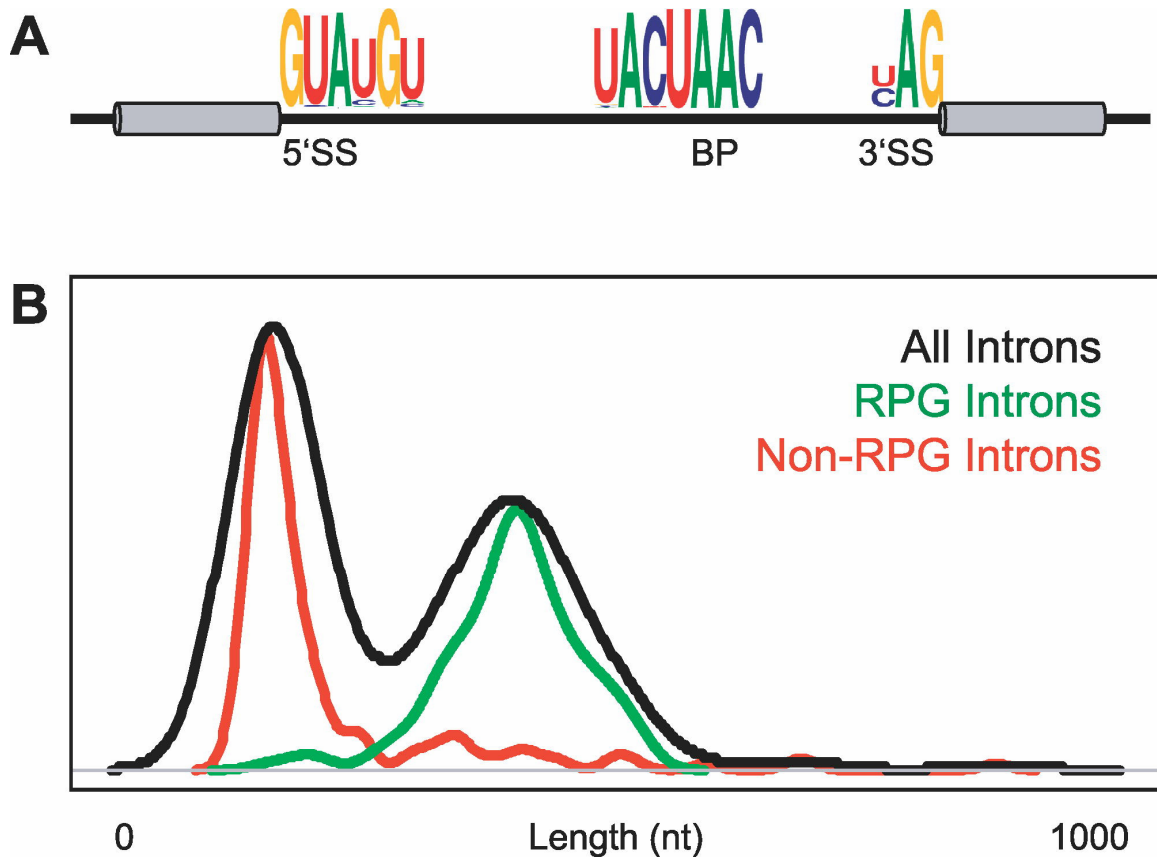


Figure 1. Properties of yeast introns.

(A) Consensus sequences found in yeast introns at 5' splice sites (5'SS), 3' splice sites (3'SS), and branch points (BP) (37). (B) Distribution of yeast intron lengths shown for all introns (black), introns only in ribosomal protein genes (green), or introns only in nonribosomal protein genes (red).

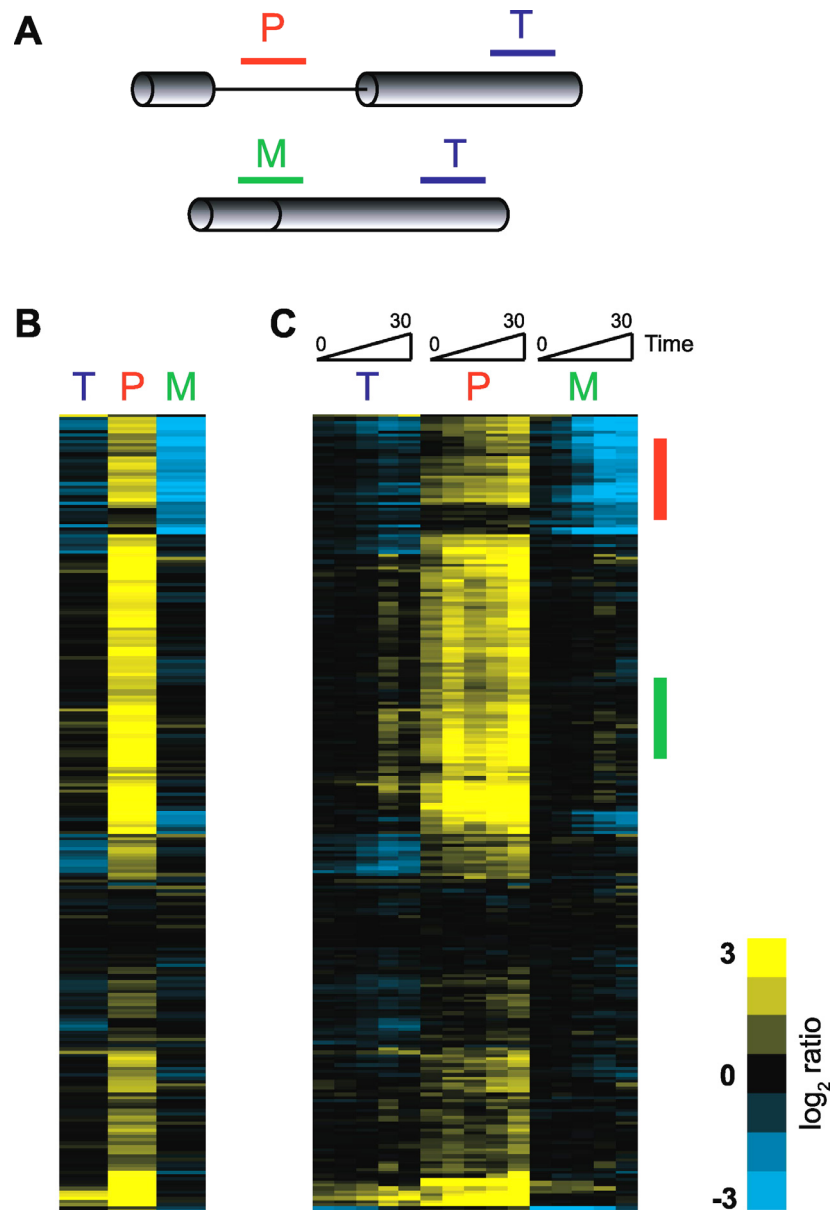


Figure 2. Monitoring genome-wide changes in pre-mRNA splicing.

(A) Oligos target either the pre-mRNA (P), the mature mRNA (M), or the total level of mRNA (T) for each intron-containing gene. (B) False color representation of the splicing response of 249 different intron-containing transcripts at a single timepoint after inactivation of Prp2 activity. Each horizontal line describes the behavior of a single transcript. (C) Time course from 0 to 30 min showing the kinetic response to inactivation of Prp2. The gene order is identical to that shown in (B).

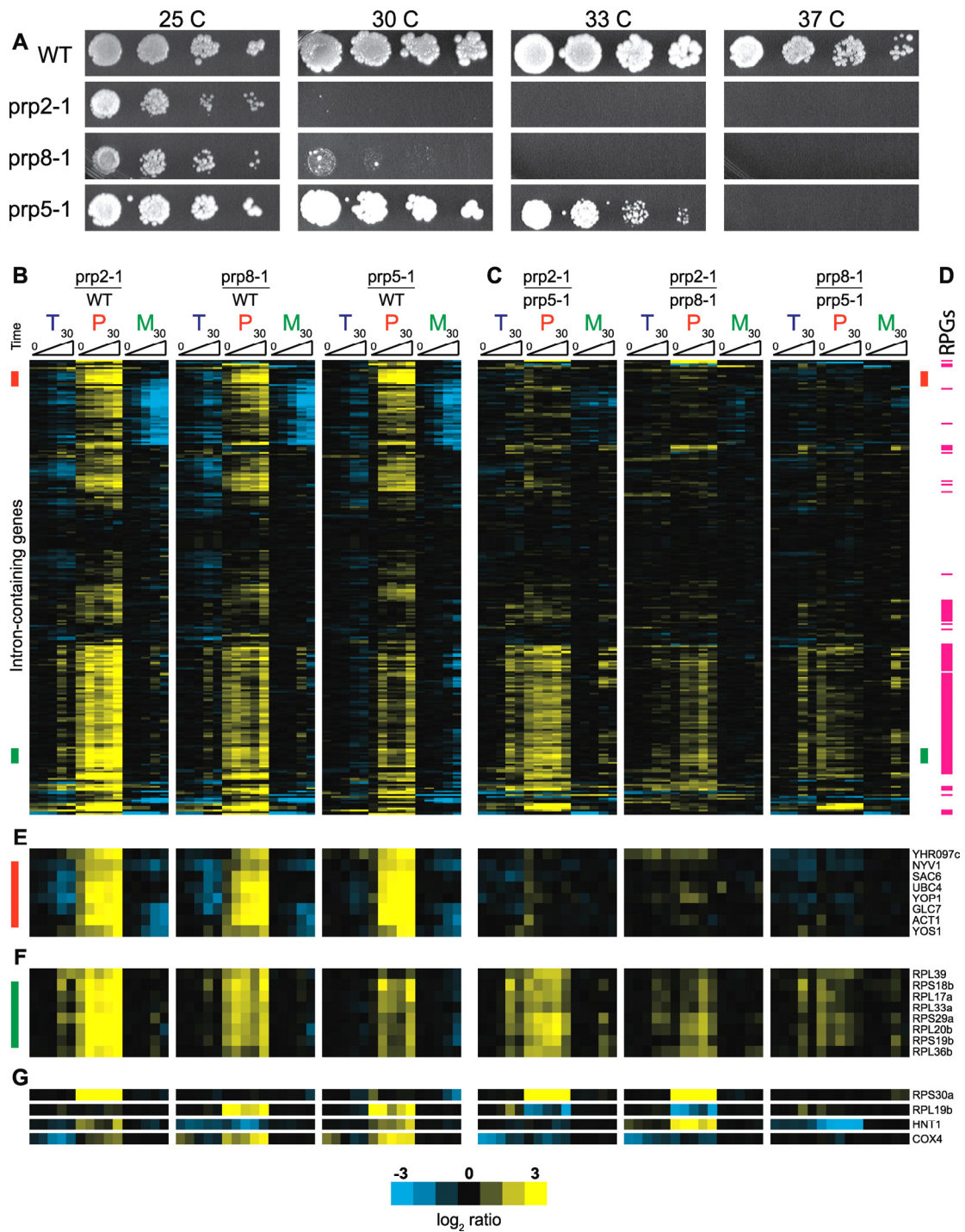


Figure 3. Comparing the cellular and molecular phenotypes of three different spliceosomal mutants.

(A) Serial dilutions of a wild-type strain and strains containing the *prp2-1*, *prp8-1*, and *prp5-1* mutations are grown at the temperatures indicated. (B) Time-resolved splicing

profiles for each of the three mutant strains compared to a wild-type strain. (C) Time-resolved profiles resulting from microarrays directly comparing the indicated mutants. The order of genes is identical to that in (B). (D) Transcripts that encode ribosomal protein genes are indicated with a red line. The order of genes is identical to that in (B) and (C). (E) The identification and behavior of a subset of genes indicated with a green bar in (B) and (C). For these transcripts the comparison experiments reveal a different level of precursor accumulation for each of the mutants. (F) The identification and behavior of a subset of genes indicated with a red bar in (B) and (C). For these transcripts the comparison experiments show an identical splicing defect for all three mutants. (G) The identification and behaviors of a subset of genes whose splicing is affected by only one or two of the three mutants studied.

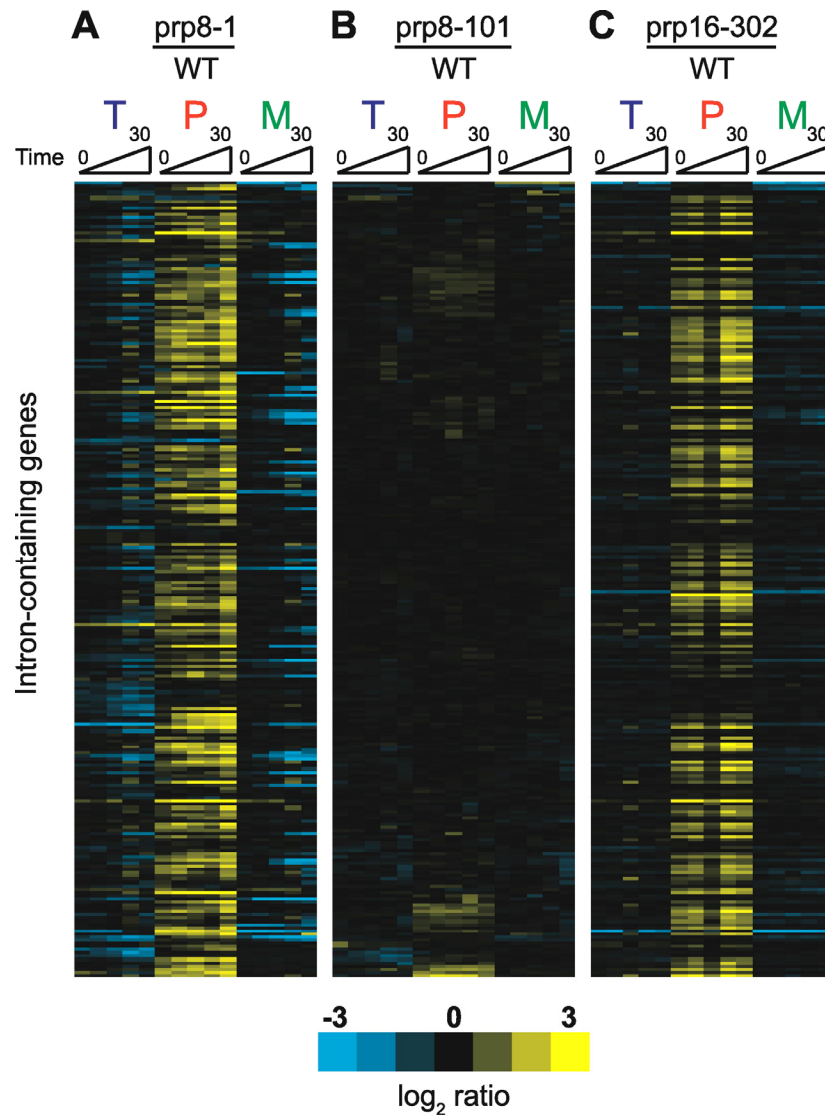


Figure 4. Two different alleles of Prp8 produce very different molecular phenotypes.

Time-resolved splicing profiles derived from comparisons of *prp8-1* (A), *prp8-101* (B), or *prp16-302* (C), each compared to a wild-type strain. The order of the genes is the same for all three profiles.

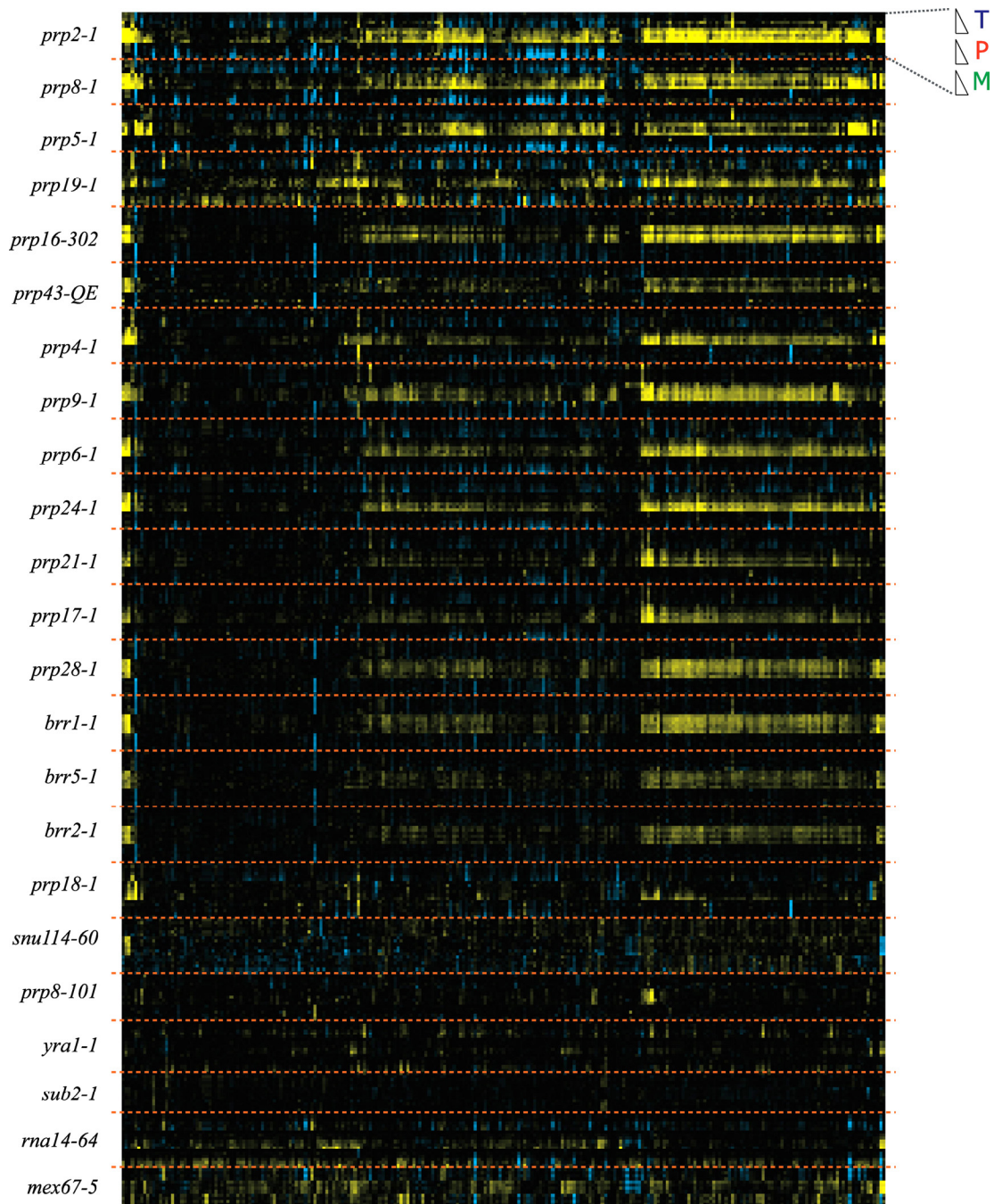


Figure 5. A global view of defects in pre-mRNA splicing.

Time-resolved splicing profiles derived from a panel of factors involved in pre-mRNA processing. The behavior of each individual transcript can be seen by following a vertical line. The particular mutations examined are shown on the left. For each mutant, the order is indicated at the top right corner of the figure.

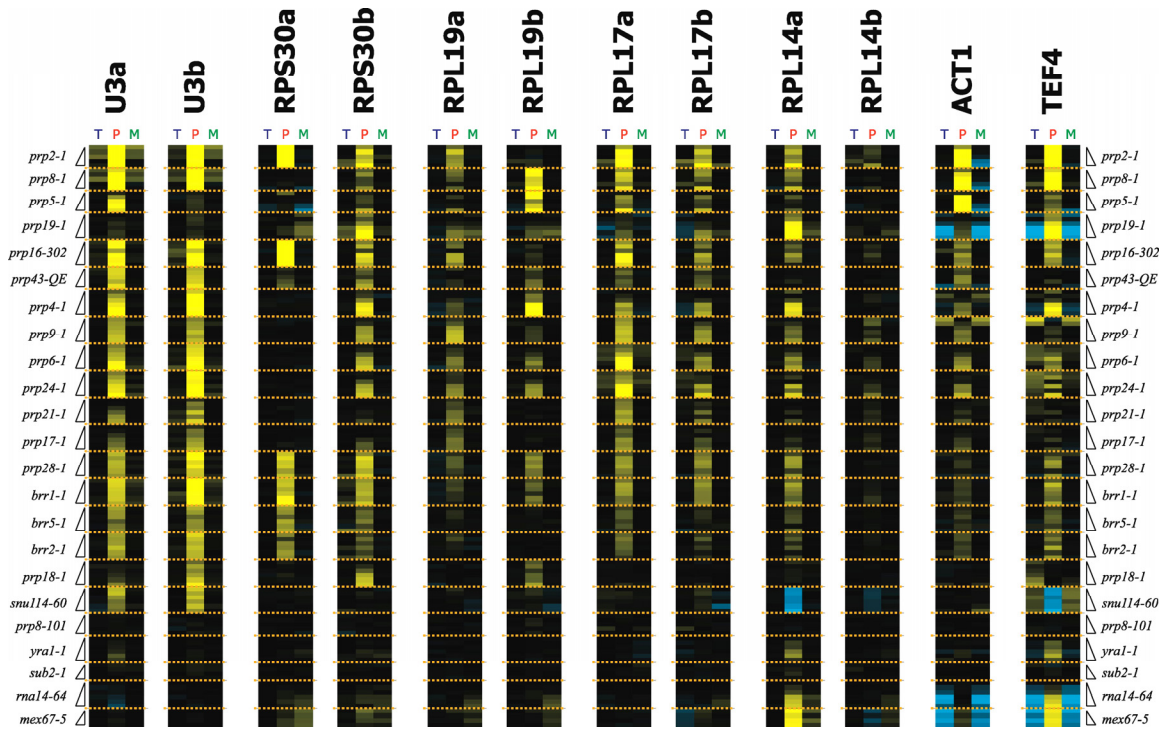


Figure 6. Transcript-specific view of defects in mRNA splicing.

Time-resolved splicing profiles for individual transcripts derived from the same panel of factors shown in Figure 5.

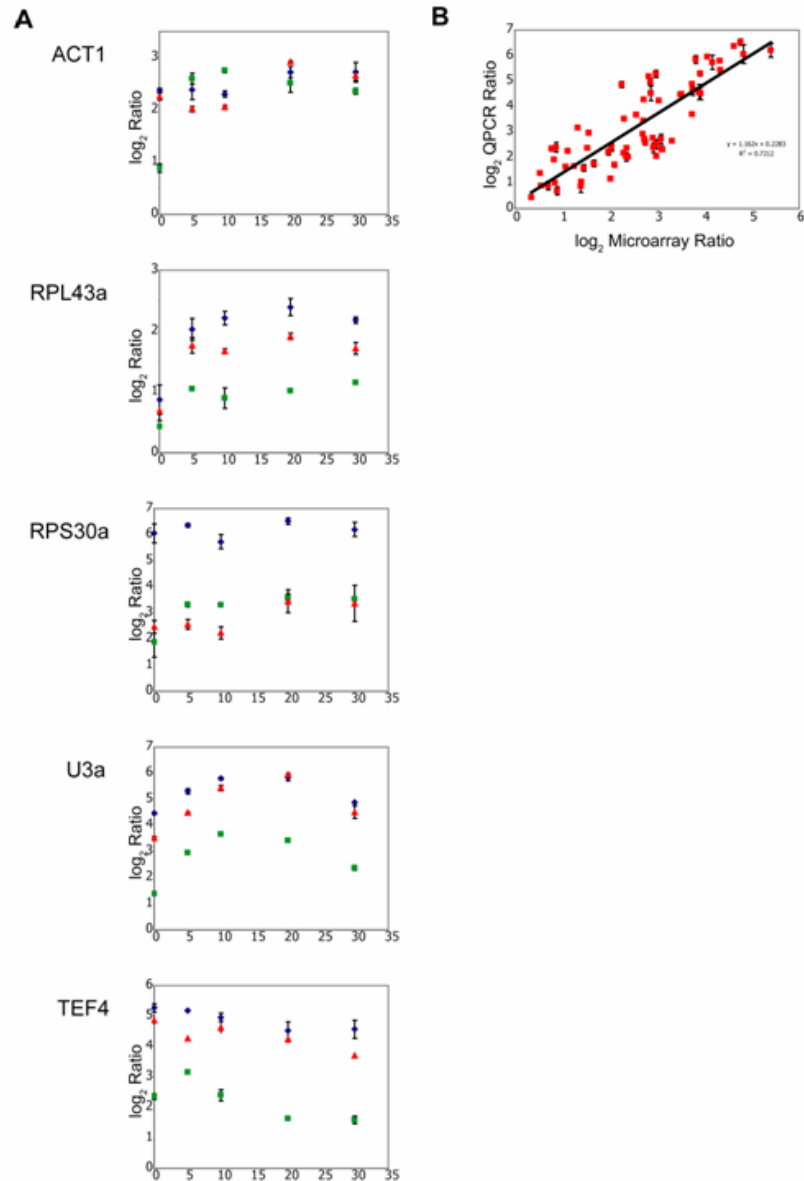


Figure 8. Validation of microarrays using quantitative RT-PCR.

(A) Quantitative RT-PCR (QPCR) experiments comparing wild-type samples with *prp2-1* (blue), *prp8-1* (red), or *prp5-1* (green) samples. Input material was normalized using the average value obtained from probes targeting both *NRG1* and *ECO1* transcripts. Values shown are derived from triplicate measurements with error bars included. (B) Comparison of QPCR and microarray values for genes shown in (A).

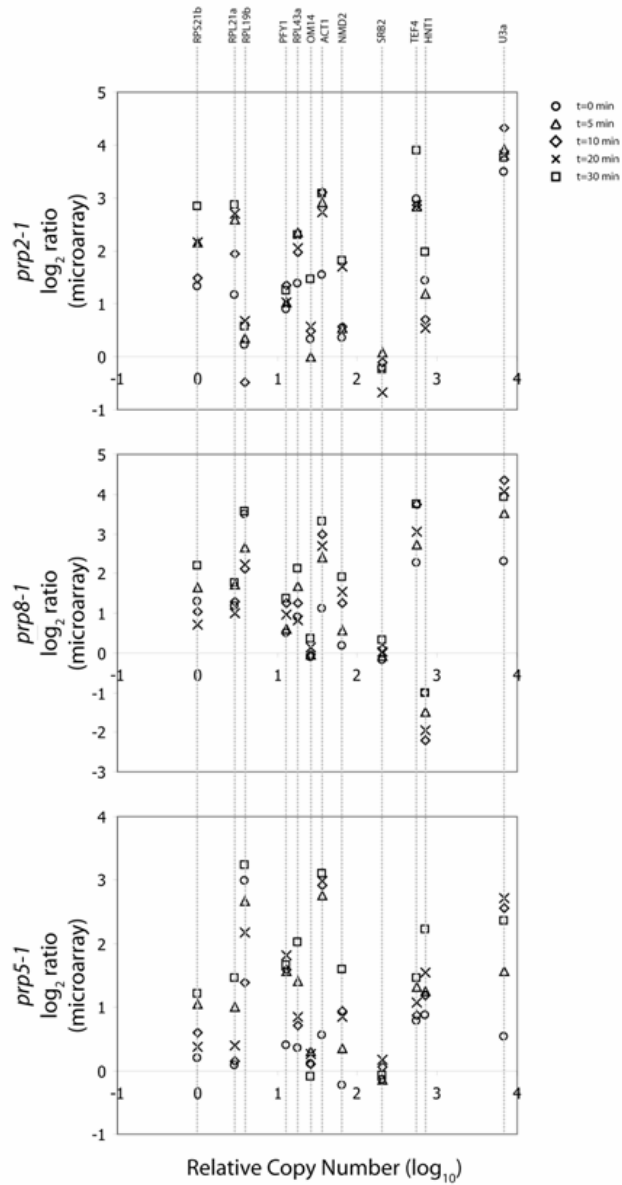


Figure 9. The extent of precursor accumulation is independent of initial precursor levels.

Ratio values corresponding to the precursor species of the indicated genes (top) derived from each of three mutant analyses are plotted against initial precursor values as determined in Table 1.

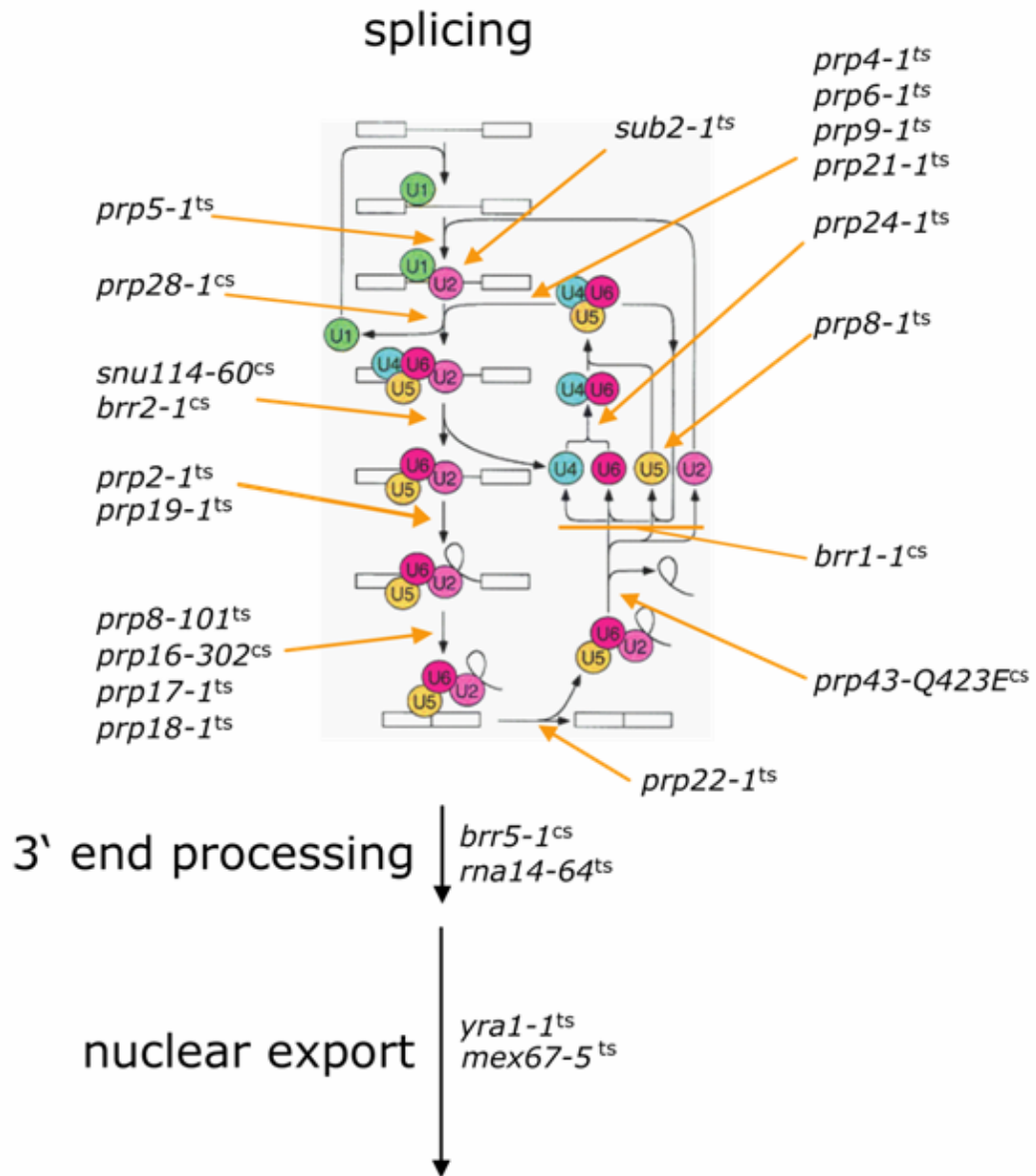


Figure 10. Splicing mutants examined.

Flow diagram of splicing and spliceosomal recycling pathways showing putative stage of action of each interrogated allele. Alleles in factors involved in 3' end processing and mRNA nuclear export were also tested in this study.

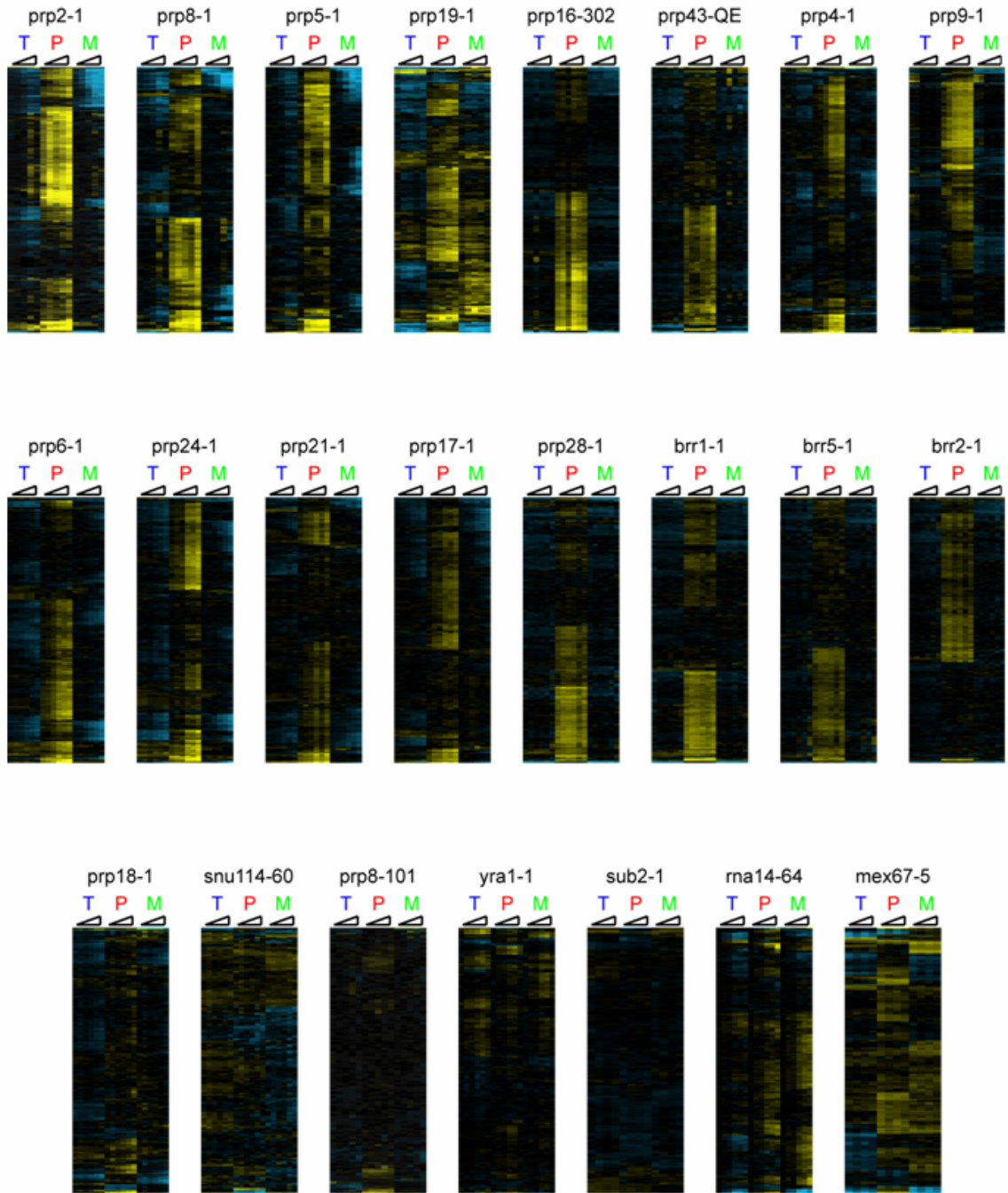


Figure 11. Individual splicing profiles of each mutant examined.

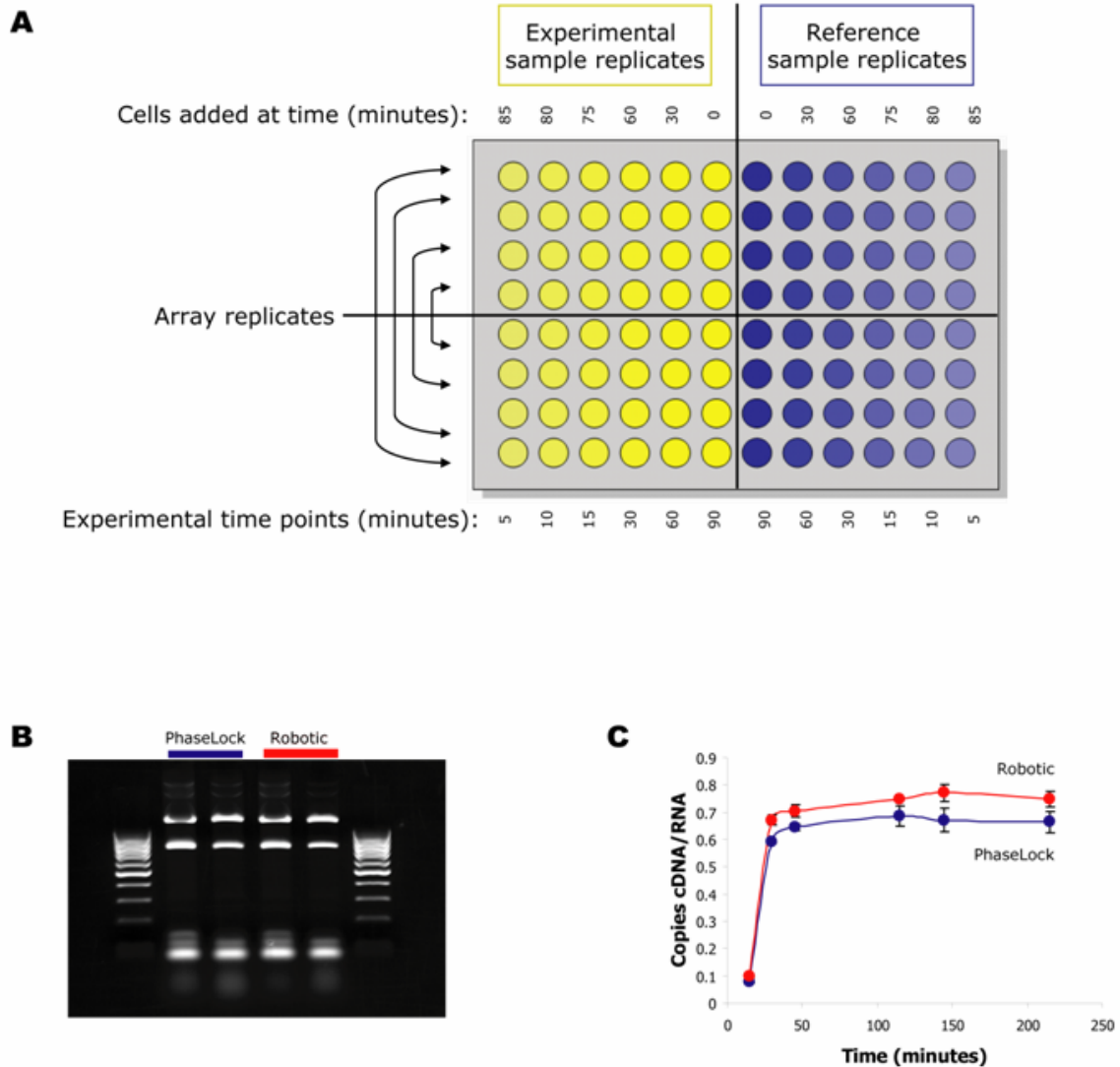


Figure 12. High-throughput sample collection and preparation.

(A) Experimental time courses of temperature shifts were performed in a 96-well format. Samples were transferred from flasks at 25 °C to a 96-well plate shaking in a 37 °C bath in a reverse time course. Matched wild-type and mutant samples were collected in the same 96-well plate to ensure consistency in time points between hybridized samples (right half and left half of the plate, respectively). Samples were collected for two biological dye-flipped array replicates in each 96-well plate (top half and bottom half of the plate). (B) Replicate samples of total RNA from wild-type cells grown to optical densities between $A_{600} = 0.5$ and $A_{600} = 0.7$ at 30 °C purified using either the standard or

the high-throughput method (see Materials and Methods) are shown here run on a 1% agarose gel stained with ethidium bromide. (C) RT efficiency of total RNA samples shown in (B).

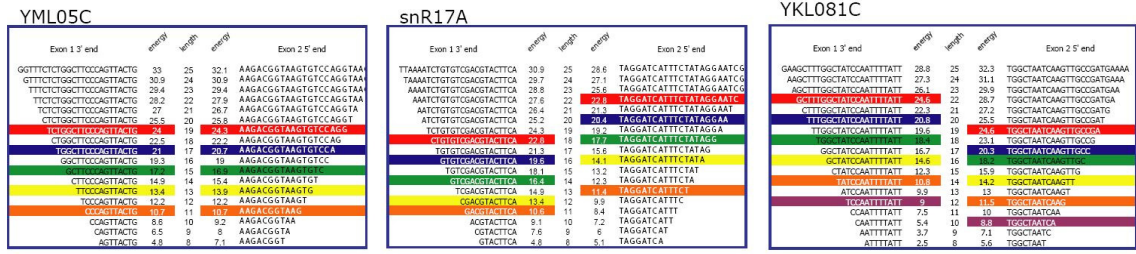


Figure 13. Design scheme for mature mRNA probes.

To design mature mRNA probes (see Figure 2, probe M), a range of possible target-binding energies between probe sequence and exon sequence flanking exon-exon junctions were experimentally tested to determine what binding energy would yield the highest signal to noise. Sample probe sequences for three of the 12 tested targets are shown here.

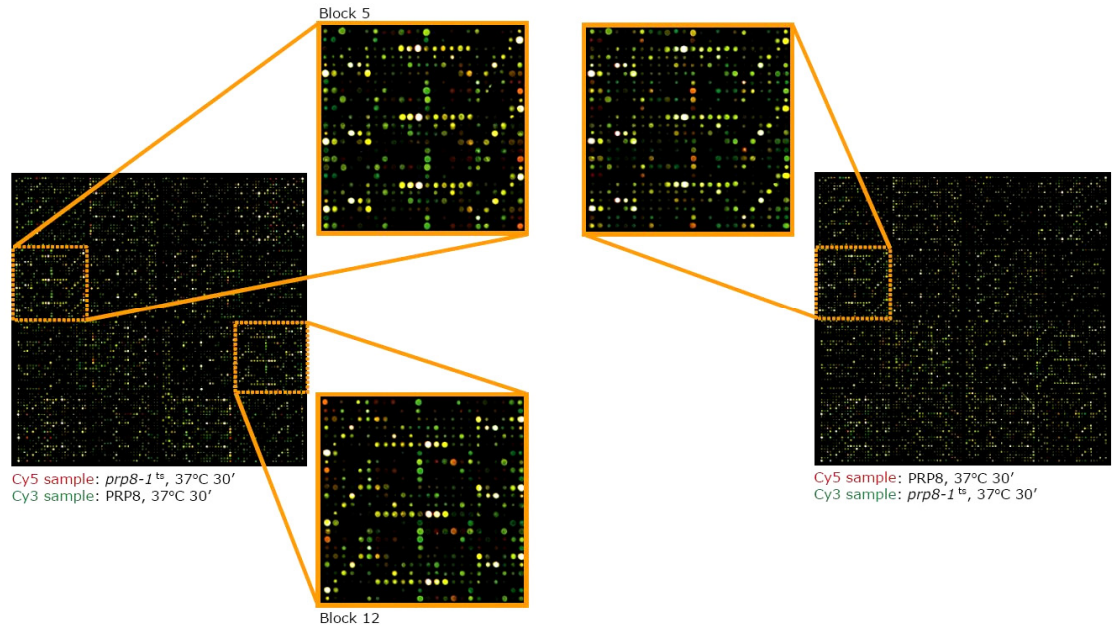


Figure 14. Microarray design.

Each biological sample is analyzed with three layers of data replication. Each probe is printed on the arrays three times by each of two different printing pins. Within a block (printed by a single printing pin) the three spot replicates are printed in the same column, separated vertically by eight rows. This vertical separation allows for assessment of top-to-bottom spatial spot ratio and intensity bias within each block. Block replicates (printed by two different pins) are oriented on the array with 180° rotational symmetry (for example, block 5 and 12). Horizontal and vertical separation of blocks allows for between-block assessment of both top-to-bottom and left-to-right spatial spot ratio and intensity bias. Finally, each experiment is assayed on two dye-flipped biological replicate arrays.

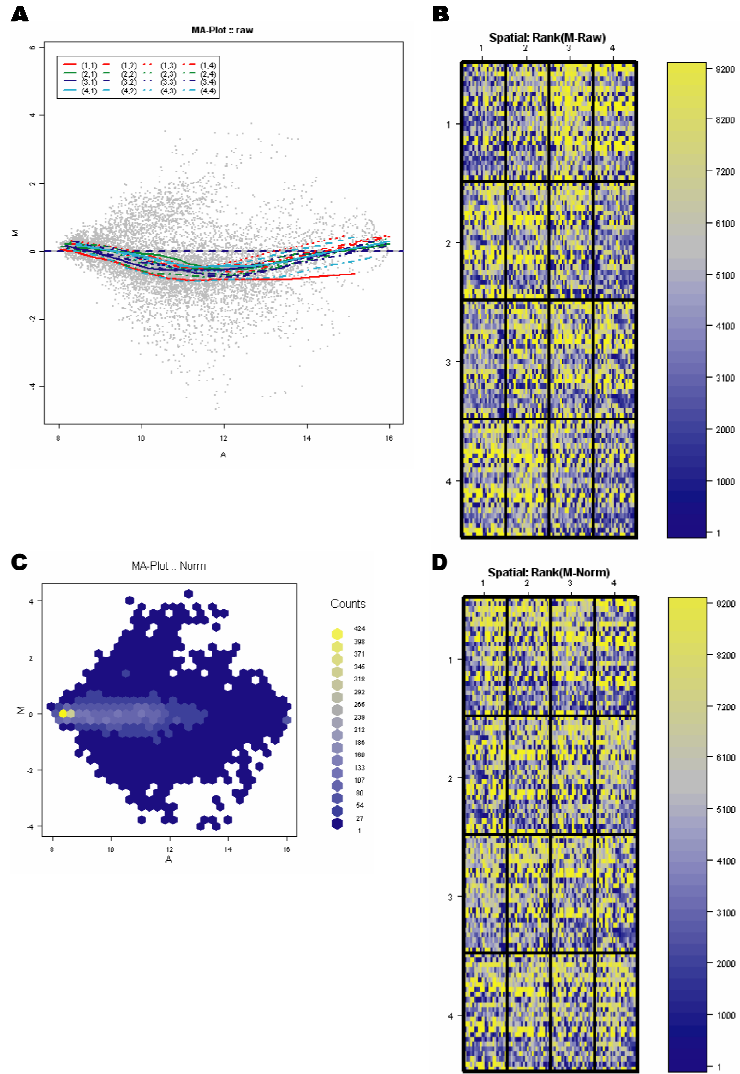


Figure 15. Global quality assessment for an individual microarray.

An example of a standard qualitative assessment of array quality, with array data from the *prp8-1* versus wild-type 30-min 37 °C temperature shift. (A) Pre-normalization spot log₂ ratios (M-values) compared to total spot intensities (A-values), with traces showing the average behavior of each of the 16 blocks corresponding to individual printing pins. (B) Spatial visualization of pre-normalization M-value ranks across the array. (C) Hexagon binning plot showing distribution and density of global Loess normalized M-values versus A-values. (D) Spatial visualization of global Loess normalized M-values across the array.

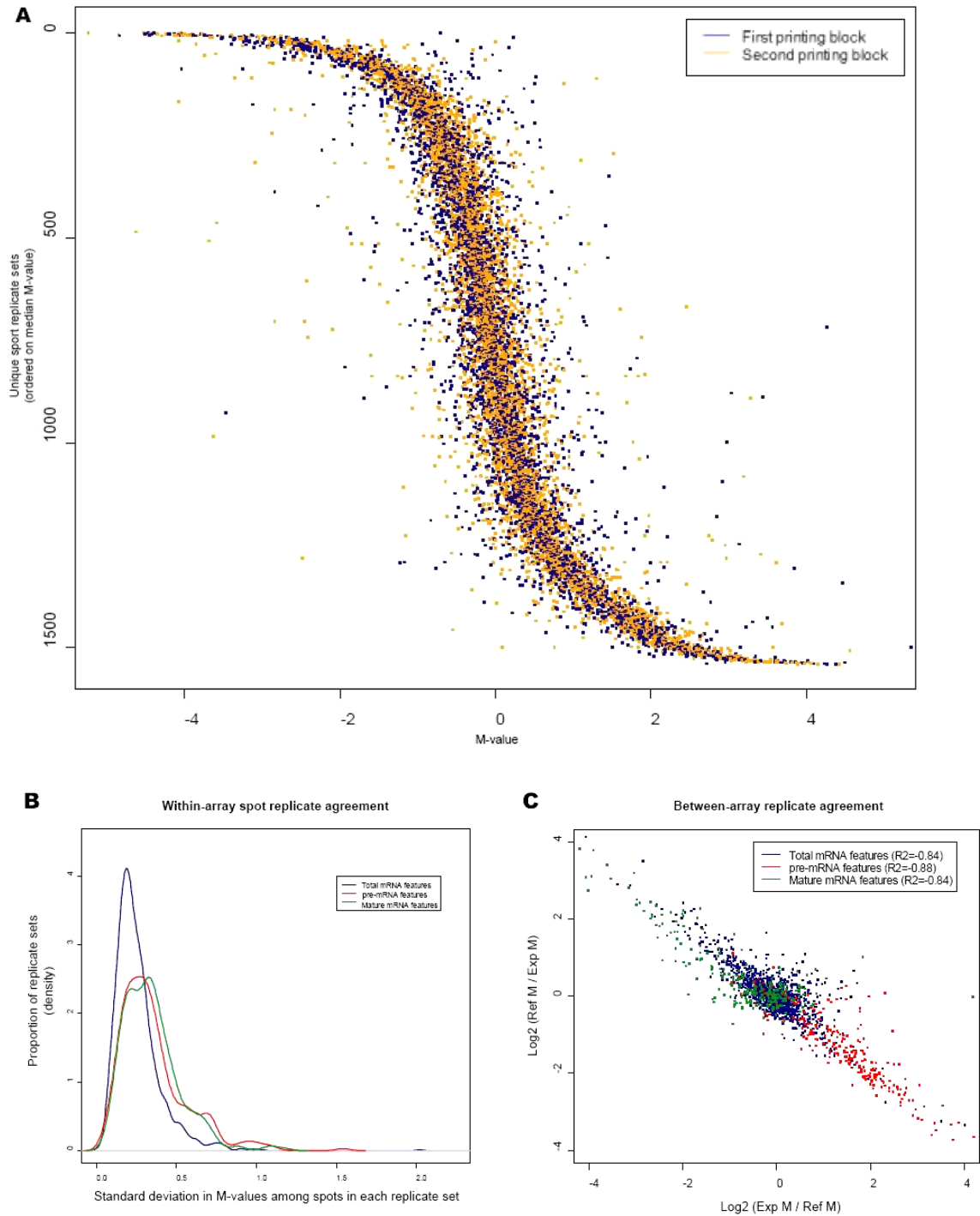


Figure 16. Quality assessment of within-array and between-array replicate agreement.

(A) Dotplot of normalized, unfiltered, log₂ ratios (M-values) for every probe on the *prp8-1* versus wild-type 30-min 37 °C temperature shift array. Each set of six replicate probes is

plotted on a single line vertically. M-values for probes printed in the first set of printing blocks (blocks 1–8) are plotted in blue, while M-values for probes printed in the second set of printing blocks (blocks 9–16) are plotted in orange. Spot replicate sets are arranged vertically based on the median M-value of each set. (B) Density trace (Gaussian smoothing) of standard deviations for each replicate set on the array shown in (A). Three different density traces are shown for the three different types of probe sets. (C) Plot of normalized M-values on the *prp8-1* versus wild-type 30-min 37 °C temperature shift array (Exp/Ref) and the matched dye-flipped replicate array (Ref/Exp). Points corresponding to total mRNA features, pre-mRNA features, and mature mRNA features are plotted in blue, red, and green, respectively. Legend shows Pearson correlations between the values for each of these types of features in the two arrays.

ORF	Gene	Intron Copies	Exon Copies	Efficiency	M/P Ratio
YJL136c	<i>RPS21b</i>	1 ± 0.1	36,000 ± 1,500	>99.9%	36,000
YBL027w	<i>RPL19b</i>	4 ± 1	110,000 ± 2,900	>99.9%	28,000
YPR043w	<i>RPL43a</i>	18 ± 1.4	100,000 ± 7,200	>99.9%	5,500
YBR191w	<i>RPL21a</i>	3 ± 1	15,000 ± 900	>99.9%	5,000
YOR122c	<i>PFY1</i>	13 ± 2	36,000 ± 3,100	>99.9%	2,800
YFL039c	<i>ACT1</i>	37 ± 8.8	17,000 ± 1,700	99.8%	450
YKL081w	<i>TEF4</i>	560 ± 39	78,000 ± 15,000	99.3%	140
YBR230c	<i>OM14</i>	26 ± 1.8	2,900 ± 680	99.1%	110
YHR077c	<i>NMD2</i>	65 ± 5.5	3,000 ± 510	97.8%	46
snR17a	<i>U3a</i>	6,800 ± 690	270,000 ± 20,000	97.5%	40
YDL125c	<i>HNT1</i>	710 ± 120	5,300 ± 710	86.6%	7.5
YHR041c	<i>SRB2</i>	210 ± 7.5	1,200 ± 100	82.2%	5.6

Table 1. Relative copy numbers of total and pre-mRNA determined by quantitative RT-PCR

Gene	Feature	ForwardPrimer	ReversePrimer
YBR191w	Exon	TGCCGTCCATCTTTCTACTT	TAATCTTTTTTCTAGATATC
YBR191w	Intron	AAACGGTCTCTGCGCCTCTG	CTGTAGATTCATAGGATTAA
YKL081w	Exon	AGGTTTGATTCCATACAACA	AAGAACCAGCGGCATGAAGG
YKL081w	Intron	TAAACTAGATGTCAAGATCG	GCCAAATAAACGAACGGGAA
YJL177w	Exon	AATTGGTTGTTACCGAAAAG	CAATTCTACCTCTTTGTCTA
YJL177w	Intron	AAAGAGCCATCCCATTGAGA	CTCCCGTCAACACATCATT
YFL039c	Exon	TCTTCCCATCTATCGTCGGT	TCCATATCGTCCCAGTTGGT
YFL039c	Intron	CCTTCCCCTTTCTACTCAA	TAAATGGGATGGTGCAAGCG
snR17a	Exon	TAGGATCATTTCTATAGGAATCG	ATAGATGGCCGAACCGCTAAG
snR17a	Intron	TATGTAATATACCCCAAACATTTACCC	AAACTCTAGTACCTAAGACTTTTAGATG

Table 2. Oligonucleotide sequences used in QPCR experiments.

ORF Name	M	t-statistic	p-value	Adjusted p-value	B-statistic
YJL001W	-3.9886	-9.0127	2.01E-19	5.09E-17	33.3086
YOL048C	-3.8762	-8.7585	1.98E-18	2.50E-16	31.0846
YBL040C	-3.6922	-8.3429	7.25E-17	6.11E-15	27.5849
YIL111W	-3.6067	-8.1497	3.65E-16	2.31E-14	26.0154
YGL087C	-3.4090	-7.7029	1.33E-14	6.73E-13	22.5282
YML067C	-3.3184	-7.4982	6.47E-14	2.73E-12	20.9958
YDR305C	-3.2750	-7.4001	1.36E-13	4.92E-12	20.2763
YGR001C_2	-3.1192	-7.0480	1.81E-12	5.74E-11	17.7715
YBL059C-A	-3.0470	-6.8851	5.78E-12	1.62E-10	16.6533
YDL137W	-2.9444	-6.6532	2.87E-11	7.26E-10	15.1078
YGR183C	-2.9071	-6.5688	5.07E-11	1.17E-09	14.5582
YHR123W	-2.8991	-6.5508	5.72E-11	1.21E-09	14.4419
YMR033W	-2.7637	-6.2447	4.25E-10	8.26E-09	12.5136
YPL031C	-2.7500	-6.2140	5.17E-10	9.34E-09	12.3248
YAL030W	-2.7044	-6.1109	9.91E-10	1.67E-08	11.6992
YGL178W	-2.6273	-5.9366	2.91E-09	4.60E-08	10.6652
YJL191W	-2.6131	-5.9045	3.54E-09	5.26E-08	10.4782
YJL041W	-2.5038	-5.6576	1.54E-08	2.16E-07	9.0725
YDL219W	-2.4941	-5.6357	1.74E-08	2.32E-07	8.9509

Table 3. Evidence for differential expression in the *prp2-1* mutant.

Evidence for differential expression of intron sequences after shifting cells carrying the temperature sensitive *prp2-1* mutation versus wild-type cells to 37 °C for 30 min. The “ORF Name” column lists the standard yeast ORF name associated with each intron. M values represent averaged, log₂-transformed ratios. The B-statistic represents the log posterior odds ratio.

ORF Name	M	t-statistic	p-value	Adjusted p-value	B-statistic
YJL001W	-3.9886	-9.0127	2.01E-19	5.09E-17	33.3086
YOL048C	-3.8762	-8.7585	1.98E-18	2.50E-16	31.0846
YBL040C	-3.6922	-8.3429	7.25E-17	6.11E-15	27.5849
YIL111W	-3.6067	-8.1497	3.65E-16	2.31E-14	26.0154
YGL087C	-3.4090	-7.7029	1.33E-14	6.73E-13	22.5282
YML067C	-3.3184	-7.4982	6.47E-14	2.73E-12	20.9958
YDR305C	-3.2750	-7.4001	1.36E-13	4.92E-12	20.2763
YGR001C_2	-3.1192	-7.0480	1.81E-12	5.74E-11	17.7715
YBL059C-A	-3.0470	-6.8851	5.78E-12	1.62E-10	16.6533
YDL137W	-2.9444	-6.6532	2.87E-11	7.26E-10	15.1078
YGR183C	-2.9071	-6.5688	5.07E-11	1.17E-09	14.5582
YHR123W	-2.8991	-6.5508	5.72E-11	1.21E-09	14.4419
YMR033W	-2.7637	-6.2447	4.25E-10	8.26E-09	12.5136
YPL031C	-2.7500	-6.2140	5.17E-10	9.34E-09	12.3248
YAL030W	-2.7044	-6.1109	9.91E-10	1.67E-08	11.6992
YGL178W	-2.6273	-5.9366	2.91E-09	4.60E-08	10.6652
YJL191W	-2.6131	-5.9045	3.54E-09	5.26E-08	10.4782
YJL041W	-2.5038	-5.6576	1.54E-08	2.16E-07	9.0725
YDL219W	-2.4941	-5.6357	1.74E-08	2.32E-07	8.9509

Table 4. Evidence for differential expression in the *prp8-1* mutant.

Evidence for differential expression of intron sequences after shifting cells carrying the temperature sensitive *prp8-1* mutation versus wild-type cells to 37 °C for 30 min. The “ORF Name” column lists the standard yeast ORF name associated with each intron. M values represent averaged, log₂-transformed ratios. The B-statistic represents the log posterior odds ratio.

ORF Name	M	t-statistic	p-value	Adjusted p-value	B-statistic
YML067C	-3.1028	-11.7890	1.55E-05	1.44E-03	3.8717
YJR094W-A	-3.0386	-11.3647	1.94E-05	1.44E-03	3.6517
YGR001C 2	-3.2343	-11.2555	2.05E-05	1.44E-03	3.5935
YHR123W	-3.0628	-11.0594	2.28E-05	1.44E-03	3.4873
YMR194W	-2.8092	-10.2964	3.51E-05	1.78E-03	3.0529
YJL191W	-2.7807	-9.8142	4.68E-05	1.97E-03	2.7594
YBL040C	-2.5692	-9.3074	6.42E-05	1.98E-03	2.4338
YDR305C	-3.5341	-9.1815	6.96E-05	1.98E-03	2.3500
YLR048W	-2.4223	-9.1631	7.05E-05	1.98E-03	2.3376
YNL162W	-2.6185	-8.9422	8.14E-05	2.04E-03	2.1873
YJL001W	-2.5613	-8.6842	9.68E-05	2.04E-03	2.0070
YJR145C	-2.2148	-8.5027	1.10E-04	2.04E-03	1.8768
YLR448W	-2.3101	-8.4978	1.10E-04	2.04E-03	1.8733
YEL012W	-2.1904	-8.4567	1.13E-04	2.04E-03	1.8434
YPL090C	-2.1657	-8.3528	1.22E-04	2.05E-03	1.7672
YGL087C	-2.1096	-8.0377	1.52E-04	2.41E-03	1.5304
YBL050W	-1.9930	-7.7232	1.92E-04	2.51E-03	1.2849
YFL039C	-1.9712	-7.6930	1.96E-04	2.51E-03	1.2609
YPL031C	-2.1704	-7.6696	2.00E-04	2.51E-03	1.2422

Table 5. Evidence for differential expression in the *prp5-1* mutant.

Evidence for differential expression of intron sequences after shifting cells carrying the temperature sensitive *prp5-1* mutation versus wild-type cells to 37 °C for 30 minutes. The “ORF Name” column lists the standard yeast ORF name associated with each intron. M values represent averaged, log₂-transformed ratios. The B-statistic represents the log posterior odds ratio.

ORF Name	M	t-statistic	p-value	Adjusted p-value	B-statistic
YGL137W	-1.2963	-5.6869	1.29E-08	3.27E-06	9.2541
YLR367W	1.2353	5.4193	5.98E-08	7.57E-06	7.7926
YLR406C	-1.0848	-4.7587	1.95E-06	1.64E-04	4.4860
YBL026W	-1.0366	-4.5474	5.43E-06	3.44E-04	3.5188
YHR039C-A	-0.9822	-4.3089	1.64E-05	6.56E-04	2.4802
YGR034W	-0.9810	-4.3034	1.68E-05	6.56E-04	2.4565
YER179W	-0.9771	-4.2867	1.81E-05	6.56E-04	2.3859
YNL147W	-0.9595	-4.2093	2.56E-05	8.10E-04	2.0626
YHR141C	-0.8817	-3.8678	1.10E-04	3.09E-03	0.7059
YJL189W	-0.8748	-3.8376	1.24E-04	3.14E-03	0.5917
YDR471W	-0.8695	-3.8144	1.37E-04	3.14E-03	0.5044
YLR388W	-0.8318	-3.6490	2.63E-04	5.55E-03	-0.1028
YDR447C	-0.7810	-3.4261	6.12E-04	1.19E-02	-0.8784
YKL186C_4	-0.7749	-3.3995	6.75E-04	1.22E-02	-0.9678
YOL127W	-0.7239	-3.1759	1.49E-03	2.52E-02	-1.6908
YML024W	-0.7012	-3.0759	2.10E-03	3.32E-02	-1.9983
YNL246W	-0.6486	-2.8454	4.44E-03	6.60E-02	-2.6696
YOR293W	-0.6420	-2.8164	4.86E-03	6.83E-02	-2.7503
YBR078W	-0.6237	-2.7363	6.21E-03	8.27E-02	-2.9692

Table 6. Evidence for differential expression in the *prp8-101* mutant.

Evidence for differential expression of intron sequences after shifting cells carrying the temperature sensitive *prp8-101* mutation versus wild-type cells to 37° C for 30 min. The “ORF Name” column lists the standard yeast ORF name associated with each intron. M values represent averaged, log₂-transformed ratios. The B-statistic represents the log posterior odds ratio.

References

1. Madhani HD, Guthrie C (1994) Dynamic RNA-RNA interactions in the spliceosome. *Annu Rev Genet* 28: 1-26.
2. Valadkhan S (2005) snRNAs as the catalysts of pre-mRNA splicing. *Curr Opin Chem Biol* 9: 603-608.
3. Nilsen TW (2003) The spliceosome: the most complex macromolecular machine in the cell? *Bioessays* 25: 1147-1149.
4. Jurica MS, Moore MJ (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 12: 5-14.
5. Staley JP, Guthrie C (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92: 315-326.
6. Ast G (2004) How did alternative splicing evolve? *Nat Rev Genet* 5: 773-782.
7. Graveley BR (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* 17: 100-107.
8. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

9. Davis CA, Grate L, Spingola M, Ares M, Jr. (2000) Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res* 28: 1700-1706.
10. Ares M, Jr., Grate L, Pauling MH (1999) A handful of intron-containing genes produces the lion's share of yeast mRNA. *Rna* 5: 1138-1139.
11. Clark TA, Sugnet CW, Ares M, Jr. (2002) Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* 296: 907-910.
12. Srinivasan K, Shiue L, Hayes JD, Centers R, Fitzwater S, et al. (2005) Detection and measurement of alternative splicing using splicing-sensitive microarrays. *Methods* 37: 345-359.
13. Kim SH, Lin RJ (1996) Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol Cell Biol* 16: 6810-6819.
14. Xu YZ, Newnham CM, Kameoka S, Huang T, Konarska MM, et al. (2004) Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *Embo J* 23: 376-385.
15. Grainger RJ, Beggs JD (2005) Prp8 protein: at the heart of the spliceosome. *Rna* 11: 533-557.
16. Collins CA, Guthrie C (2000) The question remains: is the spliceosome a ribozyme? *Nat Struct Biol* 7: 850-854.

17. Collins CA, Guthrie C (2001) Genetic interactions between the 5' and 3' splice site consensus sequences and U6 snRNA during the second catalytic step of pre-mRNA splicing. *Rna* 7: 1845-1854.
18. Umen JG, Guthrie C (1995) A novel role for a U5 snRNP protein in 3' splice site selection. *Genes Dev* 9: 855-868.
19. Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, et al. (2002) Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci U S A* 99: 5860-5865.
20. Schwer B, Guthrie C (1991) PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* 349: 494-499.
21. Noble SM, Guthrie C (1996) Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. *Genetics* 143: 67-80.
22. Jenny A, Minvielle-Sebastia L, Preker PJ, Keller W (1996) Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. *Science* 274: 1514-1517.
23. Chanfreau G, Noble SM, Guthrie C (1996) Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). *Science* 274: 1511-1514.

24. Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, et al. (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95: 717-728.
25. Park JW, Parisky K, Celotto AM, Reenan RA, Graveley BR (2004) Identification of alternative splicing regulators by RNA interference in *Drosophila*. *Proc Natl Acad Sci U S A* 101: 15974-15979.
26. Vithana EN, Abu-Safieh L, Allen MJ, Carey A, Papaioannou M, et al. (2001) A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol Cell* 8: 375-381.
27. McKie AB, McHale JC, Keen TJ, Tarttelin EE, Goliath R, et al. (2001) Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum Mol Genet* 10: 1555-1562.
28. Chakarova CF, Hims MM, Bolz H, Abu-Safieh L, Patel RJ, et al. (2002) Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum Mol Genet* 11: 87-92.
29. DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680-686.
30. Bozdech Z, Zhu J, Joachimiak MP, Cohen FE, Pulliam B, et al. (2003) Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol* 4: R9.

31. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5: R80.
32. Gentleman R (2005) *Bioinformatics and computational biology solutions using R and Bioconductor*. New York: Springer Science+Business Media. xix, 473 p.
33. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, et al. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30: e15.
34. Smyth GK, Michaud J, Scott HS (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21: 2067-2075.
35. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3: Article3.
36. de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* 20: 1453-1454.
37. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14: 1188-1190.
38. Hammell CM, Gross S, Zenklusen D, Heath CV, Stutz F, et al. (2002) Coupling of termination, 3' processing, and mRNA export. *Mol Cell Biol* 22: 6441–6457.

Acknowledgements

We thank J. DeRisi, A. Carroll, and M. Ares for technical assistance with the microarrays; J. Abelson, Q. Mitrovich, G. Wilmes, C. Maeder, T. Villa, M. Inada, and members of the C.G. lab for helpful discussions and critical comments on the manuscript. C.G. is an American Cancer Society Research Professor of Molecular Genetics.

Chapter 2: Rapid, Transcript-Specific Changes in Splicing in Response to Environmental Stress.

Abstract

While the core splicing machinery is highly conserved between budding yeast and mammals, the absence of alternative splicing in *Saccharomyces cerevisiae* raises the fundamental question of why introns have been retained in only 5% of the 6000 genes. Because ribosomal protein-encoding genes (RPGs) are highly overrepresented in the set of intron-containing genes, we tested the hypothesis that splicing of these transcripts would be regulated under conditions in which translation is impaired. Using a microarray-based strategy, we find that, within minutes after the induction of amino acid starvation, the splicing of the majority of RPGs is specifically inhibited. In response to an unrelated stress, exposure to toxic levels of ethanol, splicing of a different group of transcripts is inhibited, while the splicing of a third set is actually improved. We propose that regulation of splicing, like transcription, can afford rapid and specific changes in gene expression in response to the environment.

Introduction

Messenger RNA splicing is an essential part of eukaryotic gene expression as the coding regions of most eukaryotic genes are interrupted by noncoding introns. In higher eukaryotes, the process of splicing is utilized to regulate both qualitative and quantitative aspects of gene expression (reviewed in Black, 2000 and Blencowe, 2006). Alterations in the pattern of splice site usage in multi-intronic transcripts can produce a spectrum of protein isoforms from a single genomic locus, greatly expanding the genetic repertoire of an organism. Furthermore, when coupled to nonsense-mediated decay, alternative splicing can function as an “on/off” switch by introducing premature termination codons, thereby directing mRNA degradation (Lewis et al., 2003 and Mitrovich and Anderson, 2000).

By contrast, only 5% of genes in the budding yeast *Saccharomyces cerevisiae* are intron containing, and with few exceptions these genes have only a single intron (Spingola et al., 1999). Splice site sequences in yeast introns generally conform to a strict consensus, and documented instances of alternative splicing are rare (Davis et al., 2000). Nonetheless, a number of examples of regulated splicing have been demonstrated in yeast, including a set of transcripts that are constitutively transcribed but are efficiently spliced only during meiosis (Davis et al., 2000, Juneau et al., 2007 and Nandabalan et al., 1993). These introns, as well as the autoregulated introns in Rpl30 and Yra1, all contain nonconsensus splice sites, which are required for regulation (Eng and Warner, 1991 and Preker et al., 2002). However, because most yeast introns contain consensus splice site sequences, it has remained unknown whether splicing could function as a more general regulator of gene expression.

Interestingly, the set of intron-containing genes in yeast includes many metabolic regulators and is highly enriched for ribosomal protein genes (RPGs). Of the 139 RPGs encoded in the yeast genome, 102 are interrupted by at least one intron, making this by far the largest functional category. It has been previously established that starvation for amino acids leads to transcriptional repression of RPG synthesis, repression of rRNA synthesis, a general repression of translation, and an upregulation of enzymes involved in amino acid biosynthesis through a process controlled by the nonessential kinase Gcn2 (Chen and Powers, 2006, Cherkasova and Hinnebusch, 2003, Dever et al., 1992 and Hinnebusch, 2005). Because of the overrepresentation of translational components among the intron-containing genes, we hypothesized that the splicing of these transcripts might also be regulated in response to amino acid starvation.

Here we have taken a microarray-based strategy to examine the transcript-specific splicing changes resulting from exposure to two unrelated but environmentally relevant stresses: amino acid starvation and ethanol toxicity. We find that the splicing of the majority of RPGs is inhibited within minutes of inducing amino acid starvation. By comparison, exposure to toxic levels of ethanol, which is not known to induce a global repression of translation, has little effect on the splicing of the RPG transcripts. Rather, in response to the latter stress, the splicing of a different set of transcripts is downregulated, while the splicing efficiency of a third group of transcripts is improved. The specificity of these responses and the speed of their onset argue that splicing provides an important opportunity for regulation of gene expression in response to environmental stress. Furthermore, the capacity for transcription-independent regulation may explain the evolutionary retention of introns in these genes.

Results

A Genome-wide Approach to Examining Pre-mRNA Splicing

We have employed a microarray-based strategy to monitor quantitative changes in the splicing of ~250 intron-containing yeast transcripts in response to alterations in the environment. The microarrays contain three different oligonucleotide probes targeting each intron-containing gene, as described previously (Clark et al., 2002 and Pleiss et al., 2007). These probes allow for the independent determination of differences in the levels of the precursor mRNA species (probe P), mature mRNA species (probe M), and total transcript (probe T) between two samples (Figure 17). Whereas analysis of a traditional expression array hinges upon a single measurement of the relative abundance of a transcript in an experimental and reference sample, analysis of changes in the splicing behavior of a transcript requires the simultaneous consideration of the relative behaviors of all three types of probes. By examining the behavior of both the precursor species and the mature species against the background of changes in the total level of transcript, splicing efficiency can be distinguished from overall changes in transcript levels (Pleiss et al., 2007). Thus, a splicing profile can be generated and used to identify transcripts whose splicing is affected by an experimental treatment.

Changes in Pre-mRNA Splicing in Response to Amino Acid Starvation

We began by examining the cellular response to amino acid starvation. Amino acid starvation was induced by addition of the histidine-mimic 3-aminotriazole (3AT) (Hinnebusch, 2005). Cells were collected at several short times after either a mock treatment or exposure to 50 mM 3AT. Figure 18A shows the time-resolved splicing

profiles for each intron-containing gene obtained from microarray experiments comparing RNA isolated from cells exposed to these two treatments. Genes have been divided into two groups: RPGs and non-RPGs. Within these groups, the genes are ordered based on the similarity of the behavior of their three feature types across the amino acid starvation time course. Notably, the splicing of the majority of RPG transcripts is rapidly reduced, as evidenced by the increased abundance of precursor mRNA in starved relative to unstarved cells. Not surprisingly, because RPG transcripts exhibit a strong partitioning toward their spliced form (Pleiss et al., 2007), little decrease is seen in the level of mature mRNA for these species during the short time course of 3AT exposure. By comparison, little change is seen in the level of pre-mRNA for any of the non-RPG transcripts. Interestingly, several non-RPG transcripts do show a rapid decrease in the level of total mRNA and mature mRNA present, suggesting that the mature forms of these transcripts may be rapidly destabilized in response to amino acid starvation.

Quantitative RT-PCR Validation of Splicing Inhibition

To validate the observed response to amino acid starvation, quantitative RT-PCR experiments using both intron- and exon-specific primers were employed. As seen in Figure 19, results from the quantitative PCR assay are in good agreement with the microarray data. Both methods show accumulation of the RPG pre-mRNAs corresponding to Rpl21a and Rpl17b at very short time points after induction of amino acid starvation. As seen in Figure 23 in the Supplemental Data available with this article online, nearly equal levels of accumulation are detected when using primers that include the intron and the entire second exon, demonstrating that the accumulating product is in

fact an unspliced species. By contrast, no defect is detected in the splicing of either the U3 pre-snRNA or the Tef4 pre-mRNA.

Comparison with a General Inhibition of Splicing

Because ribosomal proteins are highly transcribed (Holstege et al., 1998), it was important to demonstrate that the observed RPG-specific phenotype truly results from an RPG-specific change in splicing and does not simply suggest that these pre-mRNAs are more likely to show accumulation given their high level of transcription. Thus, we compared the splicing profile derived from amino acid starvation to that seen after inactivation of a core spliceosomal factor. Figure 18B shows the splicing profile resulting from shifting a strain containing the temperature-sensitive *prp8-1* mutant to the nonpermissive temperature. PRP8 encodes a stable component of the U5 small nuclear RNP, and mutations in Prp8 are expected to result in defects in the splicing of most actively transcribed pre-mRNAs. In response to inactivation of *prp8-1*, the fold increases in pre-mRNA levels are nearly equal for RPG and non-RPG transcripts alike. This strongly suggests that amino acid starvation does not impart a global defect in splicing but instead that the splicing of RPG transcripts is specifically downregulated. Importantly, these experiments also demonstrate the relative stability of the mature forms of the RPG mRNAs in response to defects in splicing; whereas the decreases in mature levels for many non-RPG transcripts are apparent within minutes of inactivation of Prp8, very little decrease in the level of mature RPG transcripts can be observed during the *prp8-1* time course.

Reduced RNA Turnover Is Unlikely to Account for the Increased Pre-mRNA Levels

Considered a priori, two different mechanisms could account for the observed accumulation of RPG transcript pre-mRNA during amino acid starvation. First, a reduction in the splicing efficiency of these transcripts could lead to an increase in their pre-mRNA levels. Alternatively, some fraction of these pre-mRNAs may normally be targeted to a decay pathway prior to their processing by the spliceosome. In this case, the specific inhibition of their decay could result in an increase in pre-mRNA levels. To discriminate between these possibilities, we asked whether RPG pre-mRNAs as a class are specifically stabilized in the absence of any one of a number of different components involved in RNA turnover. Figure 18C shows the splicing profiles resulting from deletion of the following: RRP6, a component of the nuclear exosome (Houseley et al., 2006); RAI1 and RTT103, two components of the Rat1 nuclear decay pathway (Kim et al., 2004); SKI2, a regulator of 3' to 5' decay (Frischmeyer et al., 2002 and van Hoof et al., 2002); UPF1, a component of the nonsense-mediated degradation pathway (Leeds et al., 1991); or XRN1, a cytoplasmic 5' to 3' exonuclease (Fillman and Lykke-Andersen, 2005). Notably, while the genome-wide changes in mRNA levels resulting from deletion of many of these factors have previously been reported (He et al., 2003, Lelivelt and Culbertson, 1999 and Wyers et al., 2005), our experiments specifically address the behavior of the pre-mRNA species.

Importantly, there is no evidence from these experiments that RPG pre-mRNAs as a class are constitutively degraded at significant levels under normal growth conditions. By contrast, our data validate several pre-mRNAs that were previously shown to be stabilized in response to mutations in these decay pathways. For example, the RPL28 pre-mRNA is strongly stabilized in the absence of UPF1 (He et al., 1993). Also, as we

had previously shown (Preker and Guthrie, 2006), the YRA1 pre-mRNA is strongly stabilized in the absence of XRN1. Several additional pre-mRNAs are also stabilized in response to each of these mutations. While we cannot formally rule out the possibility that a different, completely nonoverlapping degradation pathway may contribute to the phenotype observed in response to amino acid starvation, our results strongly suggest that the increased levels of pre-mRNA result from a failure of these transcripts to be efficiently spliced.

Gcn2 Is Not Required for the Spliceosomal Response to Amino Acid Starvation

Given the well-defined role of Gcn2 in regulating the response to amino acid starvation, it was important to determine whether the observed splicing changes were also dependent upon the activity of this nonessential kinase. The accumulation of uncharged tRNAs resulting from amino acid starvation is known to activate Gcn2, which subsequently phosphorylates the initiation factor eIF-2 α , causing a general decrease in translation initiation (Dong et al., 2000). By a mechanism that is now well understood, this results in the paradoxical increase in translation of the transcription factor Gcn4 (Hinnebusch, 1997), which leads to the transcriptional upregulation of numerous genes involved in amino acid biosynthesis (Natarajan et al., 2001). Interestingly, as seen in Figure 19, the onset of the splicing response appears to precede the transcriptional upregulation of one Gcn4 target, HIS4, suggesting that the changes in splicing are independent of the Gcn4-mediated transcriptional response.

Using a strain deleted for GCN2, we repeated our experiment by comparing an untreated sample with one exposed to 3AT. As expected, the Δ gcn2 strain is largely unable to effect the upregulation of the biosynthetic genes in response to amino acid

starvation (Figure 20A and B). Nonetheless, the capacity of this strain to downregulate RPG splicing is nearly unchanged. Indeed, a direct experimental comparison of the amino acid starvation responses of a wild-type and a $\Delta gcn2$ strain both treated with 3AT demonstrates their highly divergent transcriptional responses but nearly identical splicing responses (Figure 20C). Interestingly, slight differences are apparent between the two strains at the longer time points, with more pre-mRNA accumulation apparent in the $\Delta gcn2$ strain. We presume this difference reflects the Gcn2-dependent global downregulation of RPG transcription known to accompany amino acid starvation. The onset of this transcriptional program presumably limits pre-mRNA accumulation in the wild-type cells.

Distinct Splicing Regulation Is Observed in Response to Ethanol Toxicity

To determine whether splicing changes affecting the RPG transcripts are specific to the cellular response to amino acid starvation or reflect a more general response to unfavorable growth conditions, we examined the effects of exposure to toxic levels of ethanol. Figure 21A shows the kinetic response of wild-type cells to ethanol toxicity. Interestingly, in response to ethanol exposure, a small number of transcripts decrease in splicing efficiency (marked with a red bar in Figure 21A and highlighted in the top of Figure 5B). As seen in Figure 21C, these transcripts show no splicing defect in response to amino acid starvation. Notably, the splicing efficiency of a second small subset of transcripts actually increases in response to ethanol (marked with a green bar in Figure 21A and highlighted at the bottom of Figure 21B). These transcripts likewise showed no change in splicing efficiency in response to amino acid starvation. A full comparison of the responses to amino acid starvation and toxic ethanol shown in Figure 22 reveals their nonoverlapping nature, demonstrating that the splicing changes are specifically tailored to each of these particular environmental stresses.

Discussion

Here we have taken a microarray-based approach to examine the capacity of pre-mRNA splicing in the budding yeast, *S. cerevisiae*, to be regulated in response to changes in environment. In response to amino acid starvation, the splicing efficiency of nearly all RPG transcripts is rapidly downregulated. Interestingly, these changes in splicing efficiency, unlike previously described responses to amino acid starvation, are not dependent upon the activity of the Gcn2 kinase. Furthermore, this response appears specific to amino acid starvation, as a different environmental stress, exposure to toxic levels of ethanol, does not globally repress RPG splicing. Instead, ethanol toxicity induces a unique spliceosomal response wherein the splicing efficiency of a different set of transcripts is downregulated while the splicing efficiency of another group of transcripts is improved.

The splicing architecture present in yeast is simple relative to higher eukaryotes. Very few genes in yeast contain multiple introns, and yeast introns tend to conform to tight consensus sequences at their splice sites, precluding the process of alternative splicing so rampant in metazoans. As such, it has been unclear whether yeast has retained the capacity to extensively regulate the activity of this step in gene expression. However, our experiments demonstrate that the splicing efficiency of distinct sets of transcripts can be rapidly and specifically modulated in response to changing environmental conditions. We have recently reported that mutations in core spliceosomal components can lead to transcript-specific defects in splicing (Pleiss et al., 2007), demonstrating that the yeast splicing architecture exhibits the physical capacity to differentiate among transcripts. Importantly, the results of our current experiments suggest that there must exist in yeast

mechanisms to differentiate splicing activity even among the many transcripts that contain strong consensus splice sites.

The Biology of Splicing Regulation

The splicing changes observed in response to amino acid starvation can be readily rationalized: under conditions in which translational resources are limiting, the spliceosome can rapidly downregulate the synthesis of new ribosomal components. Under normal growth conditions, wild-type yeast devote a significant fraction of their metabolic resources to the synthesis of new ribosomes. Many steps in this pathway are known to be tightly regulated, presumably reflecting the importance of optimizing the process of ribosome production (Warner, 1999). Our results suggest an explanation for the prevalence of introns in RPGs in that they may offer an additional level at which translational capacity can be coordinately regulated.

Moreover, the provocative result that the splicing of certain transcripts can be rapidly upregulated in response to ethanol toxicity suggests that not all yeast transcripts are processed by the spliceosome at maximal efficiency at all times. Indeed, upregulation of the splicing of these genes potentially allows for a rapid increase in protein expression without the need for de novo transcription. An analogy can be drawn with the unfolded protein response, in which induction of splicing of the Hac1 transcription factor, in this case by a nonspliceosomal mechanism, allows for a rapid response to the stress in the absence of new transcription (Sidrauski and Walter, 1997). Remarkably, one of the transcripts whose splicing is improved after ethanol exposure encodes Srb2, a nonessential component of the Mediator complex of transcription factors. Our results

suggest that expression of this conserved coactivator of transcription can also be regulated at the level of splicing and that splicing can in turn act to regulate transcription.

Mechanisms of Splicing Regulation

While it is true that yeast splice sites generally adhere to a strict consensus, all previously documented cases of regulated splicing in yeast rely on the presence of suboptimal splice site sequences. For example, a nonconsensus 5' splice site in the Rpl30 transcript is necessary to maintain a splicing-based autoregulatory feedback loop (Eng and Warner, 1991). Likewise, nonconsensus 5' splice sites in the transcripts of MER2 and MER3 are necessary for their positive regulation during meiosis (Nandabalan et al., 1993). However, splice site sequences alone are insufficient to explain the transcript-specific changes we have observed during amino acid starvation, as no unique sequences have been identified within this group of transcripts (Figure 23). Another potential explanation is based on the observation that, as a class, RPG introns tend to be significantly larger in size than introns in other yeast genes: the median RPG intron is ~400 nucleotides in length, whereas the median non-RPG intron is only ~130 nucleotides. Nevertheless, size alone is also insufficient to explain the specificity, as the splicing of non-RPGs such as Act1, whose intron contains 308 nucleotides, is unaffected during amino acid starvation. Thus, other currently unknown cis-acting features must contribute to the observed specificity.

In higher eukaryotes, members of the large SR and hnRNP families of regulatory proteins are known to play major roles in regulating alternative splicing decisions. By comparison, the spectrum of yeast proteins belonging to these families is reduced, and it remains unknown whether these proteins play any role in regulating splicing activity in

yeast. Instead, recent evidence suggests that “specificity” factors will include core components of the spliceosome itself. RNAi-mediated knockdowns in flies of several essential splicing factors have been shown to have differential effects on the splicing of alternative exons (Park et al., 2004). Likewise, our recent microarray analyses of point mutations in conserved, core splicing components (Pleiss et al., 2007) also revealed remarkably transcript-specific effects that were not readily accounted for by splice site sequences per se. In particular, we previously demonstrated that the spliceosome interacts with RPG transcripts in a fundamentally different way than it does with non-RPG transcripts. The similar behavior of RPG pre-mRNAs in response to perturbations in essential, core spliceosomal components may well underlie the behavior of these pre-mRNAs in response to amino acid starvation.

We previously proposed that the rate of cotranscriptional loading of spliceosomal components would be higher for RPG transcripts than for non-RPG transcripts (Pleiss et al., 2007). Consistent with this proposal, recent genome-wide chromatin immunoprecipitation experiments demonstrated a higher density of the U1 snRNP on RPGs than nonribosomal protein genes (Tardiff et al., 2006). Conceivably, modification of a common spliceosomal factor that affects the cotranscriptional engagement of the spliceosome could be responsible for the RPG-specific reduction of splicing efficiency seen in response to amino acid starvation. By comparison, a separate genome-wide experiment (Moore et al., 2006) showed that the transcript encoding *Srb2* is one of a small number of intron-containing transcripts that fails to engage with the spliceosome cotranscriptionally. The improved splicing seen for this transcript in response to toxic ethanol may therefore also depend upon its ability to efficiently recruit active spliceosomes. For example, these data suggest that under normal growth conditions the *Srb2* transcript is inefficiently engaged by the spliceosome but that under conditions of

ethanol toxicity spliceosomal recruitment is increased, allowing for more efficient splicing. Further characterization of the behavior of this transcript will likely provide important insights into the general mechanism by which transcripts engage spliceosomes.

Connecting the Spliceosome to Cellular Environment

An important question for the future regards the mechanisms by which environmental signals are transduced to the splicing machinery. The unique molecular profiles observed in response to two unrelated physiological stresses suggest that at least two independent pathways connect the spliceosome with the cellular environment. The rapid onset of the molecular defects in response to amino acid starvation suggests that posttranslational modifications of spliceosomal components might likely be involved in this regulation. Interestingly, recent evidence suggests a potential role for ubiquitin in regulating the splicing cycle (Bellare et al., 2006). Alternatively, phosphorylation and methylation of regulatory proteins play an important role in regulating splicing in higher eukaryotes and would be obvious candidates in yeast as well (Shin and Manley, 2004). Nevertheless, the candidate most likely to control such a signaling pathway in response to amino acid starvation, the Gcn2 kinase, is not essential for the spliceosomal response. Rather, our results imply the existence of a previously uncharacterized signaling pathway linking amino acid starvation with transcript-specific changes in pre-mRNA splicing.

Implications

The findings presented here argue strongly that pre-mRNA splicing can play an important role in regulating gene expression even in a system that lacks the infrastructure necessary for alternative splicing. Further, these results argue against the notion that most transcripts interact equally with the spliceosome, suggesting instead that transcript identity is an important factor in determining splicing activity. Given the high level of conservation of the basic splicing machinery, we predict that the mechanisms utilized by yeast to regulate pre-mRNA splicing will also be important in higher eukaryotes, where they likely underlie or augment the regulation of alternative splice site choices.

Experimental Procedures

Sample Collection and Preparation

Amino acid starvation experiments utilized either the wild-type strain (yMB1) or the Δ gcn2 strain (yMB2). The strain yMB1 was constructed by repairing the HIS3 locus in the BY4742 strain (Brachmann et al., 1998) using standard methods. The strain yMB2 was constructed by disrupting the GCN2 locus of yMB1 with LEU2. A single master culture of these strains was grown at 30°C in 250 ml of minimal medium lacking histidine (Guthrie and Fink, 2002) until its optical density was between $A_{600} = 0.5$ and $A_{600} = 0.7$. An initial 15 ml sample was collected by filtration using Millipore HAWP0025 filters (<http://www.millipore.com>) prior to initiation of the time course. The filters were immediately frozen in N₂(l). Experiments were initiated by splitting the master culture

into two 100 ml aliquots and adding either 5 ml of water or 1 M 3AT. At the appropriate times, 15 ml of cells were collected.

For experiments examining ethanol toxicity, a single master culture of the strain BY4742 was grown at 30°C in 300 ml of YEPD (Guthrie and Fink, 2002) until its optical density was between $A_{600} = 0.5$ and $A_{600} = 0.7$. Samples were then collected in a 96-well format as previously described (Pleiss et al., 2007). Briefly, 180 μ l of water or 100% ethanol was added to each well of a 96-well plate. A reverse time course was initiated by addition of 1.6 ml aliquots of culture to the appropriate wells at the appropriate times. At the completion of the time course, cells were collected by centrifugation at $5000 \times g$ for 5 min.

For experiments examining inactivation of Prp8, master cultures of both the wild-type (yEJS1) and prp8-1 (yEJS17) strain were grown at 25°C in 100 ml of complete minimal medium until their optical densities were between $A_{600} = 0.5$ and $A_{600} = 0.7$. Cells were shifted from 25°C to 37°C, and 15 ml samples were collected by filtration (as above) at the indicated times.

For experiments examining the nonessential RNA turnover mutants, samples of each mutant strain (all derivatives of BY4741 Brachmann et al., 1998) and the wild-type strain (BY4742) were grown at 30°C in 50 ml of complete minimal medium until their optical densities were between $A_{600} = 0.5$ and $A_{600} = 0.7$. Cells were then collected by centrifugation at $3000 \times g$ for 5 min.

Samples for microarray analysis were isolated and prepared as previously described (Pleiss et al., 2007). Briefly, 25 μ g of total cellular RNA was collected from each sample

and converted into cDNA for each individual microarray. Data presented as figures are derived from independent biological replicates, which were utilized to perform dye-flipped microarray replicates. Additional experiments examining distinct biological replicates have been performed and yield similar results. Microarray images were collected and analyzed as previously described (Pleiss et al., 2007).

Quantitative PCR

Total cellular RNA for quantitative RT-PCR experiments was treated with DNase I (Fermentas, <http://www.fermentas.com>) according to the manufacturer's protocol prior to use. The cDNA was then produced as previously described (Pleiss et al., 2007). Quantitative PCR was then performed using an Opticon from MJ Research (<http://www.mjresearch.com>). Error bars in Figure 19 are the result of triplicate measurements from a single biological sample. Standard deviations were only slightly higher when comparing biological replicates. A series of 10-fold dilutions of genomic DNA covering a total range of 10^6 molecules was used to generate the standard curves. Genomic DNA was purified using a ZR Fungal DNA kit (Zymo Research) according to the manufacturer's protocol.

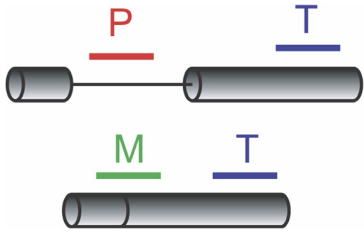


Figure 17. Splicing-specific microarrays.

Probes targeting intron-containing genes are designed to detect the precursor species (P), the mature species (M), or the total transcript (T).

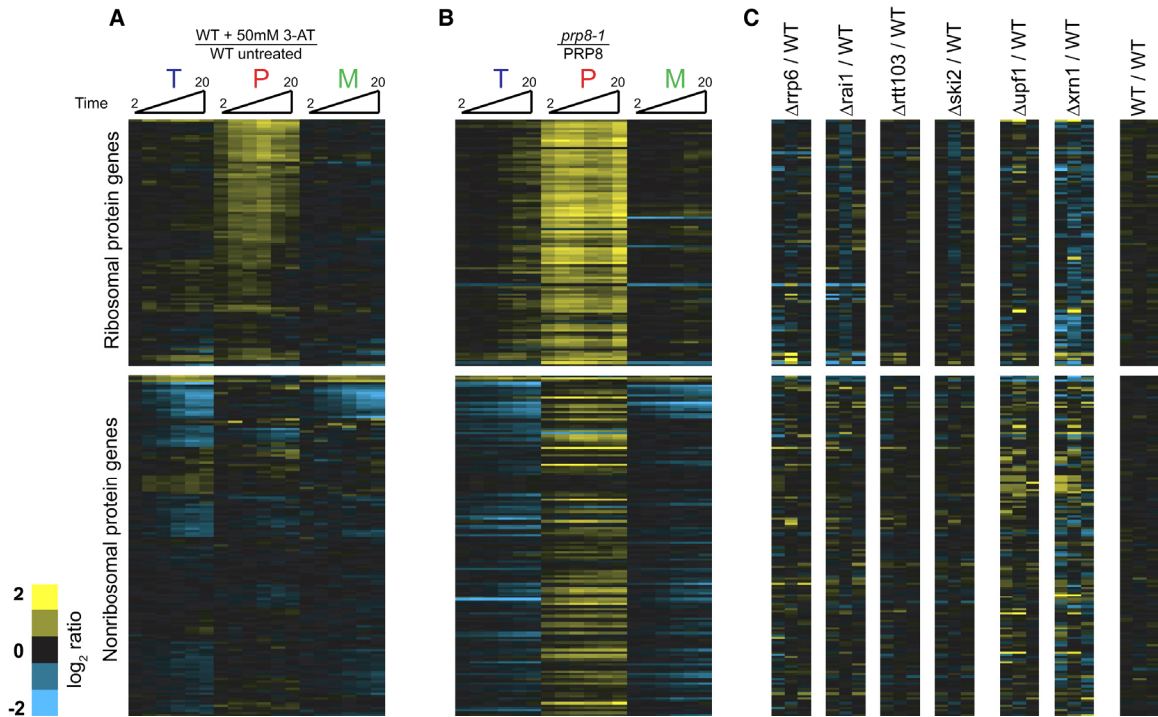


Figure 18. Regulation of pre-mRNA splicing in response to amino acid starvation.

Time-resolved splicing profiles resulting from comparisons of (A) wild-type cells either treated with 50 mM 3AT or mock treated for 2, 4, 6, 10, 15, and 20 min; (B) *prp8-1* and wild-type cells shifted from 25°C to 37°C for 5, 10, 15, 20, 25, and 30 min; and (C) wild-type cells compared to cells deleted for either *RRP6*, *RAI1*, *RTT103*, *SKI2*, *UPF1*, or *XRN1*. Also included is a comparison of two wild-type strains. The transcripts are ordered identically in all images.

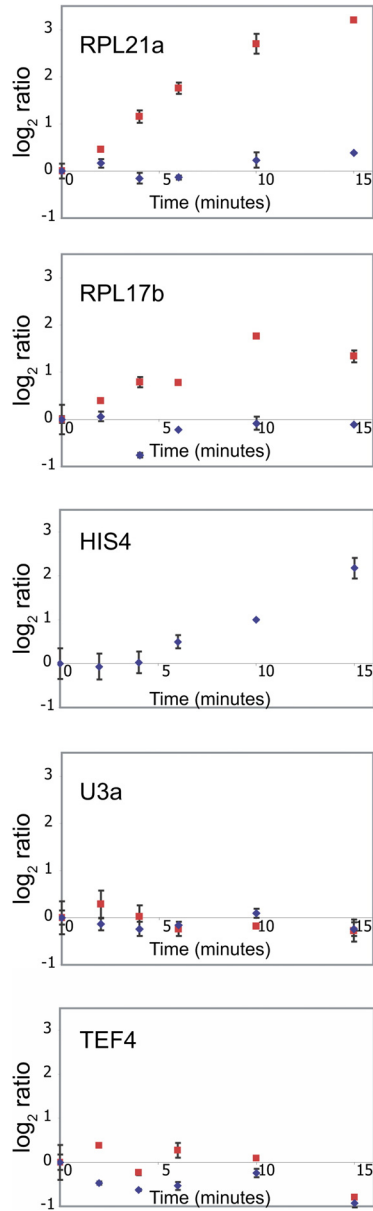


Figure 19. Quantitative RT-PCR validates the rapid, transcript-specific downregulation of splicing in response to amino acid starvation.

The behaviors of RPL21a, RPL17b, HIS4, U3 snRNA, and TEF4 were examined using primers specific to intron regions (red squares) or exon regions (blue diamonds). Error bars are the result of triplicate measurements from a single biological sample.

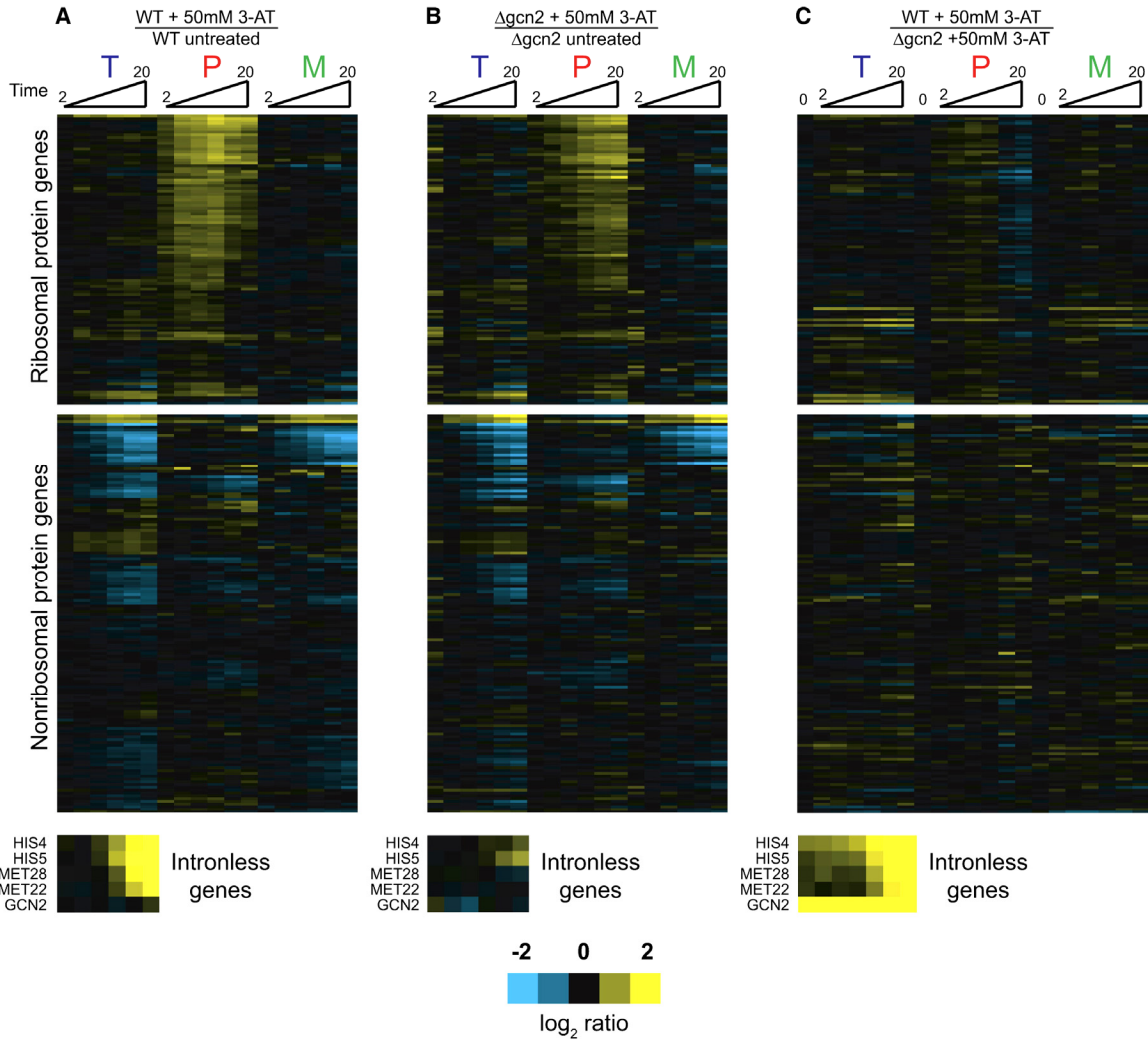


Figure 20. The splicing response to amino acid starvation does not require the activity of Gcn2.

Time-resolved splicing profiles resulting from (A) wild-type cells treated with 50 mM 3AT compared to mock-treated, (B) cells deleted for *GCN2* treated with 50 mM 3AT compared to mock-treated, or (C) wild-type cells treated with 50 mM 3AT compared to cells deleted for *GCN2* treated with 50 mM 3AT. Included in the final set is a comparison of the wild-type strain with the *GCN2*-deleted strain in the absence of 3AT

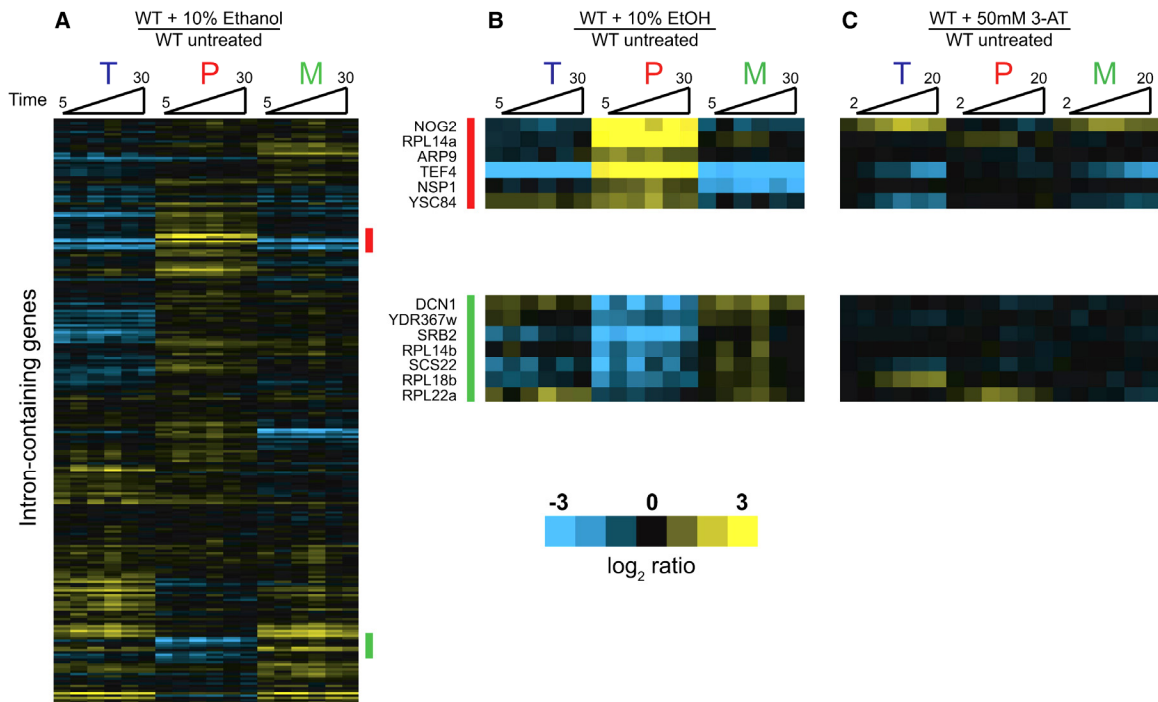


Figure 21. Regulation of pre-mRNA splicing in response to Ethanol Toxicity.

(A) Time-resolved splicing profiles resulting from comparisons of wild-type cells either treated with 10% ethanol or mock treated for 5, 10, 15, 20, 25, and 30 min. (B) Genes whose splicing is downregulated (top panel, also highlighted with a red bar in [A]) or upregulated (bottom panel, also highlighted with a green bar in [A]). (C) Behavior of genes highlighted in (B) in response to amino acid starvation.

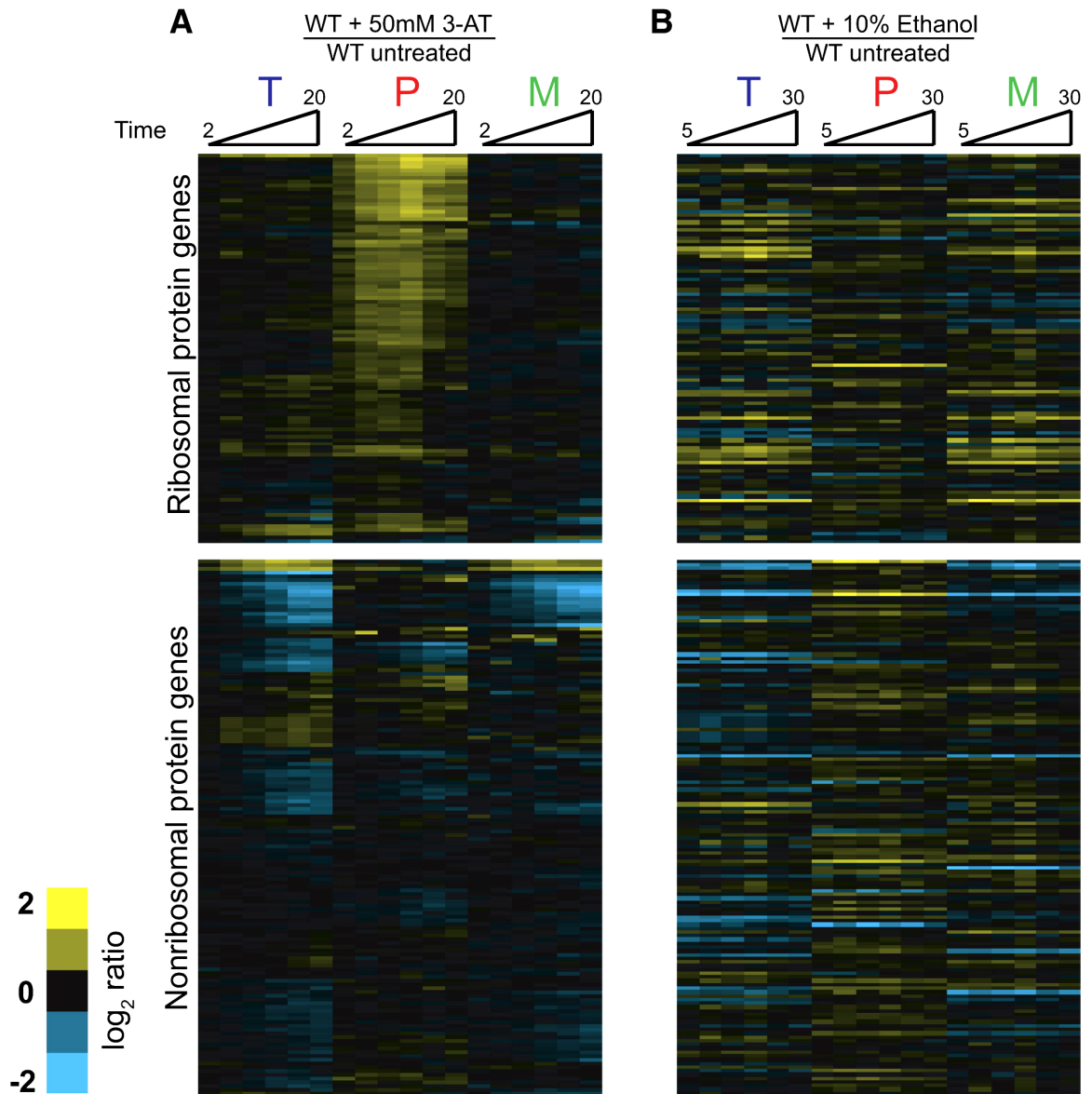


Figure 22. Comparison of splicing responses to amino acid starvation and ethanol toxicity.

The order of genes is identical to that shown in Figure 18.

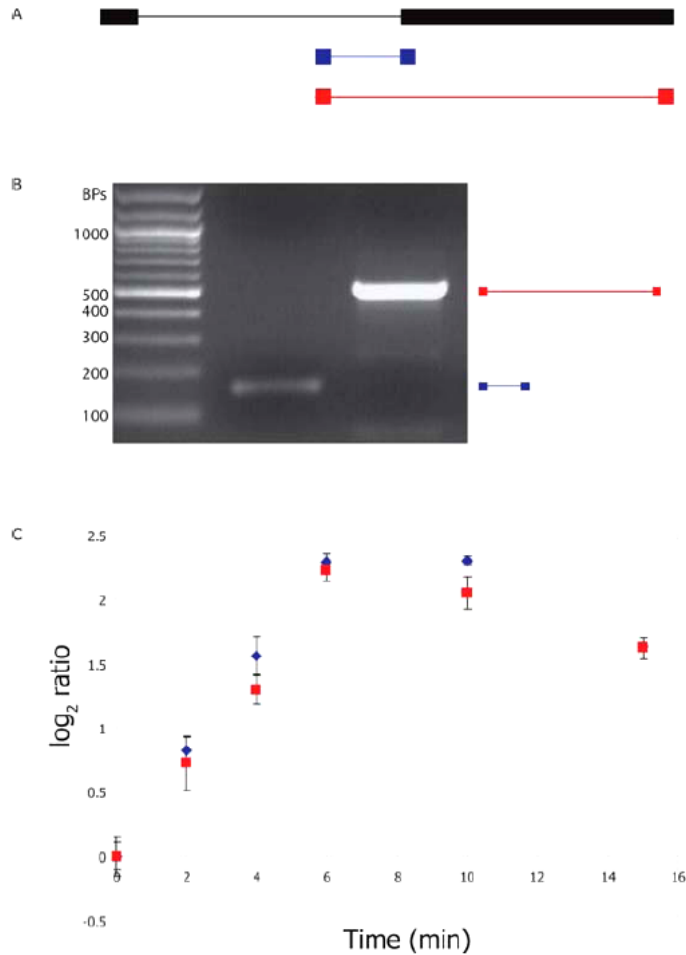


Figure 23. Accumulation of pre-mRNAs as measured using different primers.

(A) A scaled drawing of the architecture of the Rpl26a transcript is shown in black with exons denoted as boxes and the intron as a thin line. The locations of PCR primers are shown as boxes (blue and red), along with their products of amplification. (B) PCR products resulting from amplification using the primers depicted in (A). Expected product sizes are listed to the right of the gel. (C) Results of QPCR using the primers depicted in (A). Error bars are the result of duplicate measurements from a single biological sample.

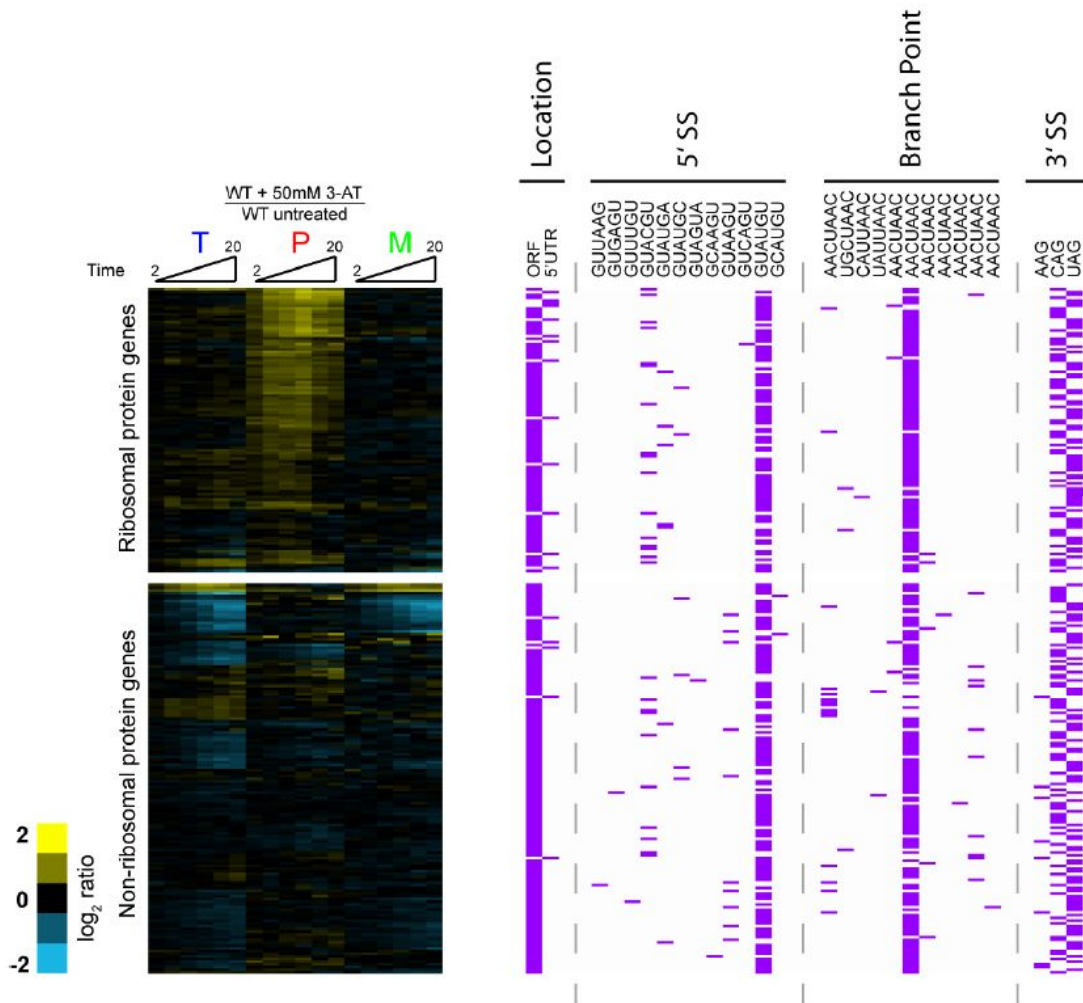


Figure 24. Analysis of splice site sequences and transcript behavior following amino

Purple hashes denote the locations of introns and the sequences present at their 5' and 3' splice sites and branch points.

References

Bellare, P., Kutach, A. K., Rines, A. K., Guthrie, C., and Sontheimer, E. J. (2006). Ubiquitin binding by a variant Jab1/MPN domain in the essential pre-mRNA splicing factor Prp8p. *Rna* 12, 292-302.

Black, D. L. (2000). Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* 103, 367-370.

Blencowe, B. J. (2006). Alternative Splicing: New Insights from Global Analyses. *Cell* 126, 37-47.

Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P., and Boeke, J. D. (1998). Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 14, 115-132.

Chen, J. C., and Powers, T. (2006). Coordinate regulation of multiple and distinct biosynthetic pathways by TOR and PKA kinases in *S. cerevisiae*. *Curr Genet* 49, 281-293.

Cherkasova, V. A., and Hinnebusch, A. G. (2003). Translational control by TOR and TAP42 through dephosphorylation of eIF2alpha kinase GCN2. *Genes Dev* 17, 859-872.

- Clark, T. A., Sugnet, C. W., and Ares, M., Jr. (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* 296, 907-910.
- Davis, C. A., Grate, L., Spingola, M., and Ares, M., Jr. (2000). Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res* 28, 1700-1706.
- Dever, T. E., Feng, L., Wek, R. C., Cigan, A. M., Donahue, T. F., and Hinnebusch, A. G. (1992). Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast. *Cell* 68, 585-596.
- Dong, J., Qiu, H., Garcia-Barrio, M., Anderson, J., and Hinnebusch, A. G. (2000). Uncharged tRNA activates GCN2 by displacing the protein kinase moiety from a bipartite tRNA-binding domain. *Mol Cell* 6, 269-279.
- Eng, F. J., and Warner, J. R. (1991). Structural basis for the regulation of splicing of a yeast messenger RNA. *Cell* 65, 797-804.
- Fillman, C., and Lykke-Andersen, J. (2005). RNA decapping inside and outside of processing bodies. *Curr Opin Cell Biol* 17, 326-331.
- Frischmeyer, P. A., van Hoof, A., O'Donnell, K., Guerrerio, A. L., Parker, R., and Dietz, H. C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* 295, 2258-2261.
- Guthrie, C., and Fink, G. R. (2002). *Guide to yeast genetics and molecular and cell biology* (Amsterdam ; Boston ; London, Academic Press).

He, F., Li, X., Spatrick, P., Casillo, R., Dong, S., and Jacobson, A. (2003). Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Mol Cell* 12, 1439-1452.

He, F., Peltz, S. W., Donahue, J. L., Rosbash, M., and Jacobson, A. (1993). Stabilization and ribosome association of unspliced pre-mRNAs in a yeast *upf1*- mutant. *Proc Natl Acad Sci U S A* 90, 7034-7038.

Hinnebusch, A. G. (1997). Translational regulation of yeast GCN4. A window on factors that control initiator-trna binding to the ribosome. *J Biol Chem* 272, 21661-21664.

Hinnebusch, A. G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. *Annu Rev Microbiol* 59, 407-450.

Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., Golub, T. R., Lander, E. S., and Young, R. A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717-728.

Houseley, J., LaCava, J., and Tollervey, D. (2006). RNA-quality control by the exosome. *Nat Rev Mol Cell Biol* 7, 529-539.

Juneau, K., Palm, C., Miranda, M., and Davis, R. W. (2007). High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proc Natl Acad Sci U S A* 104, 1522-1527.

Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedeá, E., Greenblatt, J. F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432, 517-522.

Leeds, P., Peltz, S. W., Jacobson, A., and Culbertson, M. R. (1991). The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev* 5, 2303-2314.

Lelivelt, M. J., and Culbertson, M. R. (1999). Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. *Mol Cell Biol* 19, 6710-6719.

Lewis, B. P., Green, R. E., and Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A* 100, 189-192.

Mitrovich, Q. M., and Anderson, P. (2000). Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in *C. elegans*. *Genes Dev* 14, 2173-2184.

Moore, M. J., Schwartzfarb, E. M., Silver, P. A., and Yu, M. C. (2006). Differential recruitment of the splicing machinery during transcription predicts genome-wide patterns of mRNA splicing. *Mol Cell* 24, 903-915.

Nandabalan, K., Price, L., and Roeder, G. S. (1993). Mutations in U1 snRNA bypass the requirement for a cell type-specific RNA splicing factor. *Cell* 73, 407-415.

Natarajan, K., Meyer, M. R., Jackson, B. M., Slade, D., Roberts, C., Hinnebusch, A. G., and Marton, M. J. (2001). Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Mol Cell Biol* 21, 4347-4368.

Park, J. W., Parisky, K., Celotto, A. M., Reenan, R. A., and Graveley, B. R. (2004). Identification of alternative splicing regulators by RNA interference in *Drosophila*. *Proc Natl Acad Sci U S A* 101, 15974-15979.

Pleiss, J. A., Whitworth, G. B., Bergkessel, M., and Guthrie, C. (2007). Transcript Specificity in Yeast Pre-mRNA Splicing Revealed by Mutations in Core Spliceosomal Components. *PLoS Biol* 5, e90.

Preker, P. J., and Guthrie, C. (2006). Autoregulation of the mRNA export factor Yra1p requires inefficient splicing of its pre-mRNA. *Rna* 12, 994-1006.

Preker, P. J., Kim, K. S., and Guthrie, C. (2002). Expression of the essential mRNA export factor Yra1p is autoregulated by a splicing-dependent mechanism. *Rna* 8, 969-980.

Shin, C., and Manley, J. L. (2004). Cell signalling and the control of pre-mRNA splicing. *Nat Rev Mol Cell Biol* 5, 727-738.

Sidrauski, C., and Walter, P. (1997). The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response. *Cell* 90, 1031-1039.

Spingola, M., Grate, L., Haussler, D., and Ares, M., Jr. (1999). Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *Rna* 5, 221-234.

Tardiff, D. F., Lacadie, S. A., and Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. *Mol Cell* 24, 917-929.

van Hoof, A., Frischmeyer, P. A., Dietz, H. C., and Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* 295, 2262-2264.

Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* 24, 437-440.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J. C., Dufour, M. E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B., et al. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121, 725-737.

Acknowledgements

We thank J. DeRisi, A. Carroll, and M. Ares for technical assistance with the microarrays; J. Abelson, J. Steitz, A. Johnson, H. Madhani, M. McMahon, M. Inada, and members of the C.G. lab for helpful discussions and critical comments on the manuscript. This work was supported by funds from the Damon Runyon Cancer Research Foundation (J.A.P.), N.I.H. and N.S.F. (G.B.W.), HHMI (M.B.), and N.I.H. grant GM21119 (C.G.). C.G. is an American Cancer Society Research Professor of Molecular Genetics.

Appendix 1: Exploring the Scope and Dynamics of Splicing Regulation

Preface

In chapter 2 we demonstrated that two different environmental treatments, 3-AT induced amino acid starvation and exposure to toxic concentrations of ethanol, both induced rapid, transcript-specific changes to the yeast splicing program. Immediately following 3-AT induced amino acid starvation, the splicing of a large set of ribosomal protein gene transcripts was rapidly inhibited. By contrast, following exposure to toxic concentrations of ethanol, the splicing of a small, unrelated, subset of transcripts was inhibited while the splicing of a second small subset of transcripts was improved. This study firmly established that environmental changes could be communicated to the process of splicing to alter the splicing efficiency of particular subsets of transcripts. However, it did so through the lens of just two different, unrelated, environmental conditions. Many compelling questions remain about both splicing in particular and rules that govern post-transcriptional control of gene expression in general. How prevalent, for example, are these types of splicing phenotypes? Do different types of starvation also inhibit the splicing of ribosomal protein genes, or is this phenotypes unique to 3-AT induced amino acid starvation? Are there other types of stresses which inhibit or promote the splicing of only a small number of targets?

In pursuit of these questions, we have significantly expanded our investigation of the interplay between changes in environmental conditions and the splicing efficiency of individual transcripts in yeast. Using our splicing-specific microarray platform we have now investigated the time-resolved splicing responses to glucose depletion and re-addition, nitrogen starvation, phosphate starvation, environmental amino acid depletion, cation toxicity, osmolarity stress, heat shock, and the unfolded protein response, among others.

Results & Perspectives

Many stresses induce defined changes to the splicing of a unique subset transcripts

One of the most striking results of this survey has been that many stresses appear to induce splicing changes to a unique, generally small, number of transcripts. Several examples are discussed below.

Osmotic Stress

Osmotic stress, induced by exposing cells to 1M KCl, results in two unique splicing phenotypes (Figure 29A). First, only three transcripts show a significant inhibition in splicing: GOT1, HNT1, and HNT2. Got1 is a relatively poorly characterized factor present in the early Golgi cisternae (Conchon et al., 1999), with no previously identified role in a cellular stress response. Strikingly, the two other transcripts which show splicing inhibition in response to osmotic stress encode Hnt1 and Hnt2, two of the three yeast members of the histidine triad superfamily of nucleotide hydrolase and transferase factors (Bieganowski et al., 2002; Rubio-Teixeira et al., 2002). The third member of this family, Hnt3, does not contain an intron and is not probed by our array platform. Both Hnt1 and Hnt2 have been shown to physically interact with Hsp82 (one of the two yeast Hsp90 homologs), which has been linked to protein chaperone roles in a variety of cellular processes (Millson et al., 2005). Interestingly, deletion of Hsp82 increases the sensitivity of cells to osmotic stress (Tabera et al., 2006). The dramatic inhibition of

splicing of both intron-containing members of this family of proteins suggests an interesting and novel connection between nucleotide hydrolase activity and the early response to osmotic stress.

Osmotic stress induces a second striking phenotype: improvement in the splicing efficiency of at least 48 ribosomal protein gene (RPG) transcripts. This set of RPG transcripts significantly overlaps with the set of 75 RPG transcripts whose splicing is inhibited in response to amino acid starvation induced by 3-AT (Figure 18). The only other condition under which we have seen a broad increase in the splicing efficiency of a large number of transcripts is treatment with mycophenolic acid (data not shown), a drug which decreases the transcriptional rate of elongating RNA polymerase by inducing guanylic nucleotide starvation. It was very surprising to see a significant improvement in the splicing efficiency of a large number of RPG transcripts, because our previous survey of absolute pre-mRNA and mature mRNA levels for a number of transcripts (Table 1) had suggested that RPGs are constitutively spliced very efficiently.

Cation toxicity

The splicing response to cation toxicity, induced by exposing cells to 300mM LiCl, demonstrates a very different set of changes. In this data set we see improvement in the splicing efficiency of just 4 RPG transcripts: RPS27B, RPS24A, RPS26A, and RPL27B. Additionally we see improvement in the splicing of two other factors, which have not demonstrated significant splicing changes under any of the other experimental conditions we have tested: VMA10 and QCR10. Vma10 is a component of a vacuolar proton-transporting ATP-dependent pump, which moves hydrogen into the vacuole to promote vacuolar acidification (Forgac, 1999). Although no direct link between the function of this complex and the cation toxicity response has been documented, up-

regulating its splicing efficiency in response to toxic intracellular cation levels makes elegant biological sense. Vma10 encodes the G subunit of the ATPase (V_1) domain of the complex, and is the only intron-containing member of either the catalytic or structural domain.

Cation toxicity appears to significantly inhibit the splicing of just two transcripts: LSM7 and YPR063C. Lsm7 is a component of both the nuclear and cytoplasmic Lsm complexes. YPR063C is a largely uncharacterized yeast ORF. However, in a high-throughput genetic screen it was shown to have a strong negative interaction with PMA1, which encodes a plasma membrane proton-transporting ATP-dependent pump, which moves hydrogen out of the cell (Schuldiner, 2005). The dramatic change in splicing efficiency of YPR063C in response to cation stress and the genetic link to PMA1 may be indicative of a role in ion homeostasis.

Glucose deprivation and re-addition

In a related set of experiments we have investigated the connection between transcript splicing efficiency and glucose availability (Figure 25C, D, and Figure 26). We have withdrawn glucose by shifting cells grown in rich media with glucose to rich media with no carbon source, shifting cells grown in minimal media with glucose to minimal media with no carbon source, and shifting cells grown in rich media with glucose to rich media containing glycerol. Over short time courses (<60 minutes), the splicing and expression changes induced by these treatments are all highly correlated. In each case we see splicing inhibition of just 4 RPGs: RPS27A, RPS27B, RPL14B, RPL34B. Notably, this very small set of transcripts contains both Rps27 paralogs. It is unlikely that this finding is an artifact of primary sequence similarity between these transcripts as the apparent splicing behavior of these two transcripts across our entire database of experiments is

not significantly more similar than the average between any other two RPG transcripts (Figure 28).

Interestingly, when we exposed cells to glucose that have been growing on a non-fermentable carbon source (glycerol), we see a set of splicing changes which are strongly anti-correlated with the set of glucose withdrawal experiments. In this case splicing of these 4 RPGs is improved (Figure 25D).

U3 splicing efficiency is anti-correlated with RPG expression levels across a wide range of starvation conditions.

In yeast, both copies of the U3 snoRNA, SNR17A and SNR17B, contain an intron. The U3 snRNP is an abundant complex required for ribosomal RNA maturation. In the glucose withdrawal dataset, the splicing efficiency of both copies of the U3 snRNA shows a marked improvement (Figure 26). Not surprisingly we see the converse phenotype for the U3 snoRNA in the glucose re-addition dataset.

What is striking about this phenotype, however, is that in both of these datasets U3 splicing efficiency is anti-correlated with expression levels of RPG transcripts. In fact, this trend can be observed across a wide range of datasets. Glucose deprivation, nitrogen starvation, treatment with DTT, rapamycin and wortmanin, all exhibit a rapid decrease in total levels of nearly every RPG transcript and an improvement in the splicing efficiency of both U3 snRNAs.

This pattern raises a compelling question: if these stress conditions are causing a rapid discard of the vast majority of RPG transcripts, why up-regulate the splicing of a

functional RNA involved in ribosome biogenesis? Although U3 is the only intron-containing snoRNA, we see a variety of expression changes in the total levels of other snoRNAs under these stress conditions, with some showing an increase in abundance and others a decrease.

One provocative model to explain this observation hinges on emerging evidence that among the many pairs of ribosomal protein gene paralogs in yeast, particular members of each pair may be specialized to supporting growth under different environmental conditions (Komili et al., 2007). If these findings bear out, we are led to the possibility that the ribosome is customized to suit particular growth conditions. If such a shift between specialized versions of the ribosome occurs in response to glucose withdrawal, for example, it would likely necessitate both the discard of the current cellular pool of RPG transcripts and the up-regulation of factors involved in ribosome biogenesis, such as the improvement in U3 splicing efficiency we see in this dataset. It could, therefore, be very interesting to test whether or not replacing the endogenous U3 loci with intronless copies of U3 leads to growth phenotypes under these starvation conditions.

As a final note here, the rapidity with which total RPG transcript levels plummet in the glucose starvation, nitrogen starvation, DTT treatment, rapamycin treatment and wortmanin treatment datasets is striking. Just a few minutes after shifting the cells the array data indicate that levels of most RPG transcripts have more been cut by more than 50%. This is particularly interesting given that the average half life of the RPGs under normal growth conditions is around 20 minutes (Foat et al., 2005). The kinetics of loss of total levels of RPGs are therefore indicative of a profound change in decay rate of these transcripts, suggesting that while splicing may not be operating on them as a post-

transcriptional control mechanism under these conditions, a regulated decay mechanism must be (Grigull, et al., 2004).

The severity of the challenge to cell growth is an important determinant of the splicing response

In our published data discussed in chapter 2, we demonstrated that amino-acid starvation induced by the drug 3-aminotriazole (3-AT) inhibits the splicing of a large number of ribosomal protein genes (Figure 18). In these experiments cells are grown in media lacking histidine, which leads to up-regulation of endogenous histidine biosynthesis. The drug 3-AT induces a state of amino acid starvation in this context because it is an inhibitor of endogenous histidine production. Therefore, amino acid starvation is the result of an insult to an already engaged pathway. Indeed, wild-type cells will quickly adapt to this insult by up-regulating amino acid biosynthesis. The kinetics of the splicing response are consistent with rapid adaptation to this particular type of amino acid starvation: the inhibition of RPG splicing peaks approximately 15 minutes after addition of 3-AT, and then slowly diminishes.

Interestingly, we do not see the transient inhibition of RPG splicing observed after 3-AT addition during the other starvation responses we have assayed, such as glucose deprivation, nitrogen starvation or phosphate starvation. As discussed above, although these conditions alter the splicing efficiency of a small number of transcripts, in general we see a rapid decrease in total RPG transcripts levels, rather than an alteration in splicing efficiency. We therefore decided to test the effects of different types of amino acid starvation, to assess the uniqueness of the 3-AT induced inhibition of RPG splicing.

Figure 27 shows a side-by-side comparison of the splicing response to amino acid starvation as induced by 3-AT addition to cells growing in the absence of environmental histidine, and amino acid starvation induced by shifting cells growing in media containing optimal concentrations of all amino acids to media containing the minimal concentration of only amino acids which were essential for viability of our test strain. Clearly, this latter method of inducing amino acid starvation does not lead to the regulation of RPG splicing efficiencies that we see in the 3-AT dataset. Importantly, however, at late time points in the amino acid withdrawal experiment we do see transcriptional up-regulation of the amino acid biosynthesis genes (data not shown), demonstrating that this treatment has in fact induced one form of an amino acid starvation response.

What, then, is unique about the induction of amino acid starvation by 3-AT? One possibility is, of course, that the inhibition of RPG splicing in these experiments is not so much a result of an amino acid starvation response as it is a specific, secondary effect of the drug. In an effort to address this possibility, we shifted cells lacking an essential component of the histidine biosynthetic pathway from media containing an optimal complement of all amino acids to media missing only histidine. Although this dataset is preliminary, it does show a 3-AT-like inhibition of the splicing of a large set of RPGs.

A more biologically interesting, and likely, explanation is that the 3-AT experiment is the only condition among our panel of stresses where cells have been challenged to simply up-regulate a previously engaged adaptive pathway, and therefore it constitutes a relatively mild challenge to cell growth. Although a broader survey of other relatively mild challenges to growth is needed to fully investigate this question, a compelling model would be that splicing is a particularly useful regulatory target when flux through a

specific gene expression pathway needs to be subtly adjusted. Following this model: immediately after 3-AT addition, cells experience a transient decrease in free histidine pools, so splicing is used to temporarily dial down production of new ribosomes, which, as we note in chapter 2, is a strategy that makes obvious biological sense for cells in which protein biosynthetic capacity has exceeded free amino acid pools (Warner, 1999). As cells ramp up the transcriptional circuit that controls production of the amino acid biosynthesis machinery, the splicing response is slowly abrogated.

Topology of the splicing landscape under changing environmental conditions

In total the Guthrie lab has now investigated the time-resolved splicing profiles of over 100 unique mutations in splicing factors and components of other mRNA processing machineries and of more than 20 unique environmental conditions. This dataset comprises over 2,000 individual microarray experiments all performed on the same platform, and it presents us with a unique opportunity to begin exploring trends which emerge when the splicing behavior of individual transcripts is considered across a wide range of experimental conditions.

An example of this kind of transcript “meta-analysis” is presented in Figure 28. In this analysis we have calculated the pair-wise similarities between the splicing behavior of every intron-containing transcript across experiments examining a mutation in an mRNA processing factor (Figure 28A), or experimentings examining an environmental stress condition (Figure 28B). If two transcripts exhibit very similar changes in their total, pre-mRNA and mature mRNA features on the splicing microarrays across a large number of experiments they will score as “close” in this analysis (low distance values). Conversely if two transcripts exhibit very different behaviors across a large number of experiments,

they will score as “far” (high distance values). In the second step of this analysis, the set of pair-wise distances between each pair of transcripts is then subjected to hierarchical clustering and the results used to order transcripts symmetrically along the horizontal and vertical axes of the figure. This manipulation leads to transcripts with like profiles of similarities being more closely clustered than transcripts with different similarity profiles.

Several interesting trends emerge from this analysis. Most strikingly, the topology of transcripts viewed from the perspective of responses to changes in the environment is more granular than the topology of transcripts viewed from the perspective of responses to mutations in mRNA processing factors. In the environmental response cluster, for example, a number of interesting subclusters can be identified which are significantly enriched for transcripts that encode proteins of similar function (for example, a cluster of factors involved in aerobic respiration is highlighted at the top of Figure 28B), suggesting potential splicing “regulons”.

There is however, one clear grouping of functionally related transcripts that emerges from the analysis of the response to mutations in mRNA processing factors: the ribosomal protein gene transcripts. Interestingly, the RPG transcripts split into two distinct clusters in the environmental stress analysis. Anecdotally, we can point to specific instances for why this is likely the case. For example, the 48 RPGs which show an improvement in splicing efficiency in response to osmotic stress (Figure 25) all fall into the top RPG cluster in this figure.

Although, at present, the precise biological underpinnings behind these two phenotypic groupings of RPGs are unclear, I am optimistic that it will be fruitful to compare clusters of transcripts which arise from our meta-analysis of the splicing microarray data with

other types of large scale datasets to see if meaningful correlations emerge. One particularly interesting type of data it would be interesting to begin to cross-reference are examples of single members of RPG paralog pairs which exhibit condition-specific growth phenotypes or unique sets of genetic interactions. Data genome-scale investigations of transcription factor binding sites, for example Rap1, are another potentially fruitful place to look for explanatory variables.

Materials and Methods

In general, cells were handled and microarray experiments performed as described in chapter 2. However, for each of the following environmental stresses cells were collected by vacuum filtration: shift from YPD to YP, shift from SD to SC, shift from YPD to YPG, nitrogen starvation, phosphate starvation, and the shift from SD-complete to SD-min. In each of these cases cells were collected on a nitrocellulose filter by vacuum filtration and washed with 500mL of the experimental destination media before re-suspension in that same media. Microarrays were performed as a competitive hybridizations of samples collected from cells exposed to the experimental treatment versus samples collected from cells exposed to a mock treatment.

References

Bieganowski, P., Garrison, P.N., Hodawadekar, S.C., Faye, G., Barnes, L.D., and Brenner, C. (2002) Adenosine monophosphoramidase activity of Hint and Hnt1 supports function of Kin28, Ccl1, and Tfb3. *J Biol Chem* **277**:10852-60.

Conchon, S., Cao, X., Barlowe, C., and Pelham, H.R. (1999). Got1p and Sft2p: membrane proteins involved in traffic to the Golgi complex. *EMBO J.* **18**:3934-46

Foat, B.C., Houshamandi, S.S., Olivas, W.M., and Bussemaker, H.J. (2005). Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. *PNAS* **102**: 17675-17680.

Forgac, M. (1999). Structure and properties of the vacuolar (H⁺)-ATPases. *J Biol Chem* **274**:12951-4.

Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M.D., and Hughes, T.R. (2004). Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol Cell Biol.* **24**: 5534-47.

Komili, S., Farny, N.G., Roth, F.P., and Silver, P.A. (2007). Functional specificity among ribosomal proteins regulates gene expression. *Cell* **131**: 557-71.

Millson, S.H., Truman, A.W., King, V., Prodromou, C., Pearl, L.H., and Piper, P.W. (2005). A two-hybrid screen of the yeast proteome for Hsp90 interactors uncovers a novel Hsp90 chaperone requirement in the activity of a stress-activated mitogen-activated protein kinase, Slf2p (Mpk1p). *Eukaryot Cell* **4**:849-60.

Rubio-Teixeira, M., Varnum, J.M., Bieganowski, P., and Brenner, C. (2002) Control of dinucleoside polyphosphates by the FHIT-homologous HNT2 gene, adenine biosynthesis and heat shock in *Saccharomyces cerevisiae*. *BMC Mol Biol* **3**:7

Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., Weissman, J.S., Krogan, N.J. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**:507-19.

Tabera, L., Munoz, R., Gonzalez, R. (2006). Deletion of BCY1 from the *Saccharomyces cerevisiae* genome is semidominant and induces autolytic phenotypes suitable for improvement of sparkling wines. *Appl Environ Microbiol* **72**:2351-8.

Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. *Trends in Biochemical Sciences* **24**: 437-440.

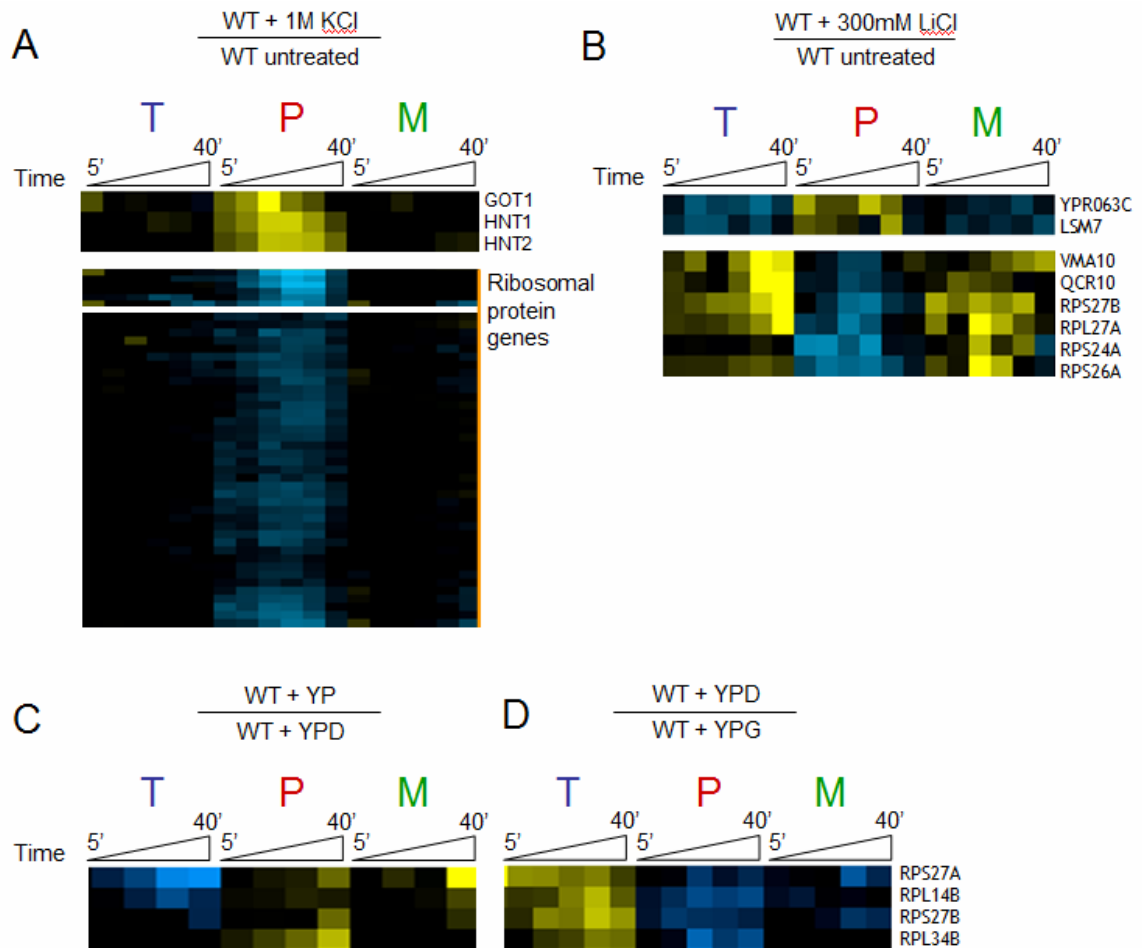


Figure 25. A variety of stresses induce defined effects on a small subset of transcripts.

Genes whose splicing is downregulated (top) or upregulated (bottom two) following treatment of wild-type cells treated with 1M KCl to induce osmotic. Genes are individually labeled on the top cluster, while the bottom two clusters represent 48 ribosomal protein genes. (B) Genes whose splicing is downregulated (top) or upregulated (bottom) following treatment of wild-type cells with 300mM LiCl to induce cation toxicity stress. (C) Genes whose splicing is downregulated by the shift from glucose containing rich media (YPD) to rich media lacking any carbon source (YP) when compared to mock treated cells. (D) Genes whose splicing is upregulated by the shift

from rich media with glycerol as the carbon source (YPG) to rich media containing glucose as the carbon source (YPD) when compared to mock treated cells.

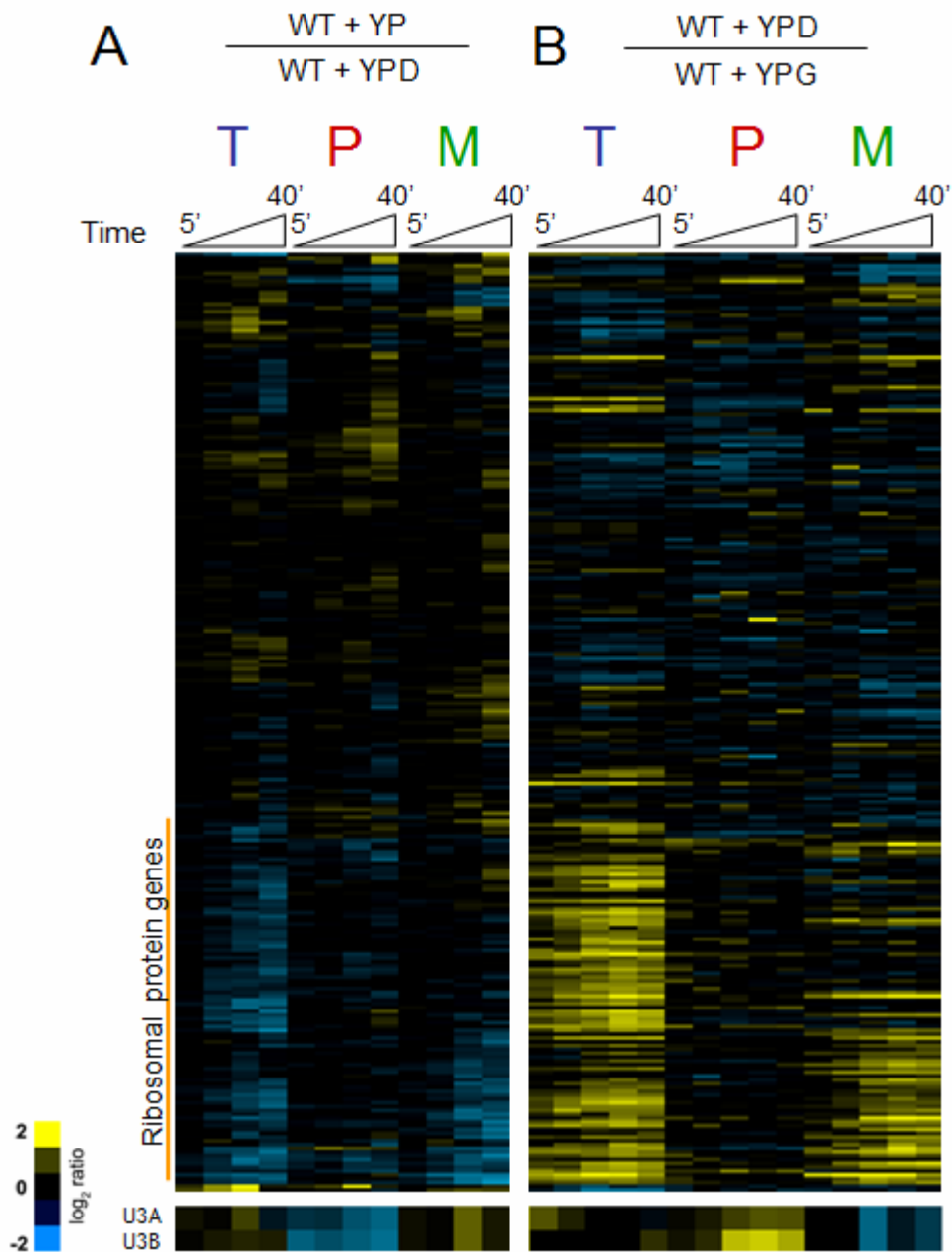


Figure 26. U3 splicing efficiency is anti-correlated with RPG expression levels.

Here the full response to glucose withdrawal (A) and glucose addition (B) are shown. Both clusters share a gene (vertical) axis. The 98 transcripts falling under the orange bar on the left hand side of the cluster are all structural components of the ribosome. The splicing responses of the U3 snRNAs are highlighted at the bottom of the figure.

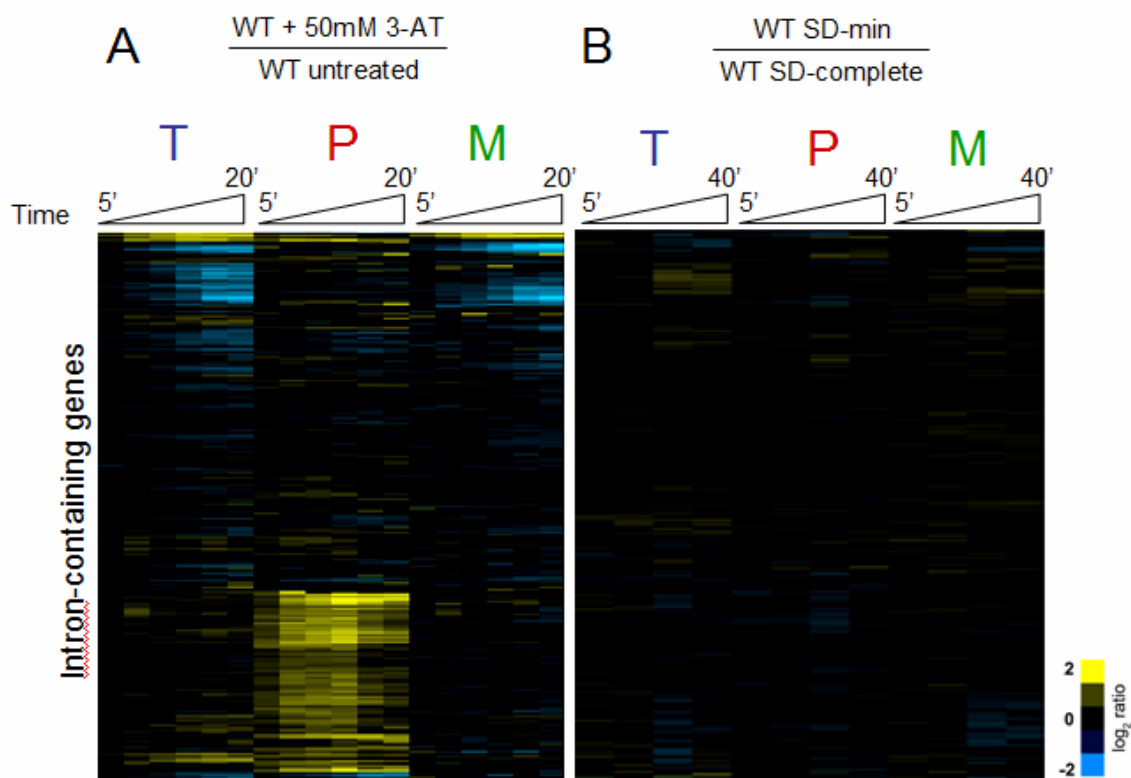


Figure 27. Comparison of amino acid starvation induced by 3-AT addition, and environmental amino acid withdrawal.

The responses to amino acid starvation induced by 3-AT (A) is shown in comparison to amino acid starvation induced by environmental amino acid withdrawal (B). These clusters share a gene (vertical) axis.

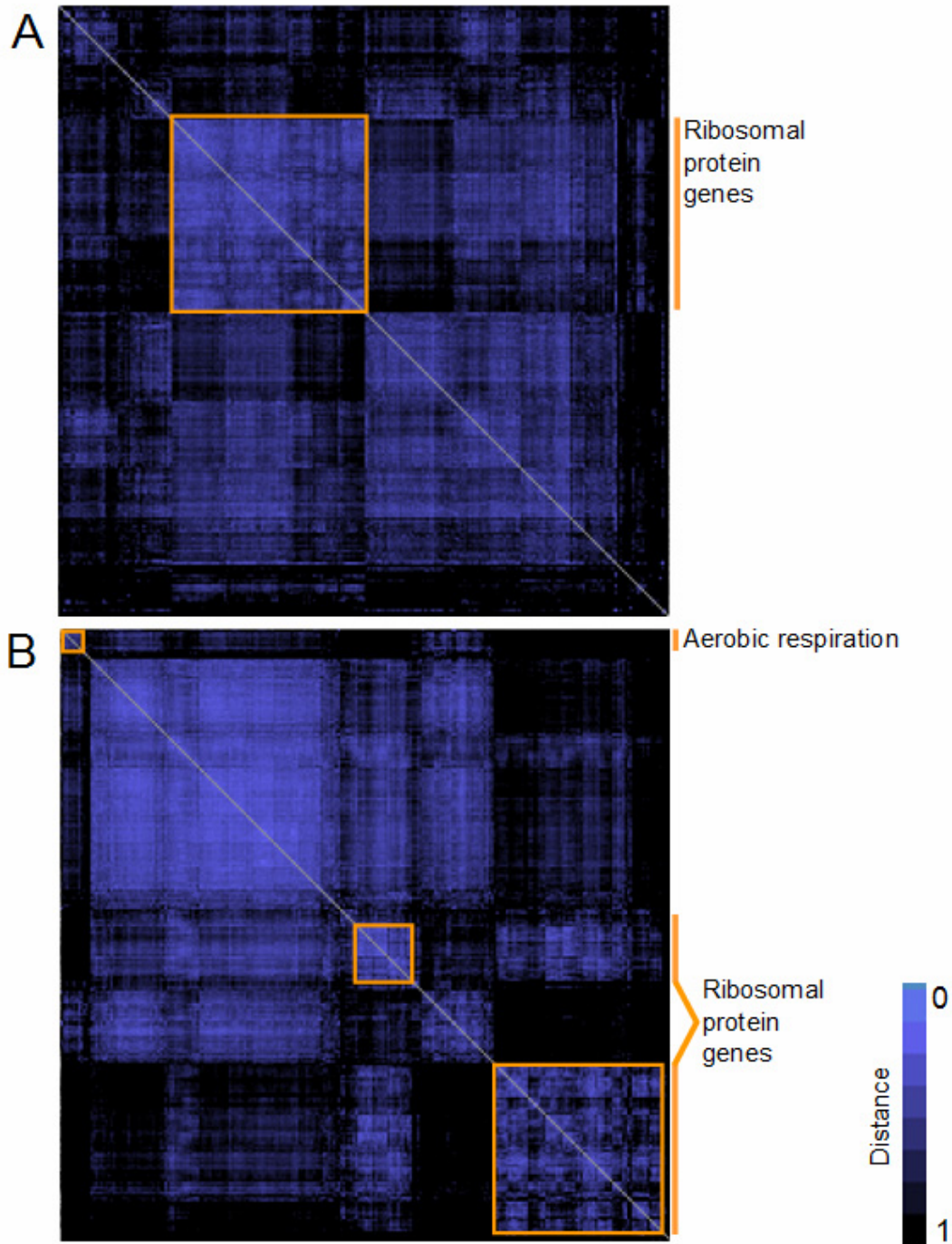


Figure 28. Analysis of similarity in splicing behavior of all intron-containing transcripts in response to mRNA processing mutations and environmental stress.

Both clusters display the pair-wise similarities between intron-containing transcripts along symmetrical vertical and horizontal axes. A light blue hash on the heat map

represents two transcripts that exhibit relatively similar behavior across experiments (low distance values), while black hashes represent transcripts with relatively dissimilar behaviors across experiment (high distances values). (A) Splicing behavior of intron-containing transcripts in microarray experiments investigating the mutations in mRNA processing factors. The highlighted orange box represents a cluster composed of transcripts encoding structural components of the ribosome. (B) Splicing behavior of intron-containing transcripts in microarray experiments investigating environmental stress conditions. The top orange box highlights a cluster of transcripts encoding factors involved in aerobic respiration, while the bottom two orange boxes highlight clusters of transcripts encoding structural components of the ribosome.

Appendix 2: Rapid Changes in the Behavior of Nuclear mRNA
Export Factors Following Environmental Stress

Preface

In eukaryotes, the complex life cycle of mRNAs affords the cell numerous opportunities for post-transcriptional regulation. Since the cell's active translational machinery is localized to the cytoplasm, it is conceptually possible that regulating the export competence of particular mRNAs could provide a powerful post-transcriptional opportunity for rapidly modulating the cell's protein expression program (Carmo-Fonseca, M., 2002). Indeed, nuclear retention of bulk poly(A) RNA has been observed following heat shock exposure in both yeast cells and mammalian tissue culture (Liu et al., 1996; Saavedra et al., 1996), while transcript-specific in situ hybridization studies in both systems have revealed that mRNAs encoding essential heat shock proteins continue to be exported to the cytoplasm (Saavedra et al., 1997). The retention of bulk poly(A), but maintenance of heat shock protein transcript export, has been proposed as a mechanism by which the expression of mRNAs encoding general housekeeping genes is rapidly suppressed to favor expression of mRNAs encoding factors essential for surviving heat shock conditions (Krebber et al., 1999).

Intriguingly, heat shock, among other stress responses, also modulates the localization of several of the major yeast shuttling mRNA export factors (Krebber et al., 1999; our results). Together these two sets of findings raise the possibility that changes in the export competence of different classes of mRNAs during a conditional response may be achieved through changes in the localization behavior of mRNP components, thereby altering composition of the pre-export mRNP complexes with which they are associated. This model would require that the mRNA export machinery be capable of exhibiting some degree of transcript-specificity, which runs counter to the prevailing idea that the mRNA export pathway is driven by a host of factors featuring generic RNA binding

domains. However, there is mounting experimental evidence that mRNA export factors exhibit non-overlapping profiles of transcript association under normal growth conditions (Kim-Guisbert, et al., 2005; Gerber et al., 2004; Hieronymus and Silver, 2003). These findings strongly suggest that some degree of transcript specificity is indeed built into the constitutive mRNA export machinery. Whether or not this apparent in vivo specificity of particular factors for specific subsets of transcripts is derived from a sequence specific interaction between the protein and RNA, or from “meta” signals like transcriptional context, chromatin state, or communication with other aspects of the nuclear processing mRNA processing machinery remains to be seen.

Nonetheless, if particular components of the mRNA export machinery do exhibit a preference for specific transcripts, is the alteration of their behavior following specific stress conditions a signifier of a role for nuclear mRNA export in post-transcriptional gene regulation? In a preliminary pursuit of this question, we sought to provide a higher resolution characterization of the localization and mRNA binding behaviors of a host of nuclear mRNA export factors following exposure to environmental stress.

Results

The localization and mRNA binding profile of Npl3 changes following salt stress

In previous studies it was shown that Npl3, an abundant, shuttling, nuclear export factor, shifts its steady state localization from the nuclear compartment to the cytoplasm following the exposure of yeast cells to heat shock, ethanol toxicity, and salt stress

(Krebber et al., 1999; Saavedra et al., 1996). Additionally, the association of Npl3 with poly(A) RNA was shown to be significantly decreased following salt stress, leading to the model that immediately following stress exposure, Npl3 disassociates from its normal nuclear cargo and is sequestered away from newly synthesized mRNAs in the cytoplasm, thereby shutting down the normal mRNA export pathway to promote translation of stress-responsive mRNAs, themselves exported by a putative alternative mechanism.

Using exposure to 500mM NaCl as a model stress we recapitulated the initial result that Npl3 rapidly re-localizes to the cytoplasm following salt stress (Figure 29A, 2 minutes). Interestingly, when we carry out the time course of Npl3 localization changes to time points concomitant with physiological adaptation to high salt conditions and the resumption of cell growth, we see that the nuclear concentration of Npl3 is ultimately re-established (Figure 29A, 30 minutes). Notably, at intermediate time points (5, 10 and 15 minutes), Npl3 shows nuclear rim staining, suggesting a possible kinetic block to Npl3 re-import into the nucleus at these time points. For simplicity, we will continue to refer to these time points as the nuclear re-import phase, but it is important to note that our current data do not formally prove that this nuclear rim staining results from a population of Npl3 in existence before the start of the stress time course (although the rapid kinetics strongly support this assumption), nor do they formally demonstrate that this staining represents a population of Npl3 on its way into the nucleus (although over expression of the Npl3 importin, Mtr10, significantly shortens the timing of this phase; data not shown). Together, this time course demonstrates that Npl3 is not permanently flushed from the nucleus following salt stress, but instead undergoes a transient shift in its bulk localization, which occurs with similar kinetics to the cellular adaptation to environmental stress and resumption of growth.

In previous work, the association of Npl3 with poly(A) RNA following stress exposure had been investigated in extracts from cells collected using centrifugation, a process which consumes approximately 30 minutes of experimental time. Given the incredibly rapid kinetics of the Npl3 localization cycle following salt stress, we re-evaluated the association of Npl3 with poly(A) RNA following stress exposure by collecting cells using vacuum filtration. This procedure allowed us to match time points in our Npl3 localization and poly(A) RNA association results to within 2-3 minutes of each other. Intriguingly, we found that Npl3 does not show an appreciable change in its association with poly(A) RNA until the nuclear re-accumulation phase of the stress induced localization cycle (Figure 29B).

This result requires an adjustment to the previously proposed model explaining Npl3 behavior following salt stress. Instead of dissociating from poly(A) RNA before leaving the nucleus, Npl3 appears to remain bound to RNA through the cytoplasmic phase of its localization cycle, and doesn't lose bulk poly(A) association until it re-accumulates in the nucleus. Importantly, though, our poly(A) RNA-protein crosslinking data does not address whether or not the poly(A) RNAs bound to Npl3 in the cytoplasmic phase are competent for translation, nor does it address whether the specific identities of the mRNAs associated with Npl3 change throughout the cycle.

To investigate this latter question, we performed a series of RNA-protein co-immunoprecipitation microarray experiments under salt stress conditions. In this procedure, cellular RNAs associated with mRNA transport factors are co-immunoprecipitated from extract using protein-specific antibodies, and are then identified using a cDNA microarray. In this experiment, mRNAs co-immunoprecipitating with Npl3

were compared to the mRNAs which co-immunoprecipitate with Nab2 by competitive hybridization on a microarray. The mRNA population co-immunoprecipitating with Nab2 served as meaningful reference sample for several reasons. First, like Npl3, Nab2 is an abundant, shuttling, mRNA export factor. Second, a significant degree of transcript-specificity is seen between the profiles of steady-state mRNA populations associated with Npl3 and Nab2 (Kim-Guisbert et al. 2005). Finally, unlike Npl3, Nab2 shows no change in its localization or bulk poly(A) RNA binding properties following salt stress (data not shown).

In a time course investigating the changes in mRNAs associated with Npl3 versus Nab2 (Figure 30), we see very few changes in the profile of transcripts associated with Npl3 5 minutes after exposure to salt stress compared to the unstressed sample. This suggests that during the initial shift from nuclear to cytoplasmic localization there is no significant change in the population of mRNAs with which Npl3 is associated. At the 10 minute time point, however, we see a shift in Npl3 association for several classes of transcripts. Some transcripts shift from de-enriched in the Npl3 sample before the 10 minute time point to enriched (Figure 30A, B), while others show the converse behavior (Figure 30C). Importantly, this first category includes both transcripts which show a decrease in total expression levels in response to salt stress (Figure 30A), and those which show little to no change in expression levels (Figure 30B).

As previously reported (Kim-Guisbert, 2005), the broadest and most unique category of transcripts which show an association bias for Npl3 under normal growth conditions are the ribosomal protein genes (RPGs) (Figure 31A). Interestingly, however, 10 minutes after salt stress exposure, the RPG preference for Npl3 is largely lost (Figure 31B). Together the microarray data strongly suggest that, during the salt stress response, Npl3

remains associated with the normal complement of its steady state cargo through its initial transition to cytoplasmic localization, but that the composition of its cargo shifts as it enters the nuclear re-import phase.

Within the margins of a reasonable analysis of significance in this data set, there were no transcripts which showed an increase in expression levels in response to salt stress which also showed an appreciable enrichment or de-enrichment in the Npl3 sample.

Cation toxicity is sufficient to induce the transient re-localization of Npl3

Exposing yeast cells to a concentration of 500mM NaCl, used here and widely in the field to induce salt stress, results in both osmolarity stress and intra-cellular cation toxicity. In an effort to pin down whether the observed changes in Npl3 behavior under high salt conditions were due to cation toxicity, osmolarity stress, or a combination of the two we exposed cells to 300mM LiCl and re-assayed the localization behavior of Npl3. This concentration of LiCl induces a cation toxicity response but does not constitute an appreciable osmolarity stress. We see that cation toxicity alone is sufficient to induce the Npl3 re-localization cycle (Figure 32).

Cation toxicity induces the transient re-localization of Yra1 and Sub2.

To better understand whether or not the re-localization behavior of Npl3 in response to cation toxicity is unique, we assayed the localization of a number of other factors implicated in nuclear mRNA export after exposure to 300mM LiCl. Although several factors showed no significant change in their cellular localization (Mex67, Nab2, Pub1;

data not shown), both Sub2 and Yra1 exhibited a stress induced re-localization cycle nearly identical in behavior and kinetics to that observed for Npl3 (Figure 33, Figure 34).

Poly(A) RNA accumulates in nuclear speckles following salt stress

In previous work, bulk poly(A) RNA has been shown to accumulate in the nucleus following heat shock, a condition which causes changes in Npl3 localization which are similar to salt stress (Krebber et al., 1999). We therefore also assayed the bulk localization of poly(A) RNA following both 500mM NaCl and 300mM LiCl stress (Figure 35). Interestingly, these two stresses do not shift the signal in a poly(A) RNA in situ stain to a flat nuclear intensity. Instead, we see discrete speckles of staining, which overlap with the DAPI signal (not shown). Although this phenotype is not as striking as the accumulation of poly(A) RNA induced by heat shock in our test strain (Figure 35), it is suggestive of some form of aberration in mRNA export.

Perspectives

These results demonstrate several intriguing and complex changes in the behavior of the nuclear mRNA export machinery following salt stress. The localization and RNA binding behavior of Npl3 requires an update to the previous model of Npl3 function under salt stress conditions: immediately following salt stress Npl3 shifts its bulk localization from the nucleus to the cytoplasm, maintaining association with its normal RNA cargo. After 10 minutes, we begin to see re-accumulation of Npl3 in the nucleus, concomitant with a significant change in the identities of transcripts with which it is associated. The biological significance of this change, however, remains unclear. The

pool of RNAs which are associated with Npl3 after the 10 minute time point are neither enriched nor de-enriched for factors known to be involved in the salt stress response, or for transcripts whose total expression levels appear to be regulated by the response. Although if a transcript were to be regulated at the level of association with the export machinery and export competence we might not expect its total level of expression to also be regulated.

The finding that both Sub2 and Yra1 also re-localize to the cytoplasm after salt stress, with kinetics that are nearly identical to those exhibited by Npl3, is surprising for two reasons. First, like Npl3 these factors are both presumed to be core components of the major Mex67-mediated export pathway. Second, unlike Npl3, neither are thought to shuttle between the nucleus and cytoplasm in the course of the mRNA transport cycle.

Although these preliminary data do not lend themselves to making specific predictions regarding the biological or mechanistic implications of these stress-induced changes in transport factor localization and mRNA binding profiles, the magnitude and rapidity of these changes are strongly suggestive of a role for modulation of the mRNA transport machinery in the response to cellular stress. I am optimistic that future work will be able to contextualize these findings and begin to reveal the dynamics and extents of the interplay between nuclear mRNA export and the environment, as our work in the preceding chapters has further revealed the interplay between splicing and the environment.

One potentially interesting question to ask would be what aspects of the transport cycle are modulated in response to stress to cause a large scale re-localization of these factors. Npl3, for example, has a well documented cycle of phosphorylation state that is

correlated with each stage of the shuttling cycle. It would be interesting to identify which steps in this pathway are altered in response to stress, and to investigate the growth survivorship implications of those changes. It would also be very interesting to identify the fate of mRNAs bound to Npl3 through the cytoplasmic phase of the stress-induced re-localization cycle. A first question would be to ask whether or not the levels of Npl3 association with polysomes changes after stress exposure. Their translational competence of individual transcripts could be assayed by polysome microarray and these data compared to the Npl3-RNA co-immunoprecipitation microarray data.

Materials and Methods

Localization of proteins by immunofluorescence

Cells were grown at least 2 doublings from a overnight starter culture to early log phase (OD 0.2 – 0.5) at 30°C before exposure to stress conditions. At the start of each time point cells were fixed and stained as previously described (deBruyn Kops and Guthrie, 2001). Images were collected with an Olympus BX-60 microscope outfitted with a Sensys CCD camera (Photometrics).

Purification of proteins UV-crosslinked to poly(A) RNA

Cells were grown at least 2 doublings from a overnight starter culture to early log phase (OD 0.2 – 0.5) at 30°C before exposure to stress conditions. In vivo crosslinking, crosslinked extract preparation, and poly(A) RNA purification were performed as previously described (Gilbert et al., 2001). At the start of each time point cells were

collected by vacuum filtration through a nitrocellulose filter, and washed with phosphate buffer cooled to 4°C. Filters rapidly transferred to dishes cooled on ice, and exposed to UV crosslinking in a procedure that consumed less than 2 minutes time from initial vacuum filtration to UV exposure. Purified poly(A) RNPs were analyzed by SDS-PAGE and immunoblotting.

Localization of poly(A) RNA by in situ hybridization.

Cells were grown at least 2 doublings from a overnight starter culture to early log phase (OD 0.2 – 0.5) at 30°C before exposure to stress conditions. At the start of each time point cells were fixed and bulk poly(A) RNA was localized by FISH as previously described in (deBruyn Kops and Guthrie, 2001). Images were collected with an Olympus BX-60 microscope outfitted with a Sensys CCD camera (Photometrics).

RNA-protein co-immunoprecipitation Microarrays

Cells were grown at least 2 doublings from a overnight starter culture to early log phase (OD 0.2 – 0.5) at 30°C before exposure to stress conditions. Immunoprecipitations, RNA preparation, cDNA labeling, microarray hybridization, and microarray platform are described in Kim-Guisbert, et al. 2005.

References

Carmo-Fonseca, M. (2002). The contribution of nuclear compartmentalization to gene regulation. *Cell* **108**, 513-521.

de Bruyn Kops, A. and Guthrie, C. (2001). An essential nuclear envelope integral membrane protein, Brr6p, required for nuclear transport, *EMBO J.* **20** (2001), pp. 4183–4193.

Gerber, A.P., Herschlag, D, Brown. P.O. (2004). Extensive Association of Functionally and Cytotopically Related mRNAs with Puf Family RNA-Binding Proteins in Yeast. *PLoS Biol* 2(3): e79

Gilbert, W., Siebel, C.W. and Guthrie, C., (2001). Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. *RNA* **7**, pp. 302–313.

Hieronymus, H. and Silver, P.A. (2003). Genome-wide analysis of RNA–protein interactions illustrates specificity of the mRNA export machinery. *Nat. Genet.* **33**: 155–161.

Kim-Guisbert, K., Duncan, C., Li, H., and Guthrie, C. (2005). Functional specificity of shuttling hnRNPs revealed by genome-wide analysis of their RNA binding profiles. *RNA* **11**, pp. 383-393.

Krebber, H., T. Taura, M.S. Lee and P.A. Silver. (1999). Uncoupling of the hnRNP Npl3p from mRNAs during the stress-induced block in mRNA export. *Genes & Dev.* **13**, 1994-2004.

Liu, Y., S. Liang and A.M. Tartakoff. (1996). Heat shock disassembles the nucleolus and inhibits nuclear protein import and poly(A)⁺ RNA export. *EMBO J.* **15**, 6750-6757.

Saavedra, C., K.S. Tung, D.C. Amberg, A.K. Hopper and C.N. Cole. (1996). Regulation of mRNA export in response to stress in *Saccharomyces cerevisiae*. *Genes & Dev.* **10**, 1608-1620.

Saavedra, C.A., C.M. Hammell, C.V. Heath and C.N. Cole. (1997). Yeast heat-shock mRNAs are exported through a distinct pathway defined by Rip1p. *Genes & Dev.* **11**, 2845-2856.

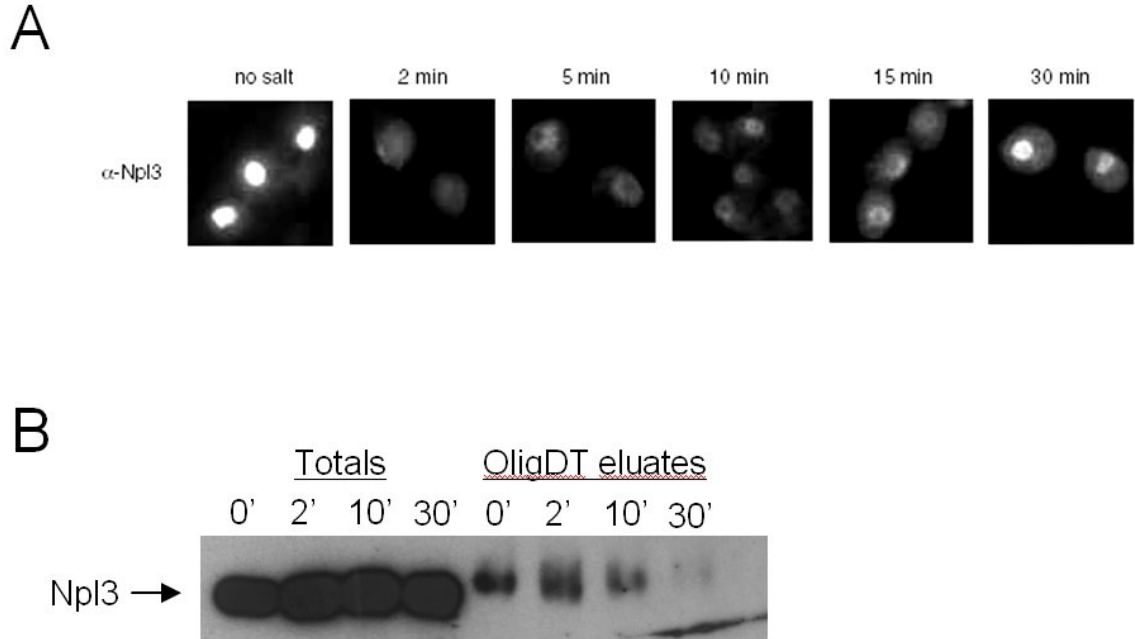


Figure 29. Changes in Npl3 localization and poly(A) RNA association following salt stress.

(A) The bulk localization of Npl3 shifts from the nucleus to the cytoplasm immediately following NaCl treatment, shows nuclear rim staining at intermediate time points, and ultimately re-accumulates in the nucleus 30 minutes after stress exposure. Wild-type W303 cells were grown to early log phase in YPD, treated with 500 mM NaCl and then fixed at the specified time points and stained using a polyclonal anti-Npl3 antibody (1:2,000). (B) A matched time course of cells collected for the poly(A) RNA-protein cross-linking assay shows that association of Npl3 with poly(A) RNA does not appreciably change until the nuclear re-accumulation phase of this cycle. In this assay cells were rapidly collected by vacuum filtration and exposed to UV cross-linking at the times specified. Levels of Npl3-copurified with poly(A) RNA on an oligo(dT) cellulose matrix were then visualized by western.

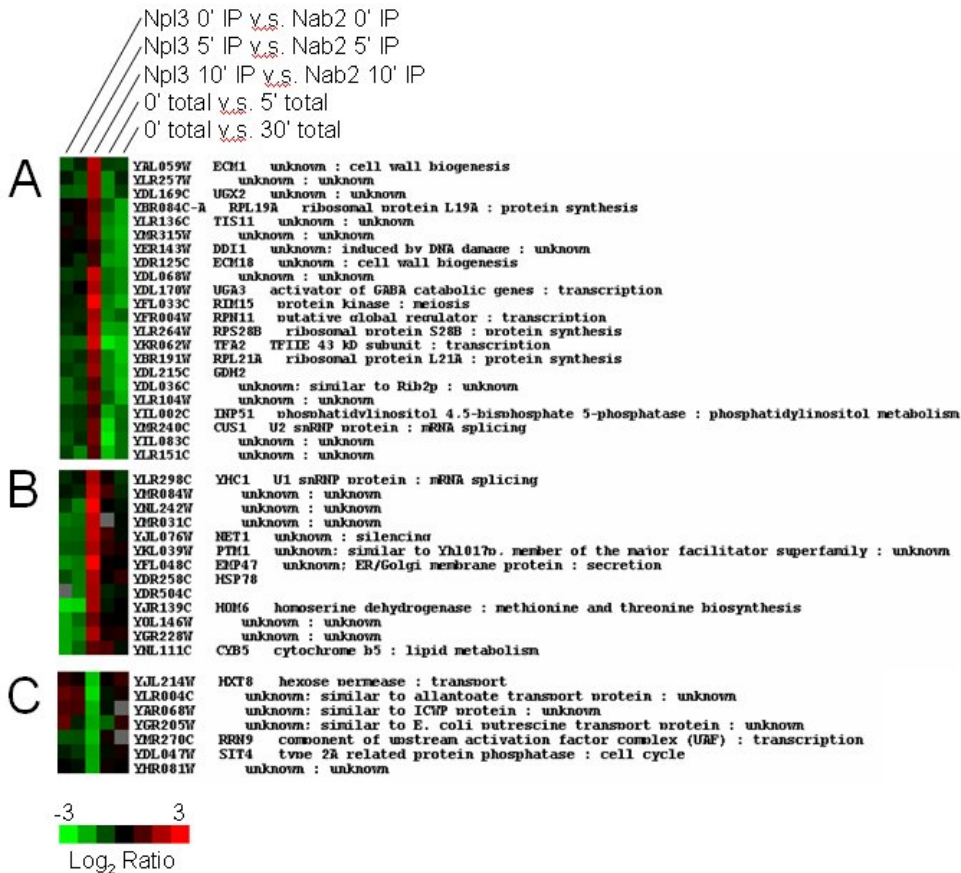


Figure 30. Changes in mRNA association with Npl3 versus Nab2 following salt stress.

From left to right columns in this cluster display enrichment of transcripts in an Npl3 co-IP (high values) versus Nab2 co-IP (low values) before treatment with 500mM NaCl (Npl3 0' IP v.s. Nab2 0' IP), 5 minutes following treatment (Npl3 5' IP v.s. Nab2 5' IP), 10 minutes following treatment (Npl3 10' IP v.s. Nab2 10' IP), enrichment in the total RNA fraction 5 minutes after treatment versus total RNA from the untreated sample (0' total v.s. 5' total), and enrichment in the total RNA fraction 10 minutes after treatment versus total RNA from the untreated sample (0' total v.s. 10' total). (A) Transcripts which display a decrease in total levels and transition from enrichment in the Nab2 IP to the Npl3 IP 10 minutes after 500mM NaCl treatment. (B) Transcripts which display no significant change in total levels and transition from enrichment in the Nab2 IP to the

Npl3 IP 10 minutes after 500mM NaCl treatment. (C) Transcripts which display no significant change in total levels and transition from enrichment in the Npl3 IP to the Nab2 IP 10 minutes after 500mM NaCl treatment.

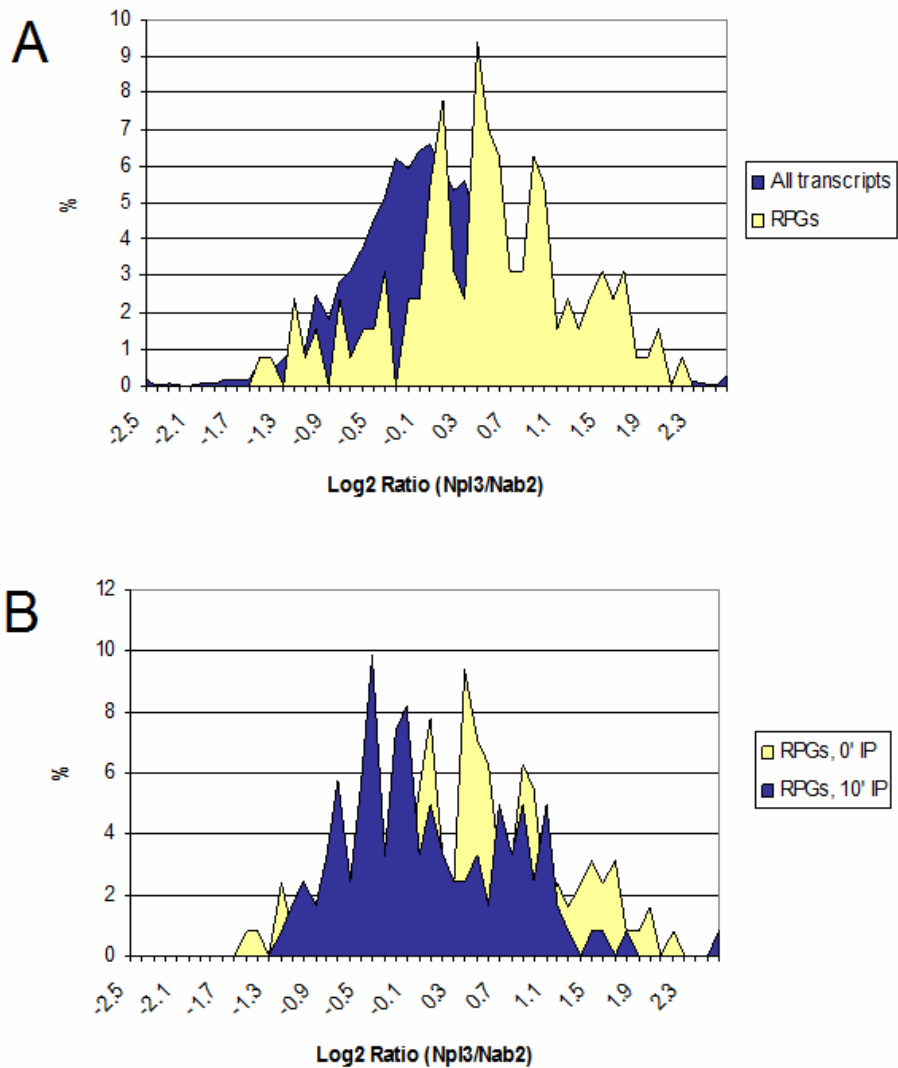


Figure 31. Changes in the association of ribosomal protein gene transcripts with Npl3 and Nab2 following salt stress.

These histograms show the distribution of ratios of Npl3 (high values) versus Nab2 (low values) transcript association in the RNA co-IP microarray experiment. (A) Under normal growth conditions, although the bulk of transcripts show a normal distribution along the Npl3 versus Nab2 association axis (blue histogram), ribosomal protein gene transcripts (RPGs, yellow histogram) exhibit an enrichment in the Npl3 IP. (B) By contrast, 10 minutes after exposure to 500mM NaCl RPG transcript preference for Npl3

is lost (blue histogram). In this panel the RPG distribution in unstressed cells is repeated for reference (yellow histogram).

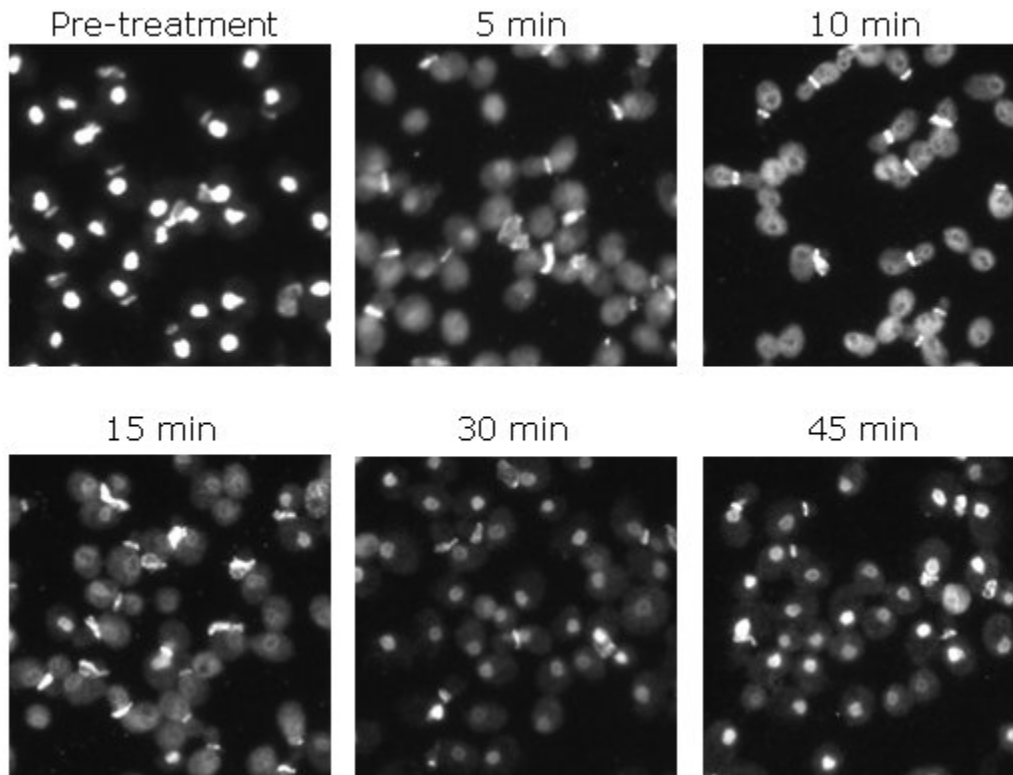


Figure 32. Changes in Npl3 localization following 300mM LiCl treatment.

Wild-type W303 cells were grown to early log phase in YPD, treated with 300 mM LiCl and then fixed at the specified time points and stained using a polyclonal anti-Npl3 antibody (1:2,000).

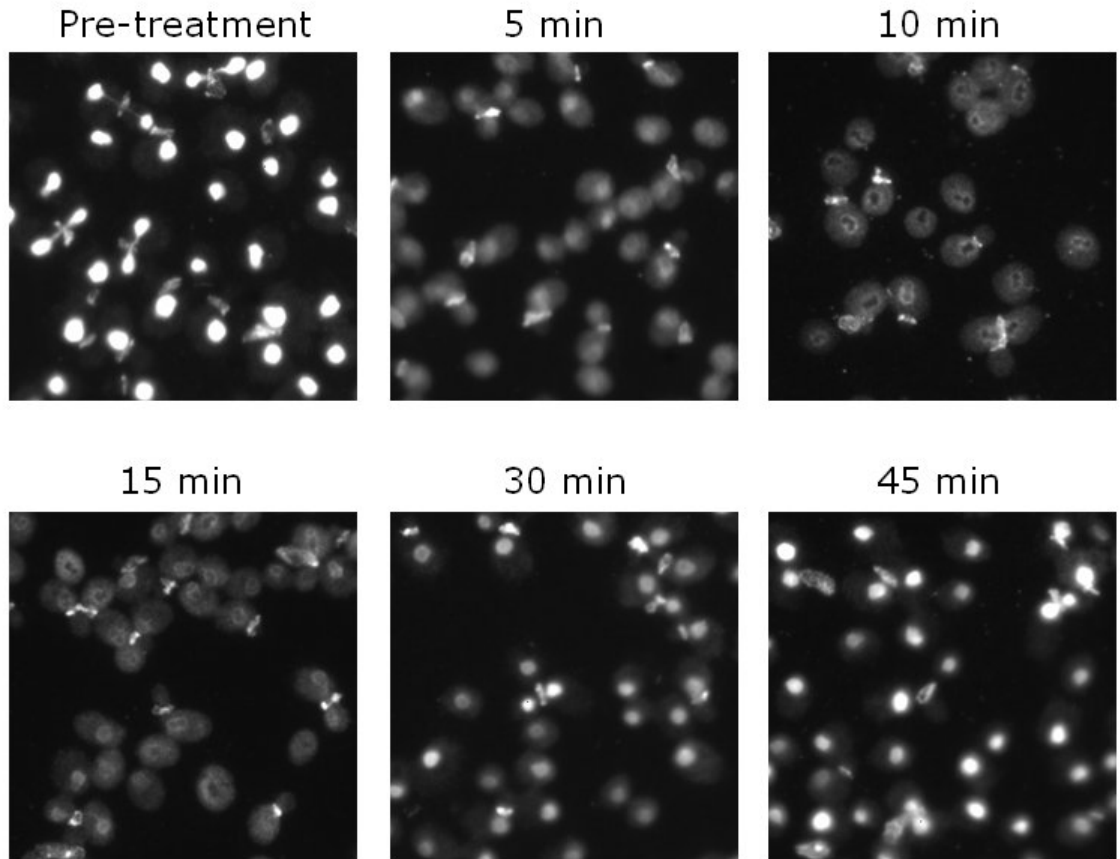


Figure 33. Changes in Sub2 localization following 300mM LiCl treatment.

Wild-type W303 cells were grown to early log phase in YPD, treated with 300 mM LiCl and then fixed at the specified time points and stained using a polyclonal anti-Sub2 antibody (1:1,000).

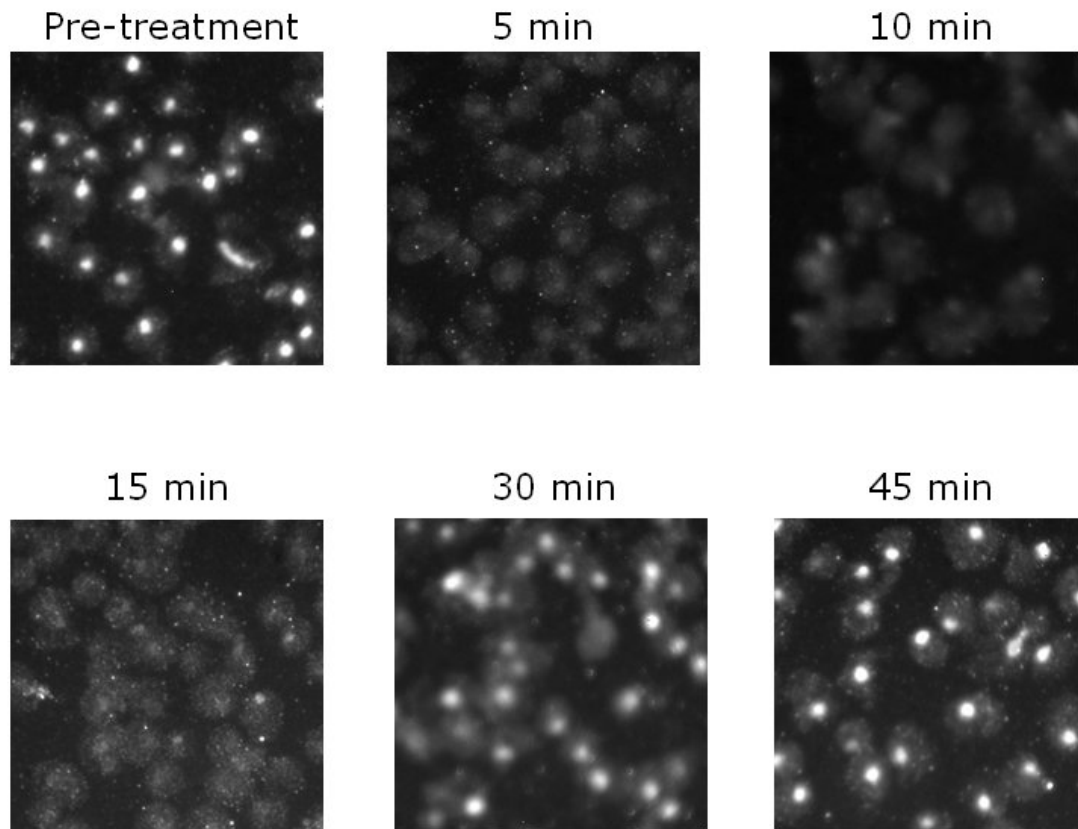


Figure 34. Changes in Yra1 localization following 300mM LiCl treatment.

Wild-type W303 cells were grown to early log phase in YPD, treated with 300 mM LiCl and then fixed at the specified time points and stained using a polyclonal anti-Yra1 antibody (1:2,000).

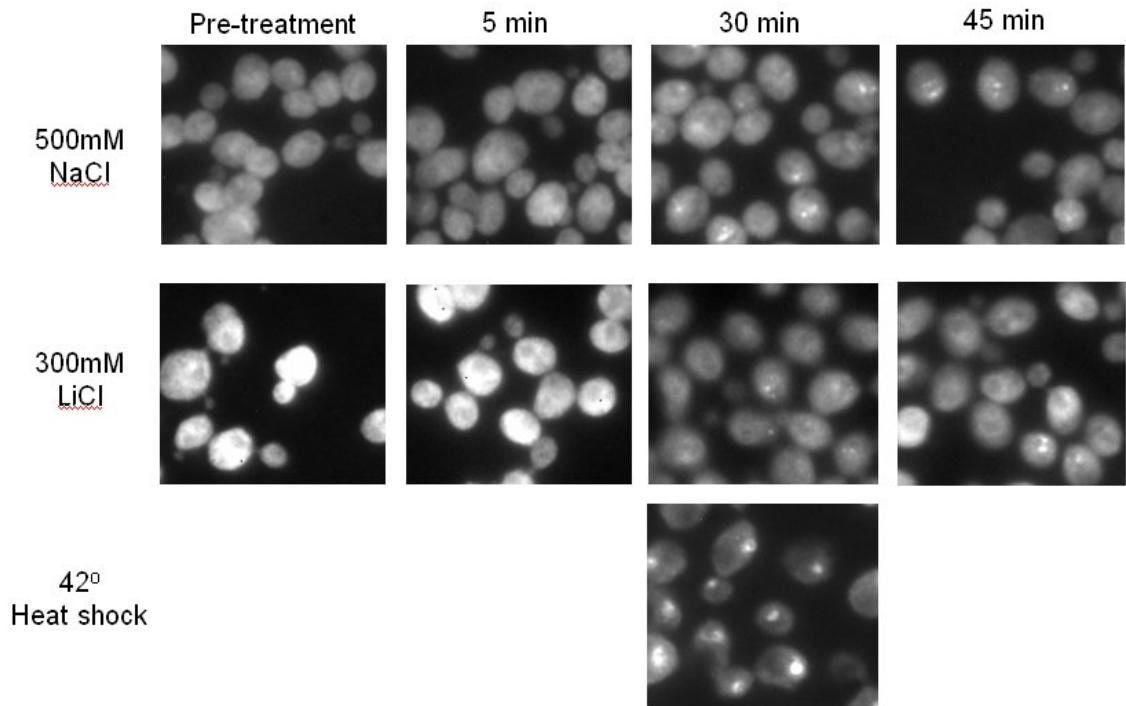


Figure 35. Changes in poly(A) RNA localization following treatment with NaCl, LiCl and heat shock.

Localization of poly(A) RNA in wild-type cells was visualized by in situ hybridization with a FITC-labeled oligo(dT)₅₀ probe following treatment with 500mM NaCl, 300 mM LiCl, and 40°C heat shock. WT W303 cells were grown to early log phase in YPD and fixed at the time points specified following initial exposure to each stress condition.

Appendix 3: Screen for Environmental Stress-Induced Growth
Phenotypes Caused by Mutations in mRNA Processing Factors

Preface

In an effort to identify specific components of the mRNA processing machinery with biologically important roles in environmental stress responses, I screened the growth phenotypes of ~350 deletions and conditional alleles in factors involved in mRNA metabolism under 26 different environmental conditions at 3 different growth temperatures. Although this effort pre-dated the development of quantitative growth phenotyping strategies, it serves as a potentially valuable database of phenotypes to use as a starting point for identifying factors with roles in both particular post-transcriptional stages in mRNA processing and environmental stress responses. The most significant phenotypes resulting from this screen are summarized in the figures that follow.

Perspective: the puzzle of cation toxicity resistance

One of the most striking set of phenotypes to emerge from this screen came from a set of deletions in non-essential components of the spliceosome (Figure 36). Each of these four mutants exhibited an increase in growth rate under cation toxicity conditions, relative to wild-type cells. The first two factors, *Isy1* and *Snt309* are both components of the Prp19-complex. The third, *Mud1*, is the yeast homolog of human U1-A and a component of the U1 snRNP. The final factor *Nst1*, although relatively uncharacterized, has been shown in several studies to have physical interactions with *Msl1*, a non-essential component of the U2 snRNP (Goossens et al., 2002). One allele of *Sub2*, *sub2-4* also showed a significant resistance to cation toxicity, as did deletion of *Nup188*.

Previously, the only well characterized mutation in yeast which lead to cation toxicity resistance was deletion of Sky1, the yeast SR-protein kinase homolog (Forment, 2002). Interestingly, one target of the Sky1 kinase, the yeast SR-protein Npl3, has recently been shown to play a role in the splicing efficiency of yeast ribosomal protein genes (Tracy Kress, work in progress). We investigated the stress phenotypes of deletions in the two other kinases in yeast most closely related to Sky1 (Yak1 and Kns1), but did not find any evidence of cation toxicity resistance in these strains (Figure 37).

So what are possible mechanisms to explain the deletion of a particular factor leading to an apparent increase in cellular fitness under cation toxicity stress? One explanation could be that the mutation predisposes cells to cope with the stress, for example by constitutively activating an adaptive pathway. In the case of cation toxicity one potential adaptation could be the constitutive up-regulation of cation efflux systems. However, in the case of the deletion in Sky1, there are several pieces of experimental evidence that make this model unlikely. First, our own microarray data show no appreciable change in the constitutive level of transcripts associated with any known components of the ion homeostasis machinery. Second, work by Wendy Gilbert in the Guthrie lab, and others (Forment, 2002), has demonstrated that intracellular cation levels are similar levels in wild-type cells and cells carrying a deletion of Sky1 following LiCl exposure.

An alternative and compelling model is that these mutations exhibit an apparent increase in fitness in the presence of high cation concentrations because cells fail to induce response pathways which are effectively repressive under laboratory conditions, but useful in more natural contexts. Specifically, high cation concentrations have been shown to induce programmed cell death in both yeast and plants (Huh et al., 2002); up to 30% of the cells in yeast cultures exposed to toxic levels of cation have shown a loss

of cell viability and nuclear DNA fragmentation characteristic of apoptosis. It would be fascinating to ask if the apparent cation toxicity resistance of these mutations in mRNA processing factors stems from a failure to induce the apoptotic response at wild-type levels.

Materials and Methods

This screen was performed by growing cultures in 2mL of media in 96-well format. A liquid handling robot was used to repeatedly back dilute cultures so that faster growing strains could be maintained in log phase growth. After dilution from an overnight stock each strain underwent between 2 and 8 doublings over the course of a grow out phase at 30°C, with cultures being maintained at a maximum OD of 0.3. Cultures were then all back diluted to an OD of 0.1, 0.016, and 0.0026 (2 X 6x serial dilutions), and robotically spotted onto media containing different stress agents.

References

Forment, J., Mulet, J.M., Vicente, O., and Serrano, R. (2002). The yeast SR protein kinase Sky1p modulates salt tolerance, membrane potential and the Trk1,2 potassium transporter. *BBA* **1565**: pp. 36-40.

Goossens, A., Forment, J., Serrano, R. (2002). Involvement of Nst1p/YNL091w and Msl1p, a U2B" splicing factor, in *Saccharomyces cerevisiae* salt tolerance. *Yeast* **19**: pp. 193-202.

Huh, G.H., Damsz, B., Matsumoto, T.K., Reddy, M.P., Rus, A.M., Ibeas, J.I., Narasimhan, M.L., Bressan, R.A., and Hasegawa, P.M. (2002). Salt causes ion disequilibrium-induced programmed cell death in yeast and plants, *Plant J* **29**: pp. 649–659.

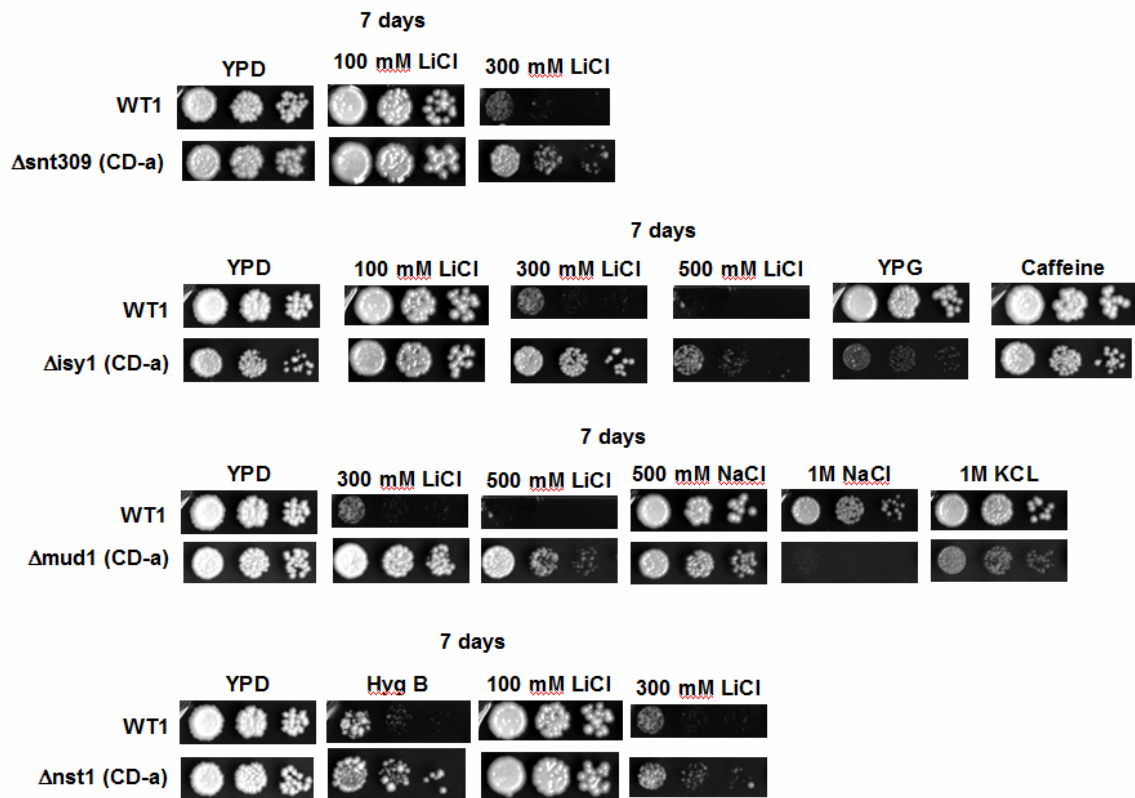


Figure 36. Stress phenotypes of deletions of non-essential splicing factors.

These images were taken after 7 days of growth at 30°C. The phenotypes of each deletion mutant (bottom rows) are shown in comparison to a matched wild-type strain (top rows).

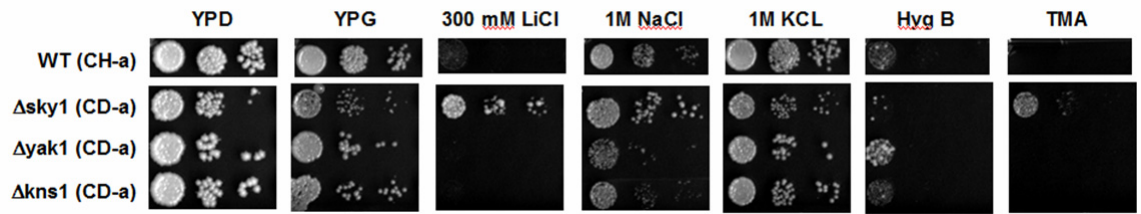


Figure 37. Stress phenotypes of deletions of Sky1 and closely related kinases.

These images were taken after 4 days of growth at 30°C. The phenotypes of each deletion mutant (bottom rows) are shown in comparison to a matched wild-type strain (top row).

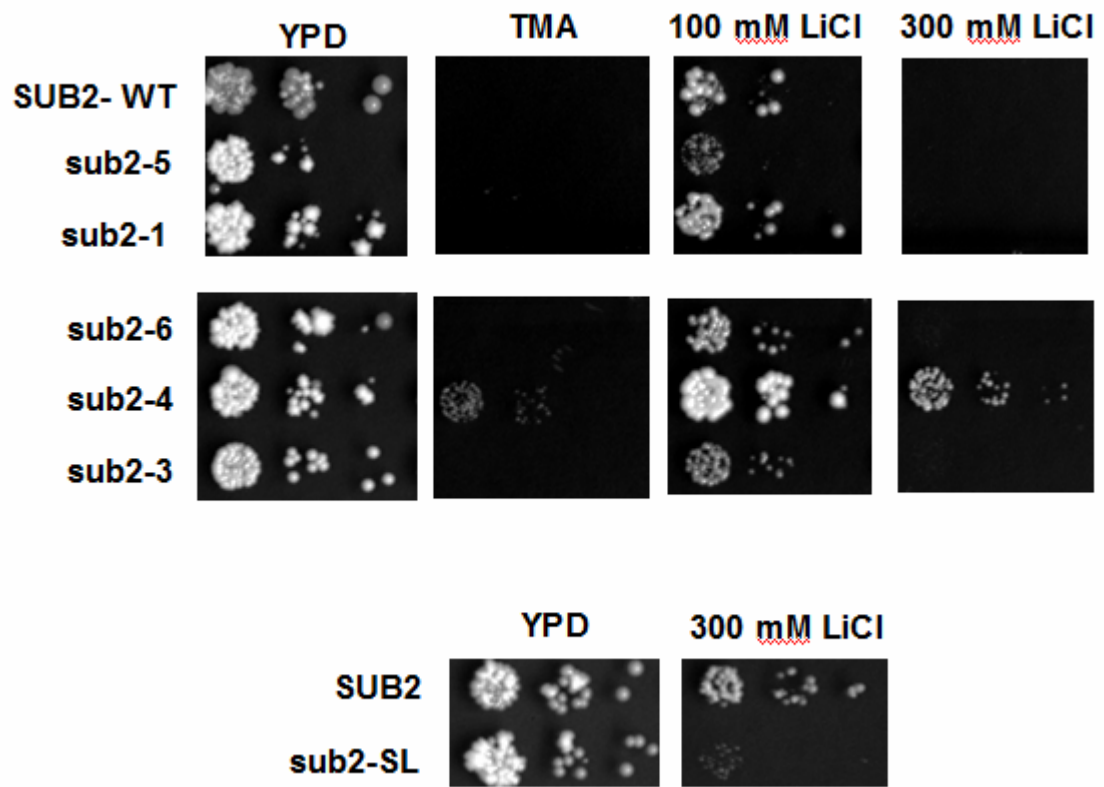


Figure 38. Allele specific stress phenotypes of Sub2 mutants.

These images were taken after 7 days of growth at 30°C. The phenotypes of each mutant (bottom rows) are shown in comparison to a matched wild-type strain (top rows).

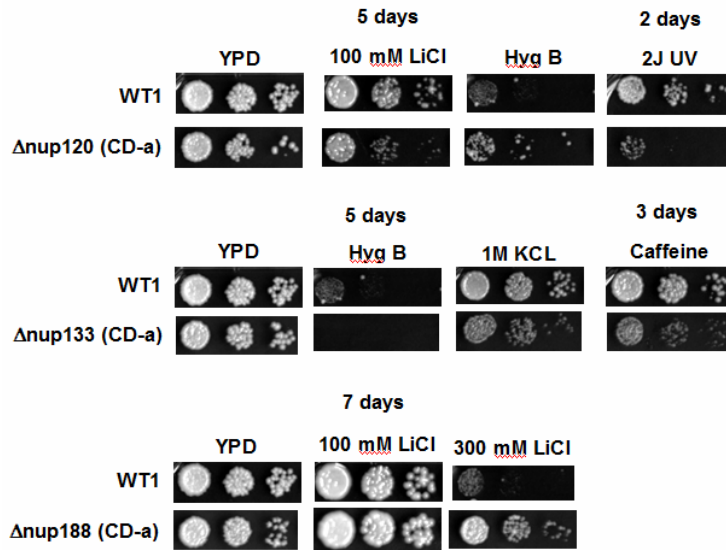


Figure 39. Stress phenotypes of deletions of nuclear pore components.

These images were taken after growth at 30°C for the number of days specified above each panel. The phenotypes of each deletion mutant (bottom rows) are shown in comparison to a matched wild-type strain (top rows).

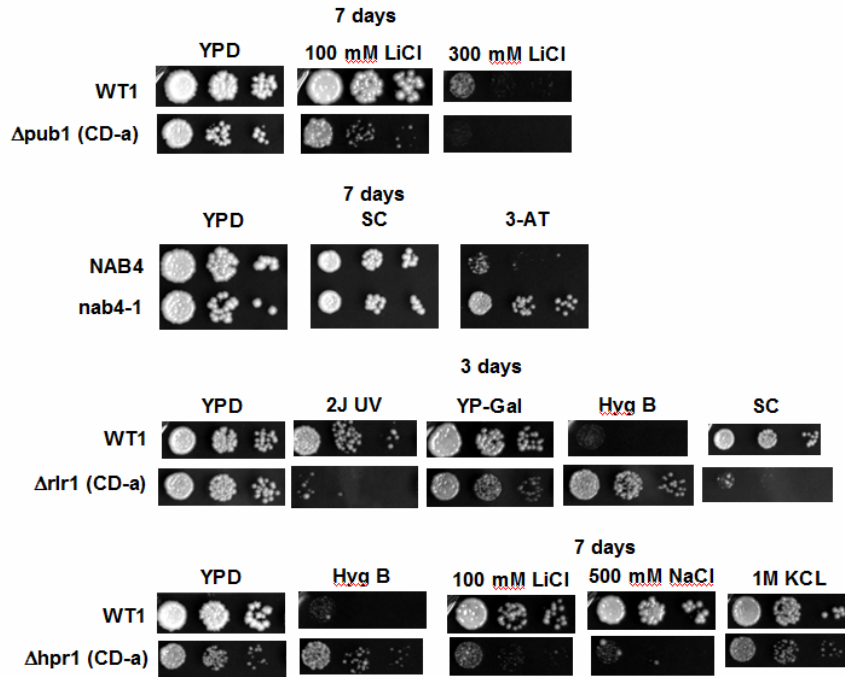


Figure 40. Stress phenotypes of factors implicated in mRNA transport and 3' end processing.

These images were taken after growth at 30°C for the number of days specified above each panel. The phenotypes of each deletion mutant (bottom rows) are shown in comparison to a matched wild-type strain (top rows).

Appendix 4: Microarray Data Analysis Toolkit

Preface

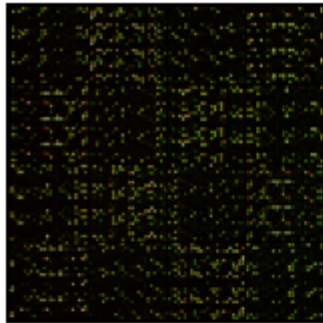
Our custom splicing-specific microarray platform has proven to be an incredibly valuable experimental tool. Development of this array platform presented us with several considerable challenges. One of the most important of these was developing a suite of data analysis techniques and supporting software that allowed us to make full use of the complexity of data produced by these arrays. Our splicing-specific microarrays, like many emerging 3rd generation microarray technologies that probe species-specific RNA variants, add an additional dimension of data complexity over expression microarrays. On traditional array platforms the behavior of each gene is represented by a single measurement under each individual experimental factor. By contrast on our arrays each gene is represented by 3 measurements: the behavior of the total mRNA population, pre-mRNA species, and mature mRNA species. Although early passes at analyzing splicing microarray data attempted to conform the data to the mold of a traditional expression array by condensing this primary dimension of the data by representing splicing behavior as “indexes” that were the ratios of the values of individual probe type measurements, this manipulation fundamentally discards important information and obscures the full complexity of splicing responses. So the challenge for us was to find new ways to represent and explore the data while maintaining each of its primary dimensions.

Although most of the higher-order analysis strategies we have developed to visualize the splicing microarray data are exemplified in the preceding chapters of this thesis, in the figures which follow we will briefly review some of the data analysis methods which have proven particularly useful and illuminating over the course of our work. Obviously because of the scale of microarray datasets in general, and our extremely large

collection of experimental factors, the data analysis side of this project has necessitated a significant amount of custom software development. Our in-house software solution is currently in its third generation. What began as a simple relational database strung together with a collection of ad hoc scripts has grown into a prototype for a fairly sophisticated data analysis framework. In the most recent incarnation of the framework we have tried to design the software suite around many of the currently vogue “web 2.0” concepts, including building a modular, component driven architecture. Ultimately this side of the software project has been an experiment in how to design a framework which creates the shortest possible path for the biologist between an abstract definition of a desired data analysis manipulation and a functional software implementation of that manipulation. The desire for such an architecture has recently arisen independently in several different commercial and open-source software projects. Whether or not we continue to evolve our custom framework, or transition our efforts to another project largely depends on the timeline of when a particular architecture emerges as the clear community standard.

I am optimistic that the time is ripe for major advances in how we as scientists implement new data analysis procedures and interact with large, complex biological datasets.

For software and technical documentation please see the project at its home on the web:
<http://code.google.com/p/gbwida/>



Two-color image from competitive hybridization of two samples on an array

Image analysis
(ex GenePix)

Block	Column	Row	ID	X	Y	Dia	F635	Median
1	1	1	1 YGL226C-A - intron	2170	9810	150	465	
1	2	1	1 YGR148C - intron	2360	9780	70	474	
1	3	1	1 YHR016C - intron	2540	9810	150	408	
1	4	1	1 YHR141C - exon	2720	9810	150	426	
1	5	1	1 YL133C - exon	2900	9810	150	405	
1	6	1	1 YJL191W - exon	3090	9810	150	408	
1	7	1	1 YOL232W - intron	3270	9810	150	481	
1	8	1	1 YGR183C - intron	3450	9810	150	558	
1	9	1	1 YHR021C - intron	3640	9810	150	454	
1	10	1	1 YHR203C - exon	3820	9810	150	505	

Ida Applications.Map

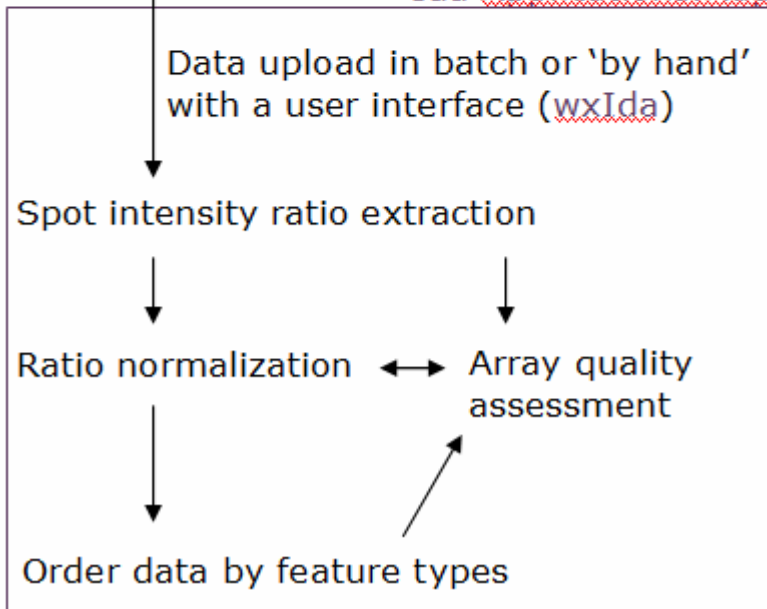


Figure 41. Schematic representation of the flow of information through our microarray data analysis procedure.

The names of specific software components are highlighted in purple.

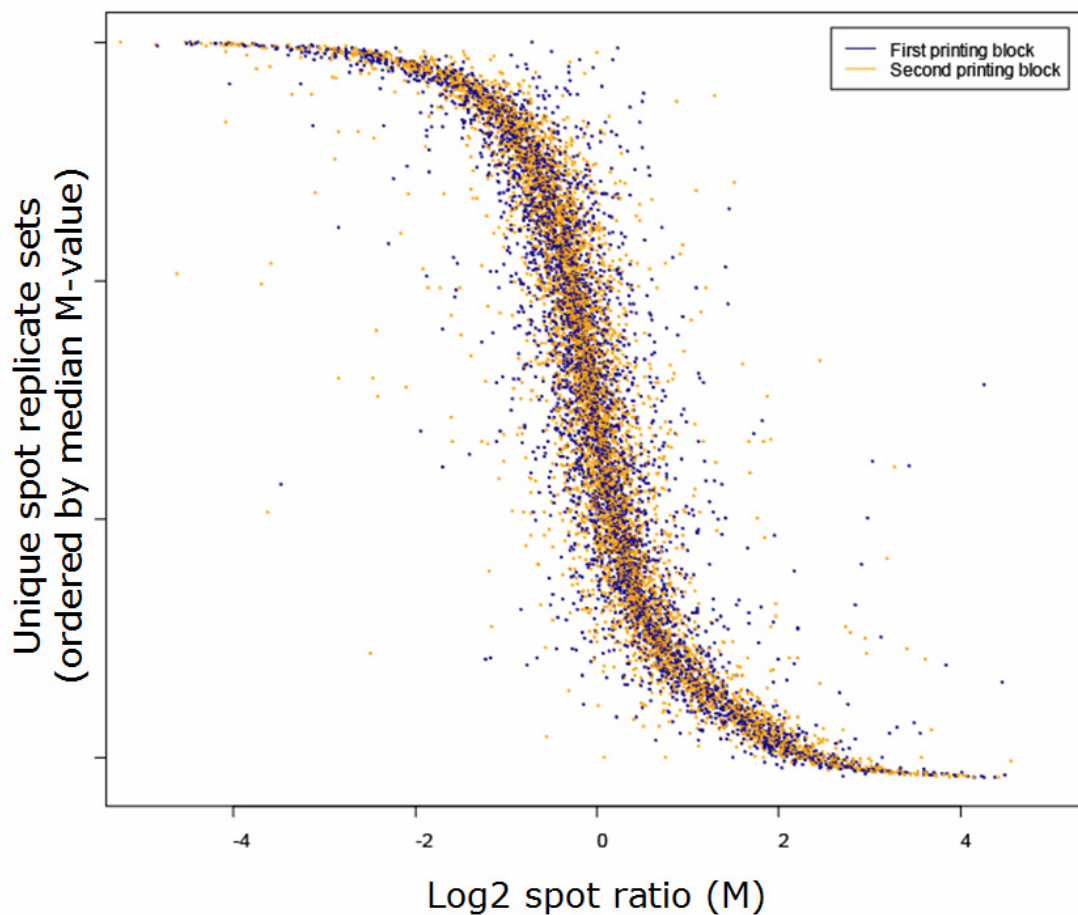


Figure 42. Dot plot of replicate spots.

Each horizontal line on this graph plots the M values for as many as six replicate spots on our splicing arrays. These lines of spots are arranged by median M value from lowest (top), to highest (bottom). Because spot replicates are printed as 3 replicates in each of 2 printing blocks, the points are colored to represent the relative block in which the spot was printed. This plot is an example assessment of within-array spot replicate agreement.

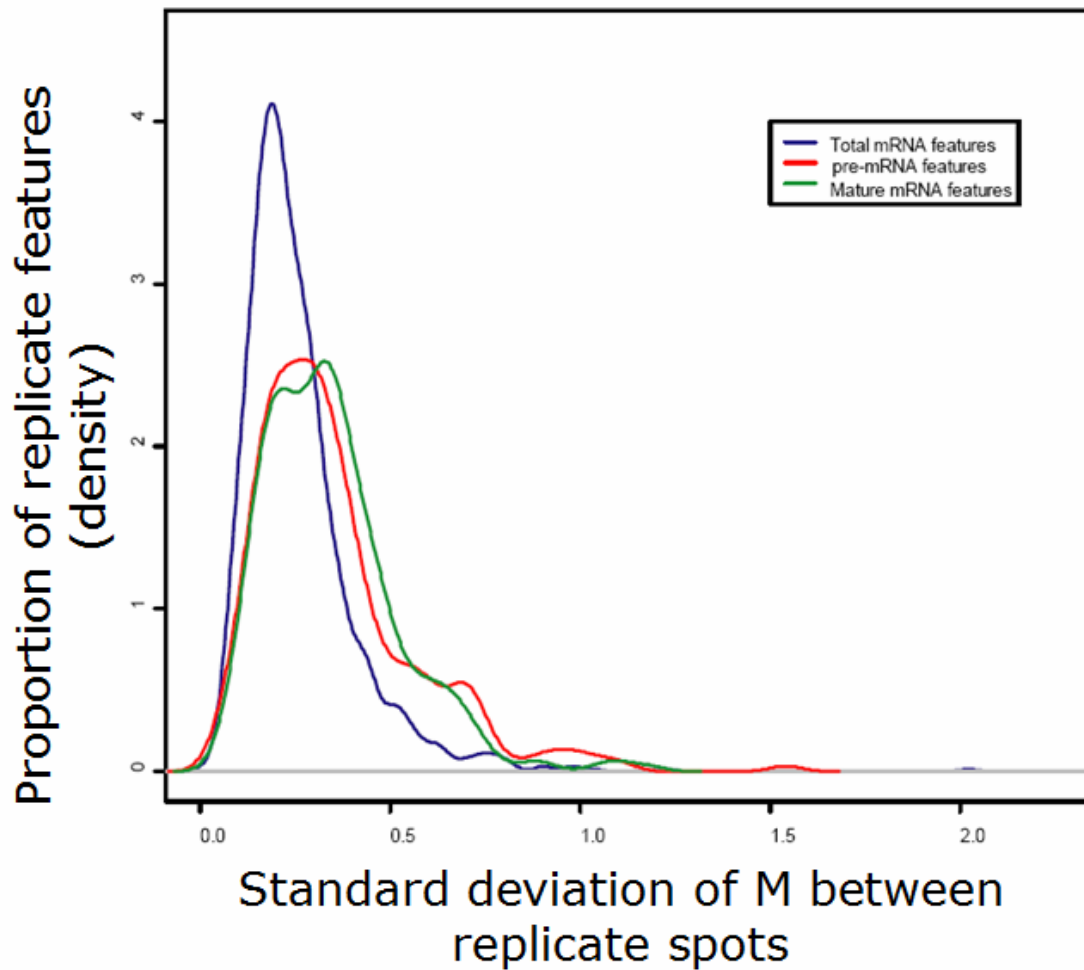


Figure 43. Variation of M values for spot replicates on a single array.

These plots show the distribution of standard deviations among spot replicate M values for data from a single array. Traces are colored by feature type. This plot is another example of assessment of within-array spot replicate agreement. Here we can evaluate the precision of each of the different types of feature probes.

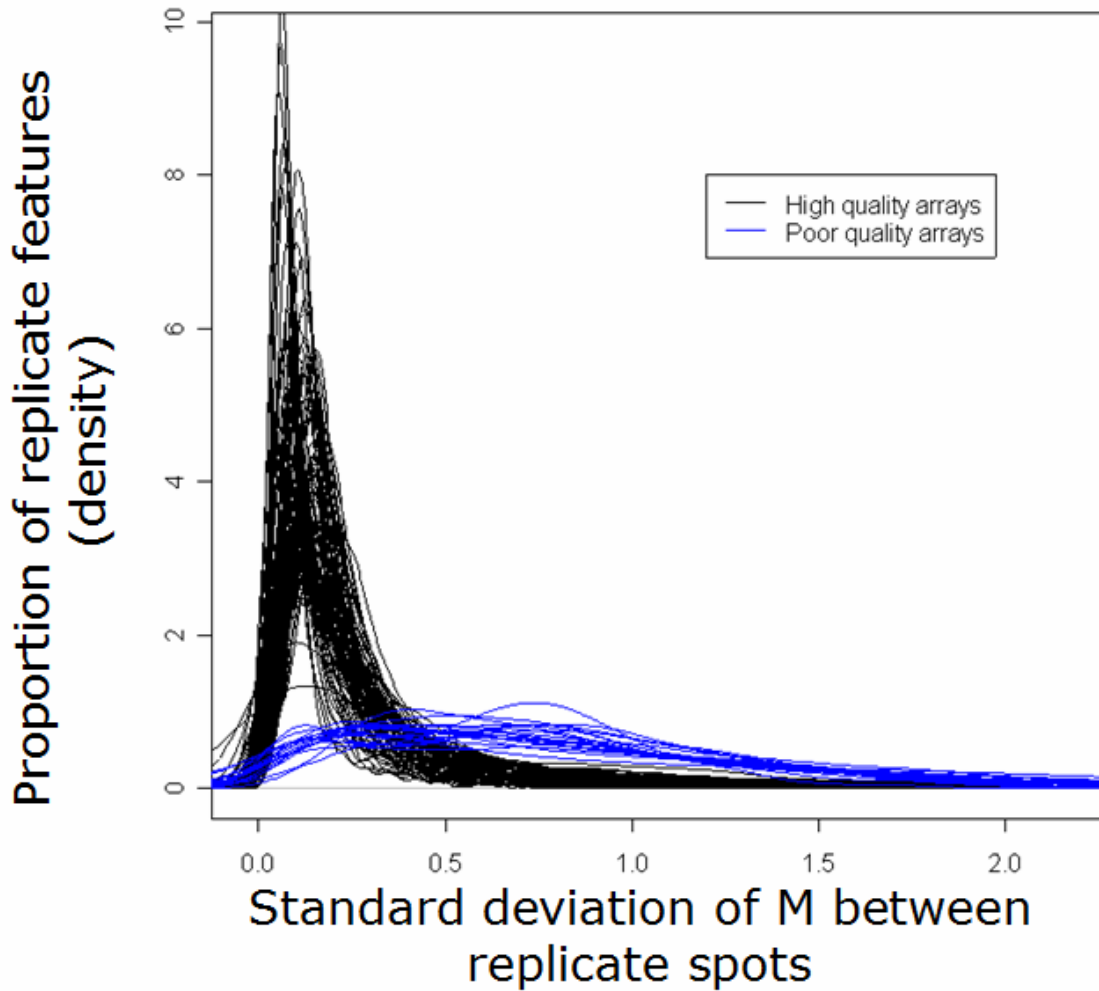


Figure 44. Variation of M values for spot replicates between arrays.

These plots show the distribution of standard deviations among spot replicate M values for data from many arrays. Traces are colored by the array quality flag. This plot is another example of within-array spot replicate agreement, but as viewed across a large number of arrays. This analysis allows us to use within-array replicate agreement as a marker of overall array quality.

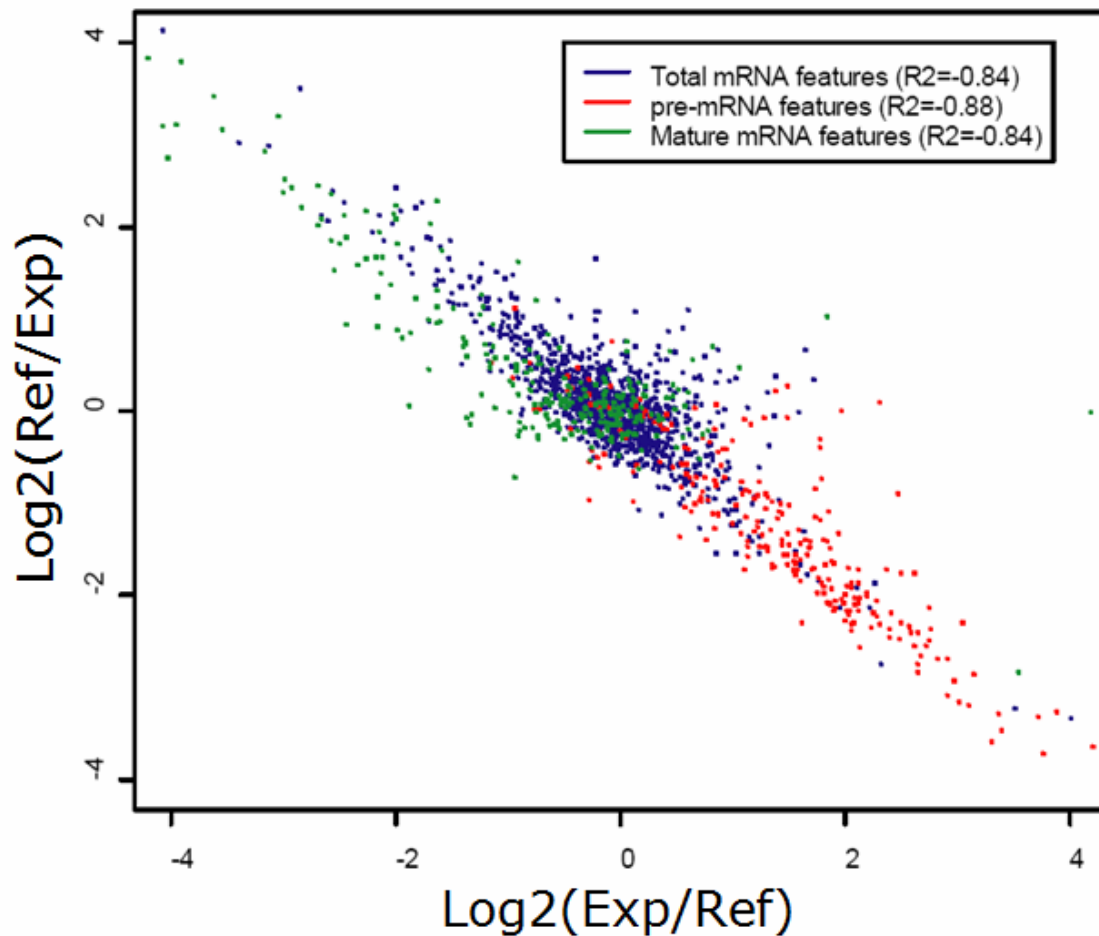


Figure 45. Correlation plot for dye-flipped replicates.

Here the ratio values for 2-dye flipped biological replicates are each plotted on an axis. Points are colored according to feature type. This plot is an example assessment of between-array replicate agreement. Because these experiments are dye-flipped with respect to each other, ideal replicates would show perfect anti-correlation.

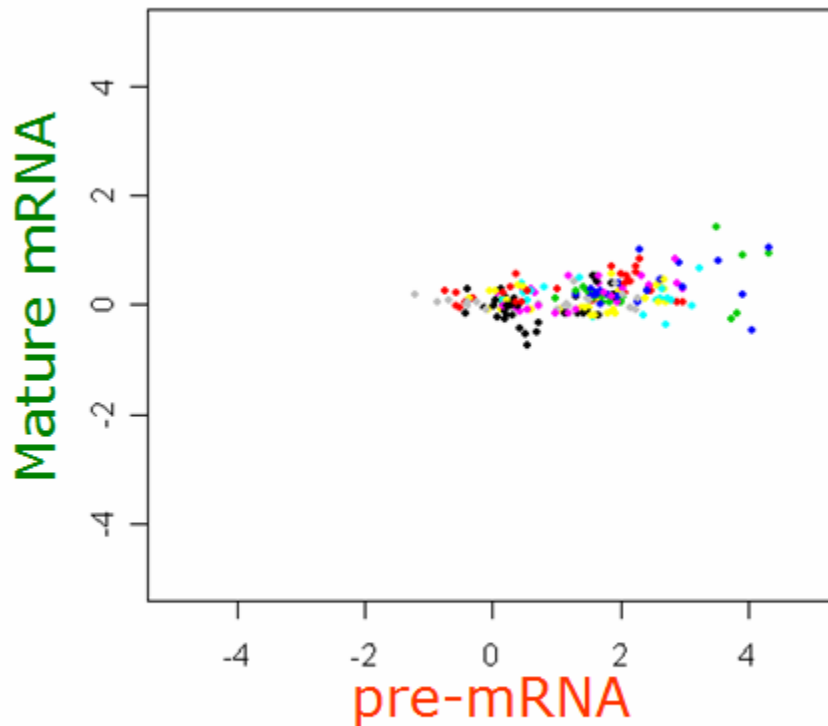


Figure 46. Feature type comparison for a given transcript.

These plots compare the M value behavior of two linked features on the array across experiments. This example shows us a clearly non-functional mature probe. These plots allow us to flag specific features on the array as unlikely to contain useful information. In general, when we plot data across a very large number of experiments, we expect these graphs to show two sub-populations values: (1) showing correlation between the behavior of both feature type, representing experiments in which changes to total transcript levels dominated, and (2) those showing anti-correlation between the mature mRNA and pre-mRNA probes, representing experiments in which splicing efficiency was inhibited or enhanced.

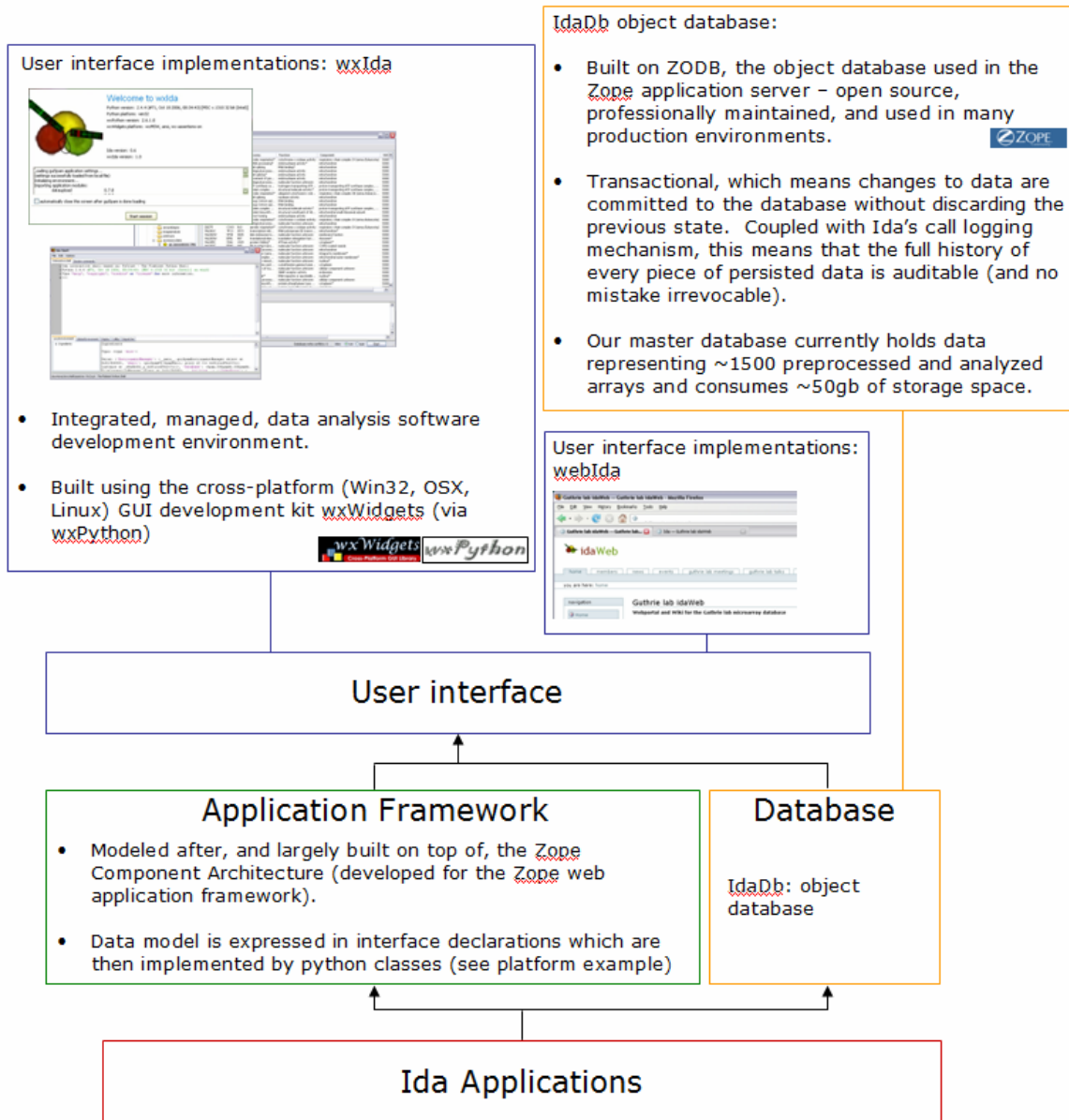


Figure 47. Overview of the `Ida` splicing microarray data analysis software suite.

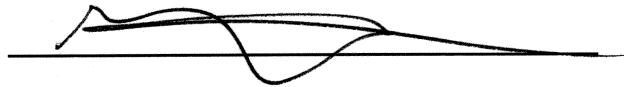
Library Release

Publishing Agreement

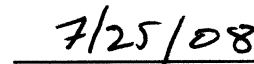
It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

A handwritten signature in black ink, consisting of several loops and a long horizontal stroke, positioned above a solid horizontal line.

Author Signature

The handwritten date "7/25/08" in black ink, positioned above a solid horizontal line.

Date