UCSF
UC San Francisco Electronic Theses and Dissertations

Title
Introns Influence Chromatin Structure and Gene Expression

Permalink
https://escholarship.org/uc/item/1pr904dk

Author
Plocik, Alex Matthew

Publication Date
2011

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Introns Influence Gene Expression and Chromatin Structure

by

Alex M. Plocik

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

[redacted]

Acknowledgements

First and foremost, I want to thank my wife Runa who has done more to help me attain my dreams than I could ever ask of her. We have accomplished a lot together. And we now have Ashar, a wonderful boy who knows how to laugh more than anything else. I hope that I can encourage Ashar to pursue his happiness the same way my parents did. My parents, Dan and Janet Plocik, saw that I loved science and encouraged me to pursue that path simply because it made me happy.

I owe much of my success in graduate school to the amazing members of the Guthrie lab. I have been very fortunate to work with such generous and exceptional people. Karen Kim-Guisbert introduced me to the lab and I am very grateful to her for this first impression. I greatly admired her strong and independant thinking and I hoped that the lab had something to do with that.

Upon joining the lab I quickly found common interests with Jeff Pleiss, Gregg Whitworth, and Megan Bergkessel. They encouraged me and they challenged me. Together they built the impressive microarray platform that I turned to again and again during my research. My work is greatly indebted to theirs.

Several post-docs have guided me on my journey through graduate school: Quinn Mitrovich, Tracy Kress, Corina Maeder, Mette Lund and Anne Anne deBruynKops were always quick to offer their attention by asking questions and providing answers. Since then, Kristen Patrick, Sarah Ledoux, Jaci Greimann, and Mike Marvin have kept this essential spirit of the Guthrie alive and I am very grateful for their help.

I've also been very lucky to have excellent peers: Argenta Price, Erica Moehle, and Kelly Nissen are destined to do great things. The students at UCSF are the most engaged group of people I have ever met. Honestly, I don't think I will ever have a more fun and exciting collaboration as I did with my classmate Jason Huff. Jason encouraged question everything and to pursue anything that seemed exciting.

I have a tremendous debt of gratitude to the faculty at UCSF. Barbara Panning, Hiten Madhani, and Keith Yamamoto generously offered sage advice as I ventured into the topic of chromatin structure, which was a completely new and unexpected direction for me. Sandy Johnson, a master of both biochemical experimentation and evolutionary thought, has brought tremendous insight to our conversations, for which I am extremely grateful.

John Abelson has been a great inspiration to me. I know no one who is happier with their career than John. He has incredible enthusiasm and an instinct about exciting science that is unmatched. When John exclaims "Now, that's a result!" it's time to start writing the manuscript.

And finally, I want to thank Christine Guthrie for her wonderful support and wisdom over the years. She has given me incredible and unpararelled freedom to pursue the questions that interested me. She has encouraged me to be the kind of scientist that I want to be, while simultaneously tempering my instincts with her wisdom.

Published Work:

Chapter 2 was published as referenced below in collaboration with Jason Huff and Keith Yamamoto.

Huff, J.T., Plocik, A.M., Guthrie, C. & Yamamoto, K.R. Reciprocal intronic and exonic histone modification regions in humans. *Nat. Struct. Mol. Biol* **17**, 1495-1499 (2010).

Abstract

Eukaryotic genes are littered with intervening sequences, or introns, that are transcribed, but must be precisely excised from a messenger RNA before it can be properly translated into protein. While introns were once regarded as "junk DNA," they are not inconsequential. However, we currently lack the knowledge to accurately predict the functions, if any, of individual introns in any organism. Here, I describe efforts to better understand the evolution and function of introns, with the vision that we will soon be able to identify important introns and predict their functions.

In Chapter 2, my colleagues and I describe an unexpected consequence of intron presence on chromatin structure, which suggests that introns have a broader influence on the biology of eukaryotes than previously appreciated. By analyzing published surveys of nucleosomes and 41 chromatin marks in humans, we show that 5' intronic and 3' exonic regions of active genes are differentially marked by characteristic chromatin marks, thus contributing substantially to the patterns of histone modifications within active genes. Intriguingly, these modification patterns were stable despite dramatic changes in the frequency of splice site usage at two alternative spliced genes. Thus, similar to promoter marks, which are relatively stable to differences in productive transcription, we propose that intronic and exonic chromatin marks reflect exon definition, rather than splicing *per se*.

In Chapter 3, I describe my work to better understand why certain introns persist in eukaryotic genomes. Using comparative genomics, I show that the ribosomal protein

genes of *Saccharomyces cerevisiae* have greatly resisted intron loss. Mimicking the effect of intron loss with directed mutagenesis, I perform experimental tests that demonstrate that these introns do not promote gene expression, but rather, provide a means for regulation. Specifically, I show that the genes encoding ribosomal protein S9, both in *S. cerevisiae* and *Drosophilia melanogaster*, autoregulate in an intron-dependent manner. Lastly, I summarize published gene expression data from diverse animals, which suggest that multiple forms of alternative splicing have evolved to autoregulate S9 gene expression. Thus, I propose that the introns of eukaryotic genes persist, in part, due to their propensity to evolve regulatory function.

Table of Contents

Chapter 1: Introduction 1

1.1 A genome for every question 1

1.2 Gene architecture and chromatin structure 1

1.3 Introns and the regulation of gene expression 3

Chapter 2: Reciprocal intronic and exonic histone modification regions in humans 8

2.1 Abstract 8

2.2 Introduction 8

2.3 Materials and Methods 10

2.3.1 Cell culture and regulated alternative splicing 10

2.3.2 Epigenomic analyses 11

2.3.3 Cross-linked mononucleosome ChIP 11

2.3.4 Quantitative PCR 12

2.4 Results 13

2.4.1 Histone modifications in 5' introns 13

2.4.2 Histone modification profiles separable from bulk nucleosome occupancy 16

2.4.3 Intronic and exonic marks associated with bone fide exons 18

2.4.4 Intronic and exonic marks are stable despite changes in splicing 19

2.5 Discussion 22

2.6 Author Contributions 27

2.7 Acknowledgements 27

2.8 Figures and Tables 28

Chapter 3: Diverse forms of RPS9 splicing are part of an evolving autoregulatory circuit 60

3.1 Abstract 60

3.2 Introduction 62

3.3 Results 64

3.3.1 Yeast ribosomal protein genes have resisted recent intron loss 64

3.3.2 Introns repress ribosomal protein gene expression..... 67

3.3.3 Drosophila RpS9 autoregulates through alternative splicing and NMD..... 69

3.3.4 Diverse forms of RPS9 alternative splicing are associated with structured and conserved RNA sequences 70

3.4 Discussion 73

3.4.1 Biased intron loss may reflect selection for functional introns..... 74

3.4.2 Intron function in the absence of alternative spicing 76

3.4.3 Evolution of the RPS9 autoregulatory circuit 78

3.5 Methods..... 80

3.5.1 Intron gain and loss analysis 81

3.5.2 S. cerevisiae strain construction..... 82

3.5.3 D. melanogaster S2 cell RNAi and transfection..... 82

3.5.4 RNA isolation and cDNA synthesis 83

3.5.5 Microarray analysis 84

3.5.6 Quantitative PCR 84

3.5.7 Assessment of alternative splicing among animal RPS9 orthologs 85

3.6 Acknowledgements..... 85

3.7 Figures and Tables..... 85

Chapter 4: Epilogue 113

4.1 Chromatin structure reflects gene architecture 113

4.1.1 Splicing is coupled to transcription 113

4.1.2 Is transcription coupled to splicing? 114

4.1.3 Dissecting the mechanisms underlying intronic and exonic chromatin marking .. 114

4.1.4 Unanswered questions 117

4.2 Regulated unproductive splicing: a recurring theme among eukaryotic RNA binding proteins 119

4.2.1 Regulated unproductive splicing is common in S. cerevisiae 119

4.2.2 Does S9 directly regulate alternative splicing? 121

Appendix A: Towards a comprehensive understanding of S. cerevisiae intron function 128

A.1 Progress toward a complete intron deletion collection 128

A.2 The effect of RPG intron deletion on gene expression 128

A.3 The effect of intron deletion on fitness 132

4.3 Figures and Tables 136

List of Tables

Table 1.1 Chromatin marks at active S. cerevisiae genes.....7

Table 2.1 Primers used in this study.....59

Table 3.1 Intron annotations and predictions.....104

Table 3.2 Primers used in this study.....112

Table 4.1 S. cerevisiae Δi strains.....144

List of Figures

Figure 2.1 PCA analysis of ChIP-seq reveals three major components.....29

Figure 2.2 Exonic and intronic marks in the human epigenome 30

Figure 2.3 Exonic and intronic marks do not merely reflect nucleosome occupancy ...32

Figure 2.4 Histone modifications are stable despite transient changes in alternative splicing 33

Figure 2.5 Histone modifications are stable despite stable changes in alternative splicing 35

Figure 2.6 Chromatin structure reflects gene architecture 37

Figure 2.7 PCA of ChIP-seq data 38

Figure 2.8 Chromatin structure reflects gene architecture (all marks) 42

Figure 2.9 ChIP-qPCR validation of ChIP-seq profiles at the STIM1 gene 47

Figure 2.10 PCA of ChIP-seq data is modestly affected by nucleosome density 48

Figure 2.11 Intronic and exonic marks do not merely reflect nucleosome density.....50

Figure 2.12 Histone modifications are more strongly influenced by bone-fide exons than exon-like composition regions (ECRs) 52

Figure 2.13 YPEL5 RT-PCR and ChIP-qPCR profiles in response to caffeine induction.54

Figure 2.14 CD45 ChIP-qPCR profiles in response to changes in hnRNPLL expression 56

Figure 2.15 Representative gene models illustrating that H3K36me3 marks are enriched near individual human exons at actively transcribed genes.....58

Figure 3.1 Biased intron loss in hemiascomycetous yeasts after the recent whole-genome duplication event.....86

Figure 3.2 RPG intron deletions cause gene-specific increases in steady-state expression 88

Figure 3.3 D. melanogaster RPS9 is autoregulated by alternative splicing coupled to NMD 90

Figure 3.4 Diverse alternatively spliced RPS9 isoforms encode PTC+ exons associated with high nucleotide conservation and predicted RNA structures.....92

Figure 3.5 Evolutionary models 94

Figure 3.6 Heatmap of annotated and predicted introns in hemiascomycetous yeast orthologs of S. cerevisiae duplicated gene pairs 95

Figure 3.7 Illustration of the intron deletion method 98

Figure 3.8 Conserved intronic regions among yeast RPS9 orthologs 99

Figure 3.9 Detailed splicing-sensitive microarray data of intron deletion mutants 102

Figure 3.10 Comparison of RPS9 spliced isoforms from 10 animals.....103

Figure 4.1 Model: Spliceosome recruitment specifies chromatin marks 119

Figure A.1 Microarray data does not support mRNA decreases reported for Δi alleles 137

Figure A.2 Total RPS9A/B mRNA expression measurements are strikingly different between two different studies 138

Figure A.3 RPGs most strongly affected by intron deletion are among the least highly expressed 139

Figure A.4 Evidence for unspliced RPL22B mRNA associated with conserved intronic regions 140

Figure A.5 Evidence for unspliced RPL18B mRNA associated with conserved intronic regions 141

Figure A.6 Fitness correlates with total RPG mRNA levels predicted by gene deletion 142

Figure A.7 Fitness does not correlate with total RPG mRNA levels caused by intron deletion 143

Chapter 1: Introduction

1.1 A genome for every question

The presence of spliceosomal introns within eukaryotic genomes profoundly influences nucleotide metabolism of both DNA and RNA. Introns increase the amount of non-coding DNA in a genome, which is only removed after conversion to RNA. In this Dissertation, I explore the impact of introns at both the level of the chromatin structure (DNA and protein) and the transcriptome (RNA). To do so, I studied two very different organisms, with very different genomes: the key difference being the number and size of introns within each genome. Like many eukaryotes, the human genome is largely composed of intronic DNA. Given the vast amount of data generated to define the human "epigenome," human genes present an unparalleled opportunity to study the impact that introns have on gene architecture and chromatin structure. On the other hand, the intron-poor genome of *S. cerevisiae* may represent a genome that has purged all but the most important introns. With its powerful genetic tools, *S. cerevisiae* is an ideal model system to study the consequences of intron loss on gene expression. Below, I briefly review the state of knowledge in each of these fields and systems prior to my contribution.

1.2 Gene architecture and chromatin structure

Eukaryotic transcription occurs in the context of chromatin. To access its substrate, RNA polymerase must contend with the nucleosome: DNA wrapped around a histone octomer. The ability of RNA polymerase to access and traverse nucleosomal DNA is

modulated by myriad post-translational modifications of the histones (chromatin marks) and through physical interactions with various transcription factors.

The technique of chromatin immunoprecipitation (ChIP) has revealed remarkably specific associations between particular chromatin marks and DNA elements associated with transcription: most notably, promoters. Two major patterns of chromatin marks have been identified: 5' and 3' marks that appear to be associated with gene architecture, specifically, regions of transcription initiation and elongation (respectively). Where functions have been determined, 5' marks and 3' marks influence transcription differently: H3K4me3 at promoters promotes initiation, while H3K36me3 in the gene body represses initiation. These basic patterns, which have been comprehensively studied in *S. cerevisiae*, are localized similarly in human, particularly at promoters (Figure 1.1).

Introns, however, profoundly influence the gene architecture, increasing the length and non-coding content of a gene. While *S. cerevisiae* genes are stereotyped, human genes are diverse in length and non-coding content. As a consequence, the elongation phase of RNA polymerase is much longer due to introns. Further, RNA polymerase traverses large regions of non-coding sequence that, theoretically, requires less fidelity than coding and regulatory sequences. Thus, 3' marks associated with transcriptional elongation may not be expected to localize identically to *S. cerevisiae* genes that are small and largely composed of coding sequence. These marks may also perform new functions in order to meet the challenges of transcribing long intron-containing genes.

We sought to assess the chromatin structure of active genes with diverse and complex gene architectures such as those found in the human genome.

To begin, Jason Huff and I turned to the ChIP assay to identify regions of the human genome that are associated with specific post-transcriptional histone modifications or histone variants. Fortuitously, DNA associated with 41 different chromatin marks had been sequenced after ChIP (ChIP-seq), which we could then use for high resolution mapping (smaller than the amount of DNA in a nucleosome) to almost any gene in the human genome. While the power of this technique was truly awesome, the amount of data was staggering (the nucleosome control alone consisted of over 150 million unique sequence reads). Our challenge was in the analysis, and perhaps more so, transforming this unwieldy data into accessible knowledge. Chapter 2 describes our work, in which we demonstrate that chromatin marks at active genes can be divided into 5' intronic and 3' exonic regions that are demarcated by the first 3' splice site. This finding uncovered a new and unexpected layer of complexity and further strengthened the perspective that chromatin structure, transcription, and splicing are intimately entwined.

1.3 Introns and the regulation of gene expression

Removal of introns requires a complex machine --- the spliceosome --- before protein-coding transcripts can be productively translated. Introns, thus, present an inherent barrier to gene expression, but the process of RNA splicing can actually promote gene expression (Le Hir, Nott, & Moore, 2003). For example, U1 snRNA interaction with the 5'

splice site of the promoter proximal intron of the human β-globin gene recruits transcription initiation factors including TFIID, TFIIB, and TFIIH (Damgaard et al., 2008).

Perhaps the best known consequence of splicing is the potential to generate multiple mRNA isoforms from a single gene by alternative splicing. In the most commonly studied cases of alternative splicing, mRNA isoforms are productively spliced and translated to generate different functional protein isoforms (Nilsen & Graveley, 2010). In others cases, alternative splicing results in regulated unproductive splicing and translation (RUST) in which mRNA isoforms that are targeted for degradation by quality control pathways, typically nonsense mediated decay (Lareau, Brooks, Soergel, Meng, & Brenner, 2007).

Due the small number of introns in the *S. cerevisiae* genome, only ~300 genes contain introns and most of these have only one. Consequently, alternative splicing in *S. cerevisiae* is notoriously constrained to a form of intron retention that, in most cases, would be expected to generate an mRNA isoform destined for degradation. Please note that I use the phrase alternative splicing where others may not: I count intron retention (an *unspliced* intron) as a form of alternative *splicing*, which may or may not also be unproductive. I make use this misnomer purposefully because I strongly believe that the use of the phrase "regulated splicing" *in leiu* of the more common phrase "alternative splicing" inhibits comparisons between *S. cerevisiae* and other eukaryotes.

The paucity of introns in *S. cerevisiae* is atypical. The earliest eukaryotes may have had as many introns as humans, while the lineage leading to *S. cerevisiae* has lost the vast

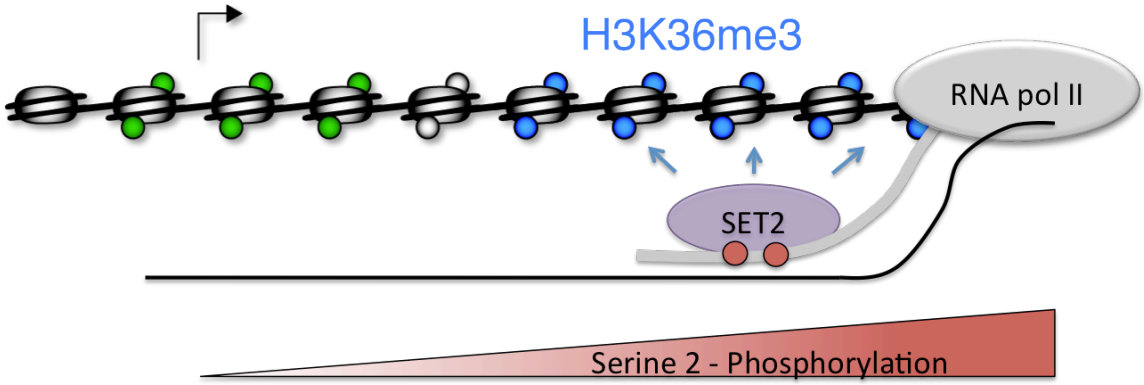
majority of these introns (Roy & Gilbert, 2005; Stajich, Dietrich, & Roy, 2007). Consequently, it has been stated that *S. cerevisiae* introns must be unimportant, and therefore, "on the way out" (Parenteau et al., 2008). Notably, however, *S. cerevisiae* introns are biased for particular gene classes, suggesting that these remaining introns may have been preserved by selection. Supporting this perspective, fully one-third of *S. cerevisiae* introns reside within ribosomal protein genes (RPGs), and upon amino acid starvation, their splicing is inhibited (Pleiss, Whitworth, Bergkessel, & Guthrie, 2007a). Intrigued by the idea that some *S. cerevisiae* introns may have resisted intron loss, I sought to systematically test the consequence of intron loss for every intron in the yeast genome. However, to embark on this ambitious task, two major innovations were needed. First, mutations intended to mimic intron loss would need to be exact. Second, I would need a platform to assess the splicing and expression of every intron-containing gene.

To accomplish the first feat, I naturally took advantage of "the awesome power of yeast genetics." *S. cerevisiae* is indeed a tremendous model system for mutational analysis; however, high-throughput gene replacement methods generally introduce extraneous mutations. Precise intron loss would require a newer more sophisticated method. Fortuitously, a chance meeting with Jon Staley provided me with the method needed. With his generous gift of reagents, I was able to create a *S. cerevisiae* strain engineered as part of a high-throughput pipeline to precisely remove introns without introducing any additional mutations.

The second feat required a facile and reliable platform to measure the expression and splicing of every intron-containing gene in *S. cerevisiae*. Here, Jeff Pleiss and Gregg Whitworth, made a tremendous contribution by building a custom made microarray platform that could be used to screen a large number of mutants quickly and cheaply (Pleiss, Whitworth, Bergkessel, & Guthrie, 2007b). In this way, I was able to analyze each mutant identically, which greatly reduced ad-hoc experimental design. While every approach has caveats, it is essential that critical issues be known. Importantly, these "splicing microarrays" had been improved through extensive use and refinement.

The first insights to come from this strategy are described in Chapter 3. By studying rates of intron loss, I confirmed that the introns of RPGs have resisted intron loss compared to other genes. Intriguingly, I found no evidence that these introns enhance their expression. Instead, I found new instances of regulated unproductive splicing. Surprisingly, both copies of RPS9 require introns for autoregulation. While such cross-regulation between RPG copies had not been reported before, the essential result fits nicely within a paradigm of autoregulation set by *RPL30* and *RPS14B* (Dabeva & Warner, 1993; Fewell & Woolford, 1999). Presumably, RPS9 genes have resisted intron loss because splicing provides a means for autoregulation, which may be particularly` important for these genes. In support of this perspective, I found that alternative splicing is common among animal RPS9 genes. Intriguingly, these genes also appear to be regulated by RUST. Thus, RPS9 genes have maintained or reinvented an autoregulatory pathway over long periods of evolutionary time.

Table 1.1 Chromatin marks at active S. cerevisiae genes



Chapter 2: Reciprocal intronic and exonic histone modification regions in humans

2.1 Abstract

While much attention has been focused on chromatin at promoters and exons, human genes are mostly composed of intronic sequences. Analyzing published surveys of nucleosomes and 41 chromatin marks in humans, we identified histone modifications specifically associated with 5' intronic sequences, distinguishable from promoter marks and bulk nucleosomes. These intronic marks were spatially reciprocal to H3K36me3, typically transitioning near internal exons. Several marks transitioned near bone fide exons, but not near nucleosomes at exon-like sequences. Thus, we interrogated splicing for a role in histone marking. Despite dramatic changes in regulated alternative splicing, histone marks were stable. Notably, these findings are consistent with a role for exon definition in specifying histone marks. In summary, our results demonstrate that the location of intragenic marks in humans can be distilled into a simple organizing principle: association with 5' intronic or 3' exonic regions.

2.2 Introduction

Nucleosomes, the basic units of eukaryotic chromatin structure, are marked by numerous post-translational modifications and compositional variants that influence DNA and RNA metabolism (Berger, 2007; Kouzarides, 2007; Loyola & Almouzni, 2007). For example, transcription by RNA polymerase II is associated with particular modifications and variants at promoters, including histone H3 lysine 4 methylations,

numerous acetylations, H2A.Z, and H3.3 (B. M. Lee & Mahadevan, 2009; Deal & Henikoff, 2011). Of these, H3K4me3 acts as a binding determinant for the TAF1 subunit of TFIID, mechanistically linking this modification to transcription initiation (Vermeulen et al., 2007).

By analogy to chromatin marks at promoters, marks positioned within genes physically coincide with various co-transcriptional processes and may influence or be influenced by RNA processing(Perales & Bentley, 2009; Zhong, Wang, Han, Rosenfeld, & Fu, 2009). For example, depletion of human BRE1A/RNF20, which is required to catalyze H2B lysine 120 ubiquitylation (H2Bub), results in decreased chromatin association of the transcription elongation and splicing factor SKIP/SNW1 (Brès, Yoshida, Pickle, & Jones, 2009). Similarly, H3K36me3 is catalyzed by a methyltransferase associated with elongating, hyperphosphorylated RNA polymerase II (B. Li, Carey, & Workman, 2007) and depletion of the human enzyme, SETD2/HYPB/KMT3A, leads to a defect in mRNA export and changes in alternative splicing (Yoh, Lucas, & Jones, 2008; Luco et al., 2010).

Most provocatively, bioinformatic analyses of data from diverse animals indicate that certain histone modifications, particularly H3K36me3, are not only elevated within genes, but specifically at exons (Andersson, Enroth, Rada-Iglesias, Wadelius, & Komorowski, 2009; Kolasinska-Zwierz et al., 2009; Nahkuri, Taft, & Mattick, 2009; Schwartz, Meshorer, & Ast, 2009; Spies, Nielsen, Padgett, & Burge, 2009; Tilgner et al., 2009). It has been suggested, though, that H3K36me3 levels simply reflect higher nucleosome occupancy at exons in Metazoa (Schwartz et al., 2009; Tilgner et al., 2009).

In any case, how the bulk of chromatin within genes compares to the one or few nucleosomes at each exon has not been fully elucidated.

To explore more broadly the relationships between chromatin and gene architecture, we began by analyzing published human epigenomic data. We detected regions located after active transcription start sites and extending through 5' introns of variable length and the first few internal exons, enriched in H2Bub, H3K23ac, and eight different methylations including H3K79me2/3. Regions in the 3' portions of active genes displayed H3K36me3 peaking at internal exons with elevated levels extending into each downstream intron. Regulated alternative splicing failed to change these histone modification profiles. Implications of these findings are discussed.

2.3 Materials and Methods

2.3.1 Cell culture and regulated alternative splicing

SW620 human colon cancer cells (ATCC CCL-227) were grown in DMEM supplemented with 10% FBS, Pen-Strep, and Normocin (Amaxa). They were grown to near confluence for the experiments in Figure 2.9 and were seeded at a density of 1 x 105 cells per cm2 24 hours prior to treatment with media supplemented with the indicated concentration of caffeine for 7 hours in Figure 2.3. BJAB and BL41 B cell lines (Oberdoerffer et al., 2008) were grown in RPMI-1640 supplemented with 10% FBS, glutamine, Pen-Strep, and Normocin (and 5 ug per ml puromycin for the BJAB lines) to a density of about 1 x 106 cells per ml. The earliest passages available for the B cell experiments were used, because the effect on CD45 alternative splicing decreased over continual passaging.

2.3.2 Epigenomic analyses

Analyses were performed and plots were generated with custom R scripts (http://www.r-project.org). Human gene models were obtained from the UCSC genome browser RefSeq (hg18) annotations (http://genome.ucsc.edu/). We offset sequencing read end-positions by 73 bases (half the width of the canonical nucleosome) to approximate the position of nucleosomal centers. MNase- and ChIP-seq reads were aggregated in non-overlapping 1-kb windows from 10 kb upstream to 10 kb downstream of TSSs and data across these genomic segments were subjected to PCA. PCA was performed by singular value decomposition with the data centered and scaled using the R function prcomp(). Figure 2.1 was rendered in PyMOL (http://www.pymol.org/). For all plots only one gene model was shown for genes with multiple models. For Figure 2.2A, 2.12, and 2.13, a Gaussian kernel density estimator was used with both the kernel width and bin size set to 146 bp (size of one canonical nucleosome). The kernel density scale was adjusted so that uniform density would be equal to 1 at all positions. We did not precisely map the positions of individual nucleosomes, because our analysis was generally on the scale of tens to hundreds of nucleosomes. However, we verified that the plots were robust to changes in nucleosome phasing and average width in the kernel density estimator.

2.3.3 Cross-linked mononucleosome ChIP

ChIP was performed as described (Edmunds, Mahadevan, & Clayton, 2008) with the following modifications. Nuclei from 1 x 107 crosslinked cells were digested with 100 units of MNase (Worthington) in 500 µl of Buffer N supplemented with 3 mM CaCl2 for

10 min at 37°C. We previously determined the concentration of MNase to be optimal for digestion to mostly mononucleosomes by agarose gel electrophoresis of the cross-link reversed input chromatin. Chromatin extract equivalent to 2 x 106 cells in 500 µl of RIPA was incubated with 4 µg of one of the following antibodies for 2 hours at 4°C: H3K79me2, H3K36me3, or H3 (Abcam ab3594, ab9050, or ab1791). Instead of Protein A/G agarose beads, 50 µl of Protein G Dynabeads (Invitrogen) were added to each tube and incubated overnight at 4°C. After the chromatin elution step, DNA was purified with QIAquick PCR columns (QIAGEN) and quantified by qPCR as described below.

To analyze ChIP-seq data, we used MNase-digested input material as a quantitative standard, thereby measuring the efficiency of immunoprecipitation, and internally controlling for digestion biases. In addition, we normalized H3K79me2 and H3K36me3 ChIP to an H3 ChIP to control for differences in nucleosome occupancy (Figure 2.9). It should be noted that ChIP-qPCR is subject to its own particular caveats, most notably the dependence on fixed primer positions. As such, MNase digestion of DNA around strongly positioned nucleosomes can negatively affect the amplification efficiency of partially overlapping amplicons. Similarly, histone modifications or variants may differentially influence the extent of MNase digestion and, consequently, bias amplification efficiency. We attempted to minimize these effects by tiling genes with primer sets designed to amplify small segments of DNA (62 bp median amplicon size, Table 2.1).

2.3.4 Quantitative PCR

PCR primers (Table 2.1) were designed with BatchPrimer3 v1.0 (http://wheat.pw.usda.gov/demos/BatchPrimer3/) from human repeat-masked sequence obtained from the UCSC genome browser (hg18). Serial dilutions of DNA ranging from 50% to 0.4% of the input chromatin were used to obtain calibration curves, measure primer efficiencies, and ensure that quantification was in a linear dynamic range. Primer sets yielding multiple amplification products or calibration curves with R-squared values of < 0.96 were excluded. For each qPCR sample, 0.5 µl of 50 µl eluted ChIP DNA was amplified in 25 µl volume reactions containing 250 µM dNTPs, 1X (NH4)2SO4 buffer (Fermentas), 0.5 µM primer, 1.5 mM MgCl2, 1.25 Units Dynazyme II (Finnzymes), and Sybr Green I fluorescent dye (Sigma). Fluorescence was measured on a BioRad CFX-96 machine using standard cycling conditions (3 min at 95°C, 40 cycles of 15 s at 95°C, 30 s at 55°C, and 15 s at 72°C). Each ChIP value was calculated as the median of 2-4 technical qPCR replicates. For each biological replicate consisting of qPCR replicates, median ChIP values for H3K79me2 and H3K36me3 were then divided by median H3 ChIP values.

2.4 Results

2.4.1 Histone modifications in 5' introns

To discover new epigenomic features related to gene expression, we first analyzed published surveys of human mononucleosomes prepared by native micrococcal nuclease digestion (MNase-seq), mononucleosome ChIP (ChIP-seq) for 20 histone methylations, 18 histone acetylations, and the histone variants H2A.Z and H3.3, as well

as cross-linked sonicated ChIP-seq of H2B lysine 120 ubiquitylation (Barski et al., 2007),(Schones et al., 2008),(Shema et al., 2008),(Zhibin Wang, Zang, et al., 2008),(Jin et al., 2009). We used principal component analysis (PCA) to reduce the complexity of this large data set, producing a smaller number of dimensions, or principal components, that retain much of the variance in the original data set. Thus, we aimed to identify subsets of the 41 chromatin marks that shared common characteristics.

To identify relationships specific to genes, we applied PCA to data from 10 kb upstream to 10 kb downstream of annotated RefSeq transcription start sites (TSSs; Figure 2.1 and Figure 2.7). We confirmed that the first principal component corresponded to those marks enriched at active promoters, including all 17 marks (H2A.Z, H2BK5ac, H2BK12ac, H2BK20ac, H2BK120ac, H3K4ac, H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me1, H3K18ac, H3K27ac, H3K36ac, H4K5ac, H4K8ac, and H4K91ac) previously identified and characterized as a "backbone"(Zhibin Wang, Zang, et al., 2008). In addition, our analysis indicated that four more acetylations (H2AK9ac, H3K23ac, H4K12ac, and H4K16ac) appeared also to be enriched at active promoters, raising the number of marks in the "backbone" to at least 21 (Figure 2.1 and 2.8). The second principal component included nucleosomes and numerous marks that are generally depleted near active promoters, but have elevated occupancy within genes (Figure 2.7 and Figure 2.8). The third principal component included high levels of H2BK5me1, H2Bub, H3K4me1, H3K4me2, H3K9me1, H3K23ac, H3K79me1, H3K79me2, H3K79me3, and H4K20me1 (Figure 2.1, 2.7, and 2.8).

Intriguingly, we noticed that several marks, particularly those enriched in the third principal component, were enriched in the 5' regions of actives genes. This was reminiscent of their spatial distribution in yeast (B. Li et al., 2007), but in humans these marks extend downstream into 5' introns of varying length, decreasing in the vicinity of exons. To assess if this is a genome-wide phenomenon, we aligned thousands of genes by TSSs and sorted them by distances from the TSSs to the first 3' splice sites, plotting each gene as a row with sequencing read density as a heatmap (Figure 2.2). Thus, this perspective simultaneously displayed three different aspects of gene architecture: TSS positions, first introns, and first internal exons. In this presentation of the three principal components on the genome (Figure 2.7), the components exhibited distinct profiles, with enrichment more apparent at genes with higher expression. Component three, and to a lesser extent component two, displayed a 5' bias that decreased near internal exons (Figure 2.7).

Analyzing individual chromatin marks representative of each principal component also revealed relationships between features of gene architecture and particular chromatin marks (Figure 2.8). As expected, promoter marks, such as H3K4me3, which were high in component one, surrounded TSSs (Figure 2.2). Nucleosomes, which were high in component two and must be present at TSSs, introns, and exons containing modified nucleosomes, appeared slightly higher near TSSs and internal exons than within neighboring introns (Figure 2.2 and refs.; Andersson et al., 2009; Nahkuri et al., 2009; Schwartz et al., 2009; Spies et al., 2009; Tilgner et al., 2009). H3K79me2, which is high in component three, showed enrichment that extended downstream of TSSs into first

introns until reaching internal exons (Figure 2.2). In contrast, H3K36me3, the most 3'-biased mark, exhibited a spatially reciprocal relationship with H3K79me2, clearly peaking at internal exons (Figure 2.2). For comparison, H3K27me1 is present within active genes (Barski et al., 2007) and is high in the second component, but exhibited little if any change with respect to exons (Figure 2.8).

2.4.2 Histone modification profiles separable from bulk nucleosome occupancy

Interpreting the distribution of histone modifications from ChIP-seq data is complicated by potential technical artifacts and nucleosome occupancy biases. Specifically, previous work has documented the sequence bias of micrococcal nuclease (MNase; McGhee & Felsenfeld, 1983), the GC bias typically present in sequencing data produced by Illumina 1G Genome Analyzers (Dohm, Lottaz, Borodina, & Himmelbauer, 2008), the higher nucleosome occupancy within exons compared to neighboring introns (Andersson et al., 2009; Nahkuri et al., 2009; Schwartz et al., 2009; Spies et al., 2009; Tilgner et al., 2009), and the differential enrichment of certain nucleosomes dependent upon chromatin preparation conditions (Jin & Felsenfeld, 2007). Therefore, we sought quantitatively controlled validation of histone modification profiles at specific genes.

To determine H3K79me2 and H3K36me3 profiles normalized for nucleosome density, we performed ChIPs of mononucleosomes analyzed by quantitative PCR (ChIP-qPCR). Using MNase-digested input material as a quantitative standard, we measured the efficiency of immunoprecipitation, thereby internally controlling for digestion biases. In addition, we normalized H3K79me2 and H3K36me3 ChIP to an H3 ChIP to control for

differences in nucleosome occupancy. It should be noted that ChIP-qPCR is subject to its own particular caveats, most notably the dependence on fixed primer positions. As such, MNase digestion of DNA around strongly positioned nucleosomes can negatively affect the amplification efficiency of partially overlapping primer sets. Similarly, histone modifications or variants may differentially influence the extent of MNase digestion and, consequently, bias amplification efficiency. We attempted to minimize these effects by tiling genes with multiple primer sets designed to amplify small segments of DNA (62 bp median, Table 2.1).

To independently assess intronic and exonic marks by ChIP-qPCR, we tiled the STIM1 gene, as it has internal exons separated by large 5' introns. We used 48 unique primer sets, focusing on the first four introns and internal exons. In agreement with ChIP-seq profiles, we observed enrichment for H3K79me2 in the first three long introns and H3K36me3 near exons (Figure 2.9). Our H3 ChIP and published MNase-seq displayed some heterogeneity along *STIM1* (Figure 2.9), which may reflect differential nucleosome occupancy or positioning, especially at exons (Andersson et al., 2009; Nahkuri et al., 2009; Schwartz et al., 2009; Spies et al., 2009; Tilgner et al., 2009). However, H3K36me3 within and near exons at *STIM1* remained enriched after normalization to H3, although reduced in dynamic range (Figure 2.9).

Sequencing is subject to the biases mentioned above, but compared to qPCR it is less sensitive to quantification differences due to unequal MNase digestion and is not reliant on fixed primer positions. ChIP-seq of histone marks showed higher occupancy centered

on exons, as also observed by MNase-seq (Figure 2.3A). Normalizing ChIP-seq data by MNase-seq data largely removed the exon-centered peak for most marks (e.g. H3K79me2), but H3K36me3 remained enriched after normalization, peaking near exons and extending on average a few kilobases downstream (Figure 2.3B). Also, performing the original PCA analysis above by first normalizing every ChIP-seq dataset by MNase-seq supported similar conclusions as PCA without normalization (Figure 2.10). Furthermore, normalizing the heatmaps in Figure 2.2 to nucleosomal data introduced some noise into the analysis (Figure 2.11), but did not alter the reciprocal pattern around internal exons. In summary, genome-wide analysis of sequencing data supported the conclusion from ChIP-qPCR (Figure 2.9) that the enrichment of H3K36me3 above nucleosomes near individual exons was a general phenomenon and robust to changes in quantification technique and normalization method. Furthermore, most histone modification profiles were not merely reflections of bulk nucleosome occupancy.

2.4.3 Intronic and exonic marks associated with bone fide exons

Internal exons can be viewed as genome annotations of spliced RNAs, minimally including 3' and 5' splice sites at the beginning and end, respectively, with strong additional evolutionary sequence constraints to encode proteins and include splicing enhancer and silencer sites. Spies et al. identified intronic exon-like sequence composition regions (ECRs) with high nucleosome occupancy, thus suggesting through computational analysis that the sequence composition of exons relative to surrounding introns is likely sufficient to generate high nucleosome occupancy (Spies et al., 2009).

Thus, ECRs are regions that contain exon-like nucleosome occupancy, but do not necessarily contain other aspects of bone fide exons.

Having demonstrated above that histone modification profiles within introns and exons were not merely reflections of bulk nucleosome occupancy, we sought to determine if the high nucleosome occupancy at exons could explain changes in histone modification patterns that occur near exons. We compared ECRs to bona fide exons to determine if the sequence composition of exons is sufficient to pattern histone modifications, as it appears to be to pattern nucleosome occupancy (Spies et al., 2009; Figure 2.10). To control for biases in number and position of ECRs relative to all exons, we compared each intronic ECR (12,687 total) to its nearest *bona fide* exon. Similar to genome-wide averages of H3K36me3 at exons (Figure 2.3), H3K36me3 peaked at ECR-matched exons. However, H3K36me3 did not peak at ECR positions (Figure 2.10). On average, marks in the third principal component were higher upstream of exons than downstream and a similar pattern was found near ECRs, although the marks variably displayed decreased spatial response to ECRs compared to exons (e.g. H3K4me1, H3K4me2, H3K9me1, and H3K23ac). For comparison, H3K27me1 showed little difference up- and downstream of exons or ECRs (Figure 2.10). Therefore, the sequence composition of exons and ECRs appears partially, if not fully, sufficient to generate a spatial decrease in histone modification going from up- to downstream for marks in the third principal component, but is not sufficient to direct an increase in H3K36me3.

2.4.4 Intronic and exonic marks are stable despite changes in splicing

Finding that the sequence composition of ECRs could not fully account for the spatial dynamics of histone modifications that change near exons, we considered a model in which RNA splicing events are mechanistically linked to co-transcriptional histone modification, thus coupling exon positions to chromatin states. Splicing is a complex, multistep process that involves recognition of splice sites and progression of spliceosome assembly, ultimately resulting in intron excision (Wahl, Will, & Lührmann, 2009). Both splice site recognition and spliceosome assembly can be regulated to produce alternative RNA splicing outcomes (Chen & Manley, 2009).

To test if RNA splicing itself could direct histone modification profiles, we examined two cases of regulated alternative splicing in which the frequency of exon inclusion was experimentally manipulable without altering the DNA/RNA sequence. We defined several criteria based on a model that predicts (i) kinetic coupling of splicing to chromatin states, characterized by (ii) a decrease in H3K79me2 (representative of marks in the third principal component) and/or an increase in H3K36me3 levels, proportional to (iii) the amount of exon inclusion. In our first case of regulated alternative splicing, YPEL5 exon 2 inclusion increases when cells are treated with caffeine (Shi, Hu, Pabon, & Scotto, 2008). This event fits our criteria, as it is (i) regulated within a few hours of treatment, (ii) sufficiently distant from constitutive internal exons to measure effects on histone modifications, and (iii) "switch-like" in that the change in inclusion in response to caffeine is dramatic.

To assess both YPEL5 expression and alternative splicing, we measured total mRNA by qPCR and the ratio of mRNA isoforms by RT-PCR in various concentrations of caffeine. After 7 hours exposure, mRNA levels were similar between 18 mM caffeine and 6 mM (Figure 2.11), while exon 2 inclusion increased > 50-fold from 1% inclusion to 66% inclusion (Figure 2.4A and 2.11). We verified that our ChIP-qPCR at *YPEL5* spatially recapitulated the profiles of H3K79me2, H3K36me3, and nucleosomes in the ChIP-seq and MNase-seq data (Figure 2.11). Notably, H3K36me3 and H3K79me2 profiles were similar at the low and high caffeine concentrations (Figure 2.4B). Therefore, the dramatic change in the splicing pattern of YPEL5 RNA was not sufficient to direct changes in these histone modifications.

As with many alternative splicing events, little is known about the mechanism of YPEL5 exon 2 inclusion. Moreover, we were constrained to short treatment periods in the experiments, as high levels of caffeine are toxic. Hence, we sought a second example of regulated alternative splicing, in which changes in alternative splicing could be maintained over many cell divisions. In lymphocytes, the RNA-binding protein hnRNPLL binds to CD45 pre-mRNA to promote cell-type specific exon-skipping of multiple CD45 exons (Oberdoerffer et al., 2008; Topp, Jackson, Melton, & Lynch, 2008; Z. Wu et al., 2008). As hnRNPLL is non-essential, its levels can be manipulated in cell lines, resulting in alternative CD45 splicing differences that are stably maintained. Although B cells typically express long isoforms of CD45, the B cell lymphoma lines BJAB and BL41 express different levels of hnRNPLL, and consequently, different CD45 protein isoforms (Oberdoerffer et al., 2008).

We obtained a set of matched BJAB and BL41 cell lines in which hnRNPLL levels are stably decreased by short hairpin RNA interference or increased by hnRNPLL over-expression, respectively (Oberdoerffer et al., 2008). BJAB and BL41 cells incidentally express 100-fold different levels of CD45 mRNA, so we compared each cell line to its matched line where hnRNPLL levels were specifically modulated without changing total CD45 mRNA levels (Figure 2.5A). RT-PCR analysis to assay variable inclusion of exons 4-6 confirmed that depletion of hnRNPLL in BJAB cells increased inclusion, while overexpression of hnRNPLL in BL41 cells decreased inclusion (Figure 2.5B). H3K36me3, H3K79me2, nucleosome levels at CD45 measured by ChIP-qPCR were generally similar to published ChIP-seq and MNase-seq data in T cells (Figure 2.12). If RNA splicing were sufficient to change histone modifications, increased exon inclusion at exons 4-6 in hnRNPLL-depleted cells should increase H3K36me3 and/or decrease H3K79me2 levels. In fact, we observed no significant difference in histone modifications between matched pairs of either BJAB or BL41 cells expressing different CD45 isoforms (Figure 2.5C, D).

2.5 Discussion

We used PCA to help identify groups of histone marks (Figure 2.1), among 41 modifications and variants, which displayed distinct profiles within active genes (Figure 2.7). In addition to the previously defined group, mostly acetylations, which makes up the histone modification "backbone" at promoters (Zhibin Wang, Zang, et al., 2008), we identified a group of mostly histone methylations biased for 5’ introns and roughly reciprocal with H3K36me3 near exons (Figure 2.2). Intronic and exonic histone modification enrichment was distinguishable from bulk nucleosome occupancy (Figure

2.3). We found that bone fide exons, which are associated with a multitude of information used for splicing, better explained many histone modification patterns than exon-like nucleosome occupancy alone (Figure 2.10). The genome-wide phenomenon of intronic and exonic marks suggests that a general mechanism associated with splicing influences histone marking, as speculated previously (Kolasinska-Zwierz et al., 2009). Thus, we tested if levels of RNA splicing influenced the profiles of these histone modifications. Neither "switch-like" changes in splicing frequency of a single exon in YPEL5 (Figure 2.4), nor stable changes in the inclusion of multiple sequential exons in CD45 (Figure 2.5) resulted in significant changes in H3K79me2 or H3K36me3. Importantly, changes in H3K36me3 and H3K79me2 levels occur at immediate early genes within one hour of transcriptional induction (Edmunds et al., 2008), yet we observed no change after several hours of alternative splicing induction for YPEL5 or after numerous cell divisions of splicing differences in CD45.

Our results rule out a simple correlation between the extent of RNA splicing and histone modification levels. Instead, we suggest that co-transcriptional events promoting exon definition may influence histone modifications independent of the extent of splicing. Intriguingly, regulated skipping of CD45 exon 4 is achieved by stalling spliceosome assembly at an early step following exon definition (House & Lynch, 2006); therefore, intermediate spliceosome assembly may be sufficient to decrease intron-enriched modifications and/or increase H3K36me3 at CD45 exon 4, regardless of splicing outcomes. Such a scenario may be more common than previously thought, as diverse RNA binding proteins have been implicated in regulating alternative splicing during a

transition from exon definition to intron definition (Bonnal et al., 2008; Feng, Chen, & Manley, 2008; Sharma, Kohlstaedt, Damianov, Rio, & Black, 2008). As it is now recognized that promoter marks are relatively stable (Raisner et al., 2005; Kouskouti & Talianidis, 2005) and may reflect transcription initiation (Guenther, Levine, Boyer, Jaenisch, & Young, 2007) rather than productive transcription, we suggest by analogy that intronic and exonic marks are relatively stable and may reflect exon definition, rather than productive splicing.

If exon definition or another feature related to splicing were required to influence histone modifications, we would expect a decrease in 5' intron marks and an increase in H3K36me3 specifically at *bone fide* exons, but not at ECRs, which have similar sequence composition, but are not spliced. Indeed, ECRs do not direct increased levels of H3K36me3 as at exons or decreased levels of certain 5' intronic marks (i.e. H3K4me1, H3K4me2, H3K9me1, H3K23ac; Figure 2.10). In contrast, some 5' intronic marks (e.g. H3 lysine 79 methylations) have nearly equivalent decreases near exons and ECRs, suggesting that the local sequence composition of exons (and ECRs) may be sufficient to influence these histone modification profiles. Exons and introns are subject to a range of sequence constraints to maintain proper splicing, translation, and non-coding regulation. Thus, *cis*-acting sequences, functioning as DNA or RNA, could signal to change histone modifications. For example, SR protein binding sites at exons could affect chromatin marking in a process that has been termed “reverse” coupling (Zhong et al., 2009). Additionally, elevated nucleosome occupancy at exons (Andersson et al., 2009; Nahkuri et al., 2009; Schwartz et al., 2009; Spies et al., 2009; Tilgner et al., 2009)

may play a role in signaling to change histone modifications, independently or in conjunction with *cis*-acting sequences. Extensive mutagenesis will be essential to distinguish between these possibilities and identify necessary sequences.

How distinct histone modification patterns are determined remains largely unknown. In agreement with Spies et al., we find that histone modification patterns do not merely reflect nucleosome enrichment at exons, indicating the need for additional steps of regulation beyond nucleosome occupancy (Figure 2.3). We propose a model in which human introns and exons modulate the activity and/or localization of histone modifying enzymes, several of which are known to be associated with the elongating RNA polymerase. Conceivably, such regulation could reflect changes in the phosphorylation of Ser2 residues within the RNA polymerase II C-terminal domain (CTD), which occur during elongation and influence co-transcriptional processes (Brès, Yoh, & Jones, 2008). For example, H2Bub requires the ongoing action of the P-TEFb CTD Ser2 kinase (Pirngruber et al., 2009), and the Set2 H3K36 trimethyltransferases contain a domain that binds directly to the CTD, preferring at least 2 contiguous repeats phosphorylated at Ser2 and Ser5 (M. Li et al., 2005; Vojnic, Simon, Strahl, Sattler, & Cramer, 2006). Spatial similarities among the intronic marks could reflect cross-regulation between histone modifications (Latham & Dent, 2007). H2Bub is catalyzed by the BRE1 ubiquitin ligase and is localized in the 5' regions of active genes in humans (J. Kim, Hake, & Roeder, 2005),(Zhu et al., 2005; Minsky et al., 2008). Our analysis, which shows extensive overlap between H2Bub and H3K79me2/3, is consistent with direct stimulation by H2Bub of H3 lysine 79 methyltransferase activity by DOT1L and a DOT1L-

containing protein complex (McGinty, Kim, Chatterjee, Roeder, & Muir, 2008; Mohan et al., 2010). Also, H2Bub can directly stimulate human SET1-mediated methylation of H3 lysine 4 *in vitro*, so H2Bub within 5' introns may help to increase H3K4me1 and H3K4me2 locally (J. Kim et al., 2005; J. Kim et al., 2009).

We propose that intronic and exonic intragenic histone modification regions, along with the "backbone" at promoters constitute at least three distinct chromatin zones at active human genes (B. M. Lee & Mahadevan, 2009; Zhibin Wang, Zang, et al., 2008), analogous to those found in yeast(B. Li et al., 2007; T. Kim & Buratowski, 2009): i) the "backbone" or promoter zone contains H3K4me3, ii) a 5' zone includes H3K4me2 (and H3K79me2 in humans), and iii) a 3' zone contains H3K36me3 (Figure 2.6). However, most yeast genes are intronless and only a few nucleosomes in length, whereas human genes generally contain many introns and the 5' zone is expanded variably into long introns. H3K36me3 is high at exons, which are frequently grouped at the 3' end of human genes (Andersson et al., 2009) due to the tendency of genes to have long 5' introns (Bradnam & Korf, 2008). We find in some cases, these different zones may interdigitate, *e.g.* there is switching back and forth between H3K79me2 and H3K36me3 within *STIM1* (Figure 2.9). This implies that these zones are not strictly 5' and 3' with respect to genic positioning, but rather that introns and exons act to organize intragenic histone modifications. Depending on the exon-intron structure of long intron-containing genes in humans (Figure 2.6), the resulting spatial mixtures of modifications could differentially regulate the expression of their respective genes.

Intronic and exonic histone modifications can, in principle, affect transcription elongation, co-transcriptional RNA processing, and/or termination and export (Perales & Bentley, 2009). For example, H3K36me3-modified nucleosomes may directly influence splicing, as evidenced by changes in alternative splicing upon depletion of human SETD2 or of the H3K36me3-binding protein MRG15 (Luco et al., 2010). Furthermore, H3K36me3 levels on average are slightly lower at alternatively spliced exons compared to nearby constitutive exons (Kolasinska-Zwierz et al., 2009; Schwartz et al., 2009; Hon, Wang, & Ren, 2009). An open question for future studies remains: how are these marks established and maintained at particular introns and exons within the same transcribed unit? A working knowledge of such localization mechanisms will be paramount to understanding temporal or spatial regulation of these marks.

2.6 Author Contributions

J.T.H. and A.M.P. designed and performed the analyses and experiments. J.T.H., A.M.P., C.G., and K.R.Y. wrote the manuscript.

2.7 Acknowledgements

We thank the following for kindly providing reagents and data: Shalini Oberdoerffer and Anjana Rao for B cell lines, Noah Spies and Chris Burge for ECR locations, and Efrat Shema and Moshe Oren for H2Bub ChIP-seq data. We also thank Hiten Madhani and Joan Steitz for critical reading of the manuscript and the Guthrie, Yamamoto, and Panning groups for helpful discussions. J.T.H. and A.M.P. were supported by individual ARCS Foundation Scholarships. Research support was provided by NIH grants GM21119

to C.G. and CA020535 to K.R.Y. C.G. is an American Cancer Society Research Professor of Molecular Genetics and K.R.Y. is a consultant with Merck & Co., Inc.

2.8 Figures and Tables

Figure 2.1 PCA analysis of ChIP-seq reveals three major components

3D stereoscopic image of PCA of all ChIP-seq and MNase-seq data 10 kb up- and downstream of transcription start sites. Gray points represent the PCA loadings for each of the 41 histone marks and nucleosome occupancy (Nucs). Principal components (Comp.) 1, 2, and 3 are projected to the lower right, lower left, and top, respectively, with the orientation of the axes shown as gray bars radiating from the origin. Individual points representative of each component are colored: H3K4me3 (orange) for Comp. 1, Nucs (purple) for Comp. 2, and H3K79me2 (green) for Comp. 3. For reference, marks enriched at promoters are colored light orange and H3K36me3 is colored blue. Figure 2.7 shows the fully labeled and separated PCA loadings and the transformed PCA scores plotted on the genome with respect to transcription start sites.

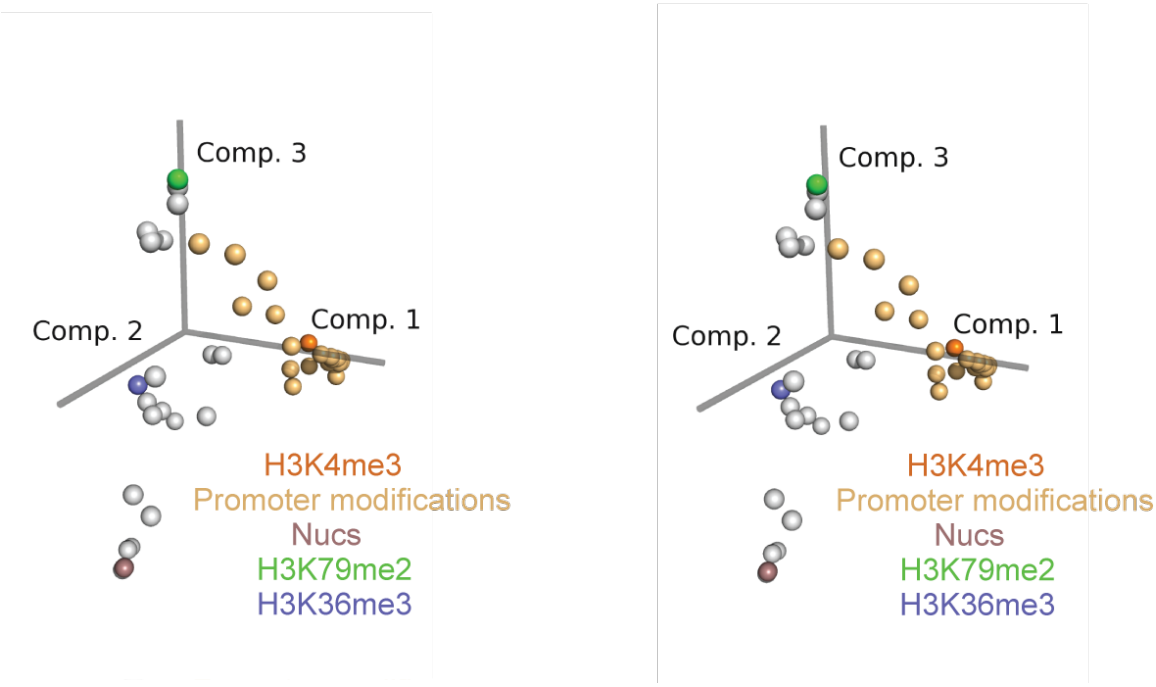


Figure 2.2 Exonic and intronic marks in the human epigenome

Heatmaps show the density for H3K4me3, nucleosomes (Nucs), H3K79me2, and H3K36me3 ChIP- and MNase-seq, plotted with respect to transcription start sites (bent arrows) with genes sorted by distances to the beginning of the first internal exons (i.e. first 3' splice sites). For clarity only the 4286 annotated RefSeq genes that are in the top 50% of expression and with the 50% longest distances to the first internal exon are shown. Full plots are shown in Figure 2.8. Each gene is displayed as a row and columns are 1-kb bins from 10 kb upstream to 50 kb downstream of transcription start sites (bottom). The value of each bin is shaded by the number of ChIP- or MNase-seq reads per kb, scaled to cover data between 0 reads (white) and the 99th percentile of each sequencing experiment (black), indicated at the top right of each heatmap. At far right the aligned transcription start sites (broken line) and the positions of first internal exons (gray line) are shown as a visual reference.

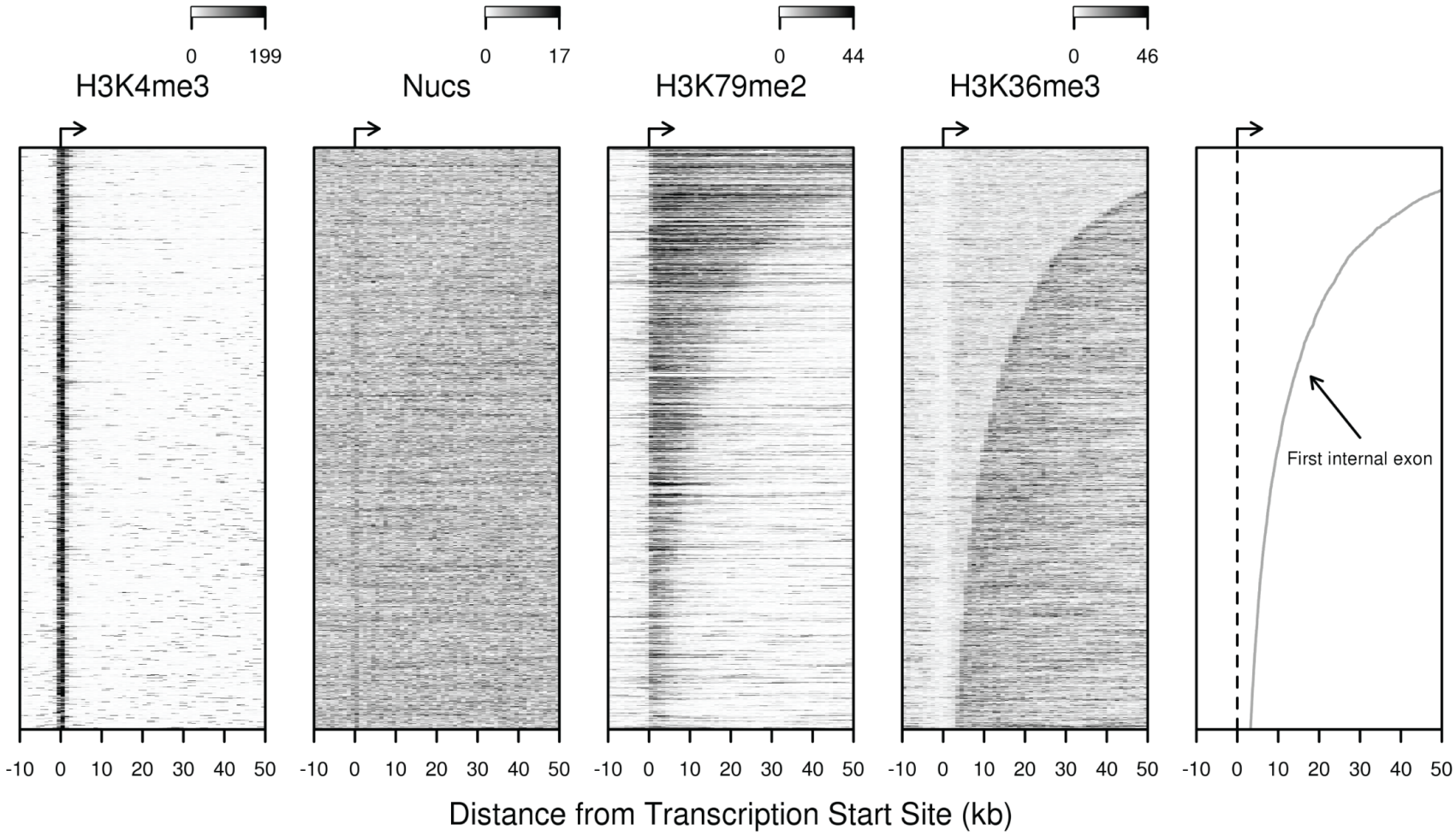

0
199
H3K4me3
0
17
Nucs
0
44
H3K79me2
0
46
H3K36me3
First internal exon
-10 0 10 20 30 40 50
Distance from Transcription Start Site (kb)

Figure 2.3 Exonic and intronic marks do not merely reflect nucleosome occupancy

Effect of spatially normalizing histone modifications (ChIP-seq) to nucleosome occupancy (MNase-seq). *(A)* The read density of ChIP-seq (H3K79me2, green; H3K36me3, blue) and MNase-seq (Nucs, purple) 10 kb up- and downstream of exons was estimated with a kernel density estimator (see Methods for details). A uniform density distribution would be equal to 1. *(B)* The ChIP-seq densities were divided by the MNase-seq density from *(A)*, thus giving modification density normalized as a log2(ratio) relative to nucleosome occupancy (H3K79me2/Nucs, green; H3K36me3/Nucs, blue). In both panels the average size of the centered exon is depicted with a vertical gray bar.

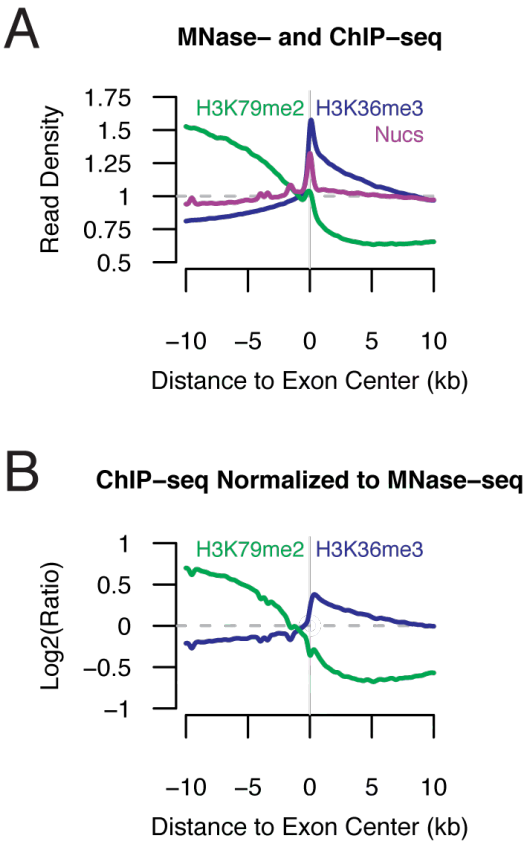


Figure 2.4 Histone modifications are stable despite transient changes in alternative splicing

Histone modification profiles at the alternatively spliced exon of *YPEL5* are similar between caffeine-treated SW620 cells expressing different YPEL5 mRNA isoforms. *(A)* Increasing caffeine concentrations result in higher exon inclusion of YPEL5 exon 2. RT-PCR of SW620 cells treated for 7 hours with 0, 6, or 18 mM caffeine. Alternatively (red) and constitutively (black) spliced exons are numbered and diagrammed to the right of the corresponding RT-PCR product. *(B)* Normalized H3K79me2 and H3K36me3 profiles in SW620 cells treated with 0 (orange), 6 (green), and 18 (blue) mM caffeine for 7 hours. Points indicate log2(ratios) of H3K79me2/H3 (top) and H3K36me3/H3 (bottom) ChIP-qPCR measurements, relative to the lowest value, for each of two SW620 biological replicates. Vertical gray lines provide a visual reference for the 3' splice site of each internal exon. Arrowheads indicate the positions of qPCR amplicons along the gene (bottom). The alternatively spliced cassette exon 2 is indicated in red. The gene schematic and markers are drawn to scale according to their genomic coordinates except for the second (green) and third (blue) data points at each amplicon, which are offset to the right to facilitate comparison between conditions.

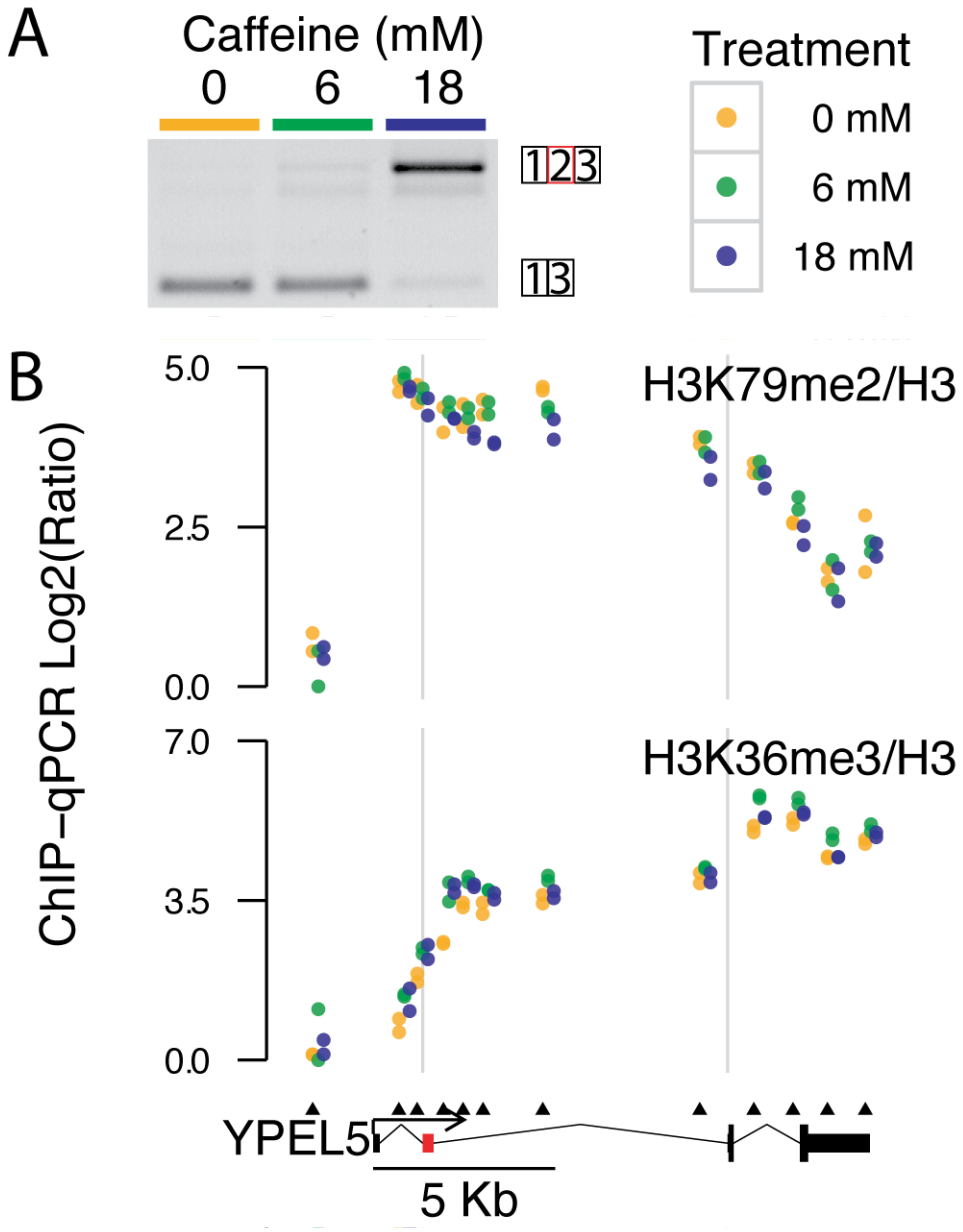

A
Caffeine (mM)
0
6
18
1 2 3
1 3
Treatment
0 mM
6 mM
18 mM
B
ChIP–qPCR Log2(Ratio)
5.0
2.5
0.0
H3K79me2/H3
7.0
3.5
0.0
H3K36me3/H3
YPEL5
5 Kb

Figure 2.5 Histone modifications are stable despite stable changes in alternative splicing

Histone modification profiles at the alternatively spliced exons of CD45 are similar between B cell lines stably expressing different CD45 mRNA isoforms. In all panels, colored points and lines refer to matched BJAB B cell lines expressing a short hairpin targeting knockdown of either RFP as a control (blue) or hnRNPLL (green) or to matched BL41 B cell lines overexpressing hnRNPLL (orange) or mock-treated as a control (red). (A) BJAB and BL41 B cell lines differ in the relative abundance of CD45 and hnRNPLL mRNA. Points indicate mean log2(ratios) +/- 1 SD of CD45/RPLP0 and hnRNPLL/RPLP0 qRT-PCR measurements from (n=4) biological replicates (not detectable; ND). (B) Confirmation that passaged BJAB and BL41 B cell lines express different CD45 isoforms. Alternatively (red) and constitutively (black) spliced exons are diagrammed to the right of the corresponding RT-PCR product. (C) and (D) Normalized H3K79me2 and H3K36me3 profiles in matched B cell lines. Points indicate the mean log2(ratios) +/- 1 SD of H3K79me2/H3 (top) and H3K36me3/H3 (bottom) ChIP-qPCR measurements, relative to the lowest value, from (C) BJAB (n=5) and (D) BL41 (n=6) biological replicates. Positions of the 3' splice sites and qPCR amplicons along the gene are displayed as vertical grey lines and arrowheads, respectively. All positions are drawn to scale, except for a gap in the second intron and the second point at each amplicon, which is offset to the right to facilitate comparison.

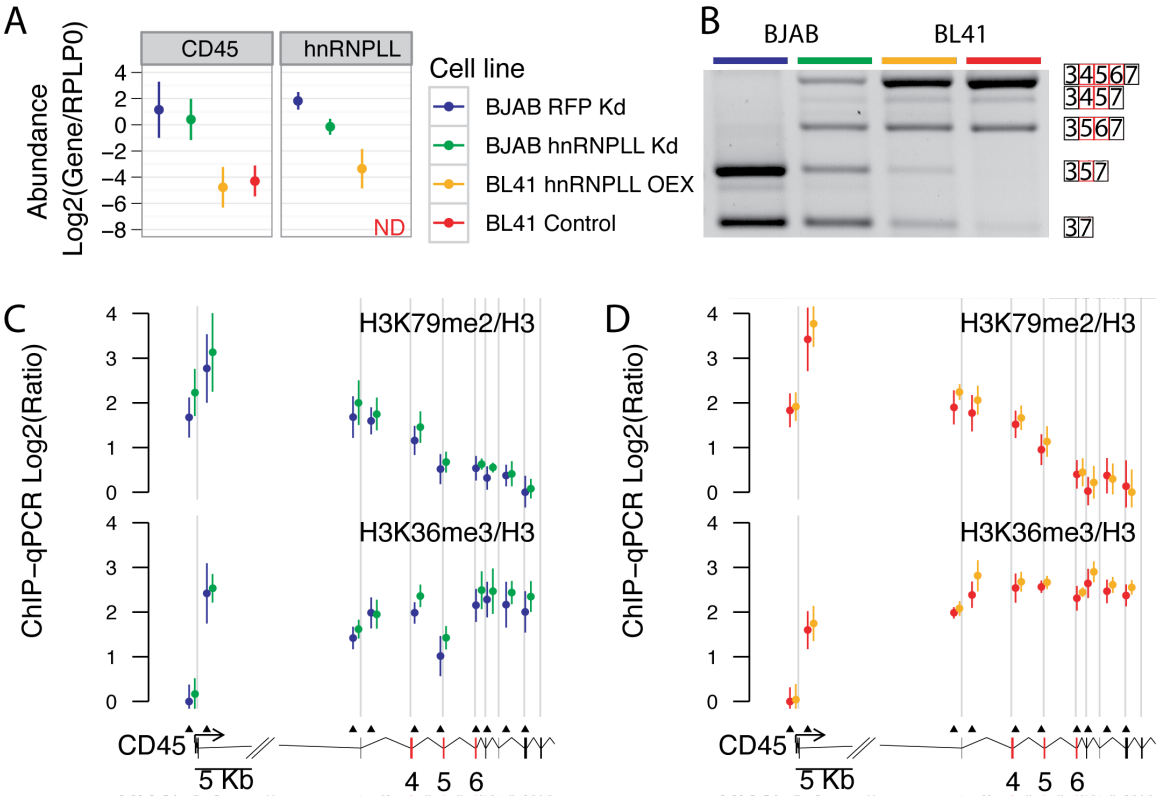

A
Abundance
Log2(Gene/RPLP0)
CD45
hnRNPLL
ND
Cell line
BJAB RFP Kd
BJAB hnRNPLL Kd
BL41 hnRNPLL OEX
BL41 Control
B
BJAB
BL41
3456 7
3457
3567
357
37
C
ChIP–qPCR Log2(Ratio)
H3K79me2/H3
H3K36me3/H3
CD45
5 Kb
4 5 6
D
ChIP–qPCR Log2(Ratio)
H3K79me2/H3
H3K36me3/H3
CD45
5 Kb
4 5 6

Figure 2.6 Chromatin structure reflects gene architecture

Intragenic modification zones reflect gene architecture. *(A)* H3K4me3, *(B)* H3K79me2, and *(C)* H3K36me3 ChIP-seq reads plotted on two genes (*UGCGL1* and *SMURF1*) of similar length, but different gene architectures. A black line in each density plot shows where a uniform distribution of sequence reads would be. The proposed promoter, 5' intronic, and 3' exonic histone modification zones are diagrammed below each gene.

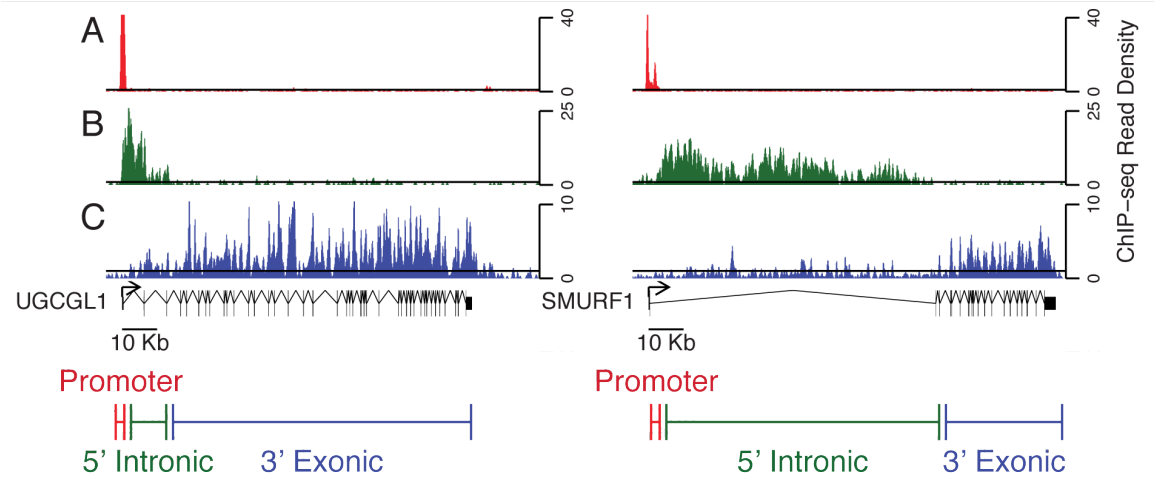
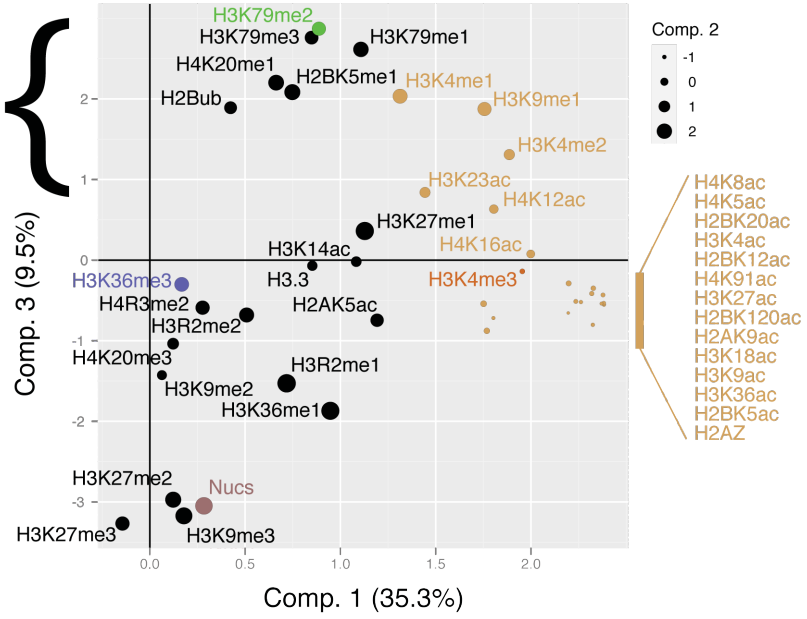
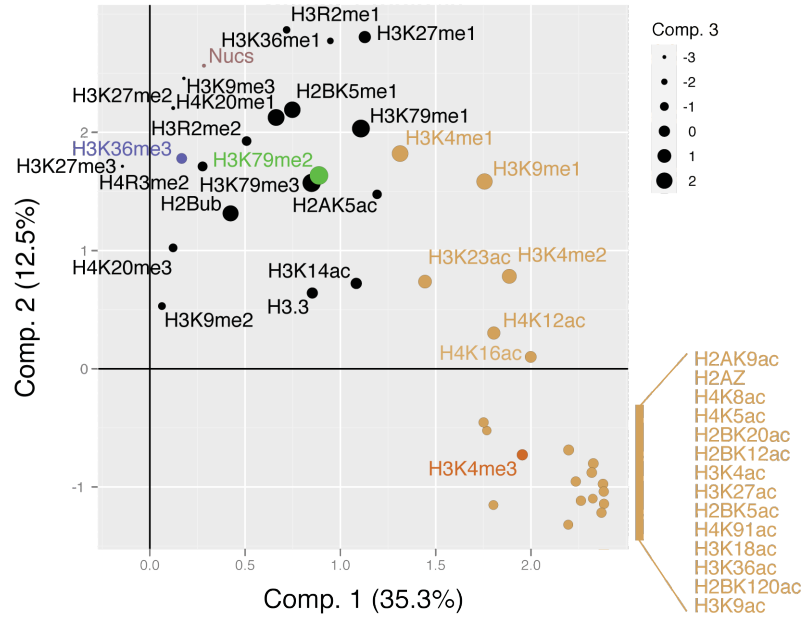


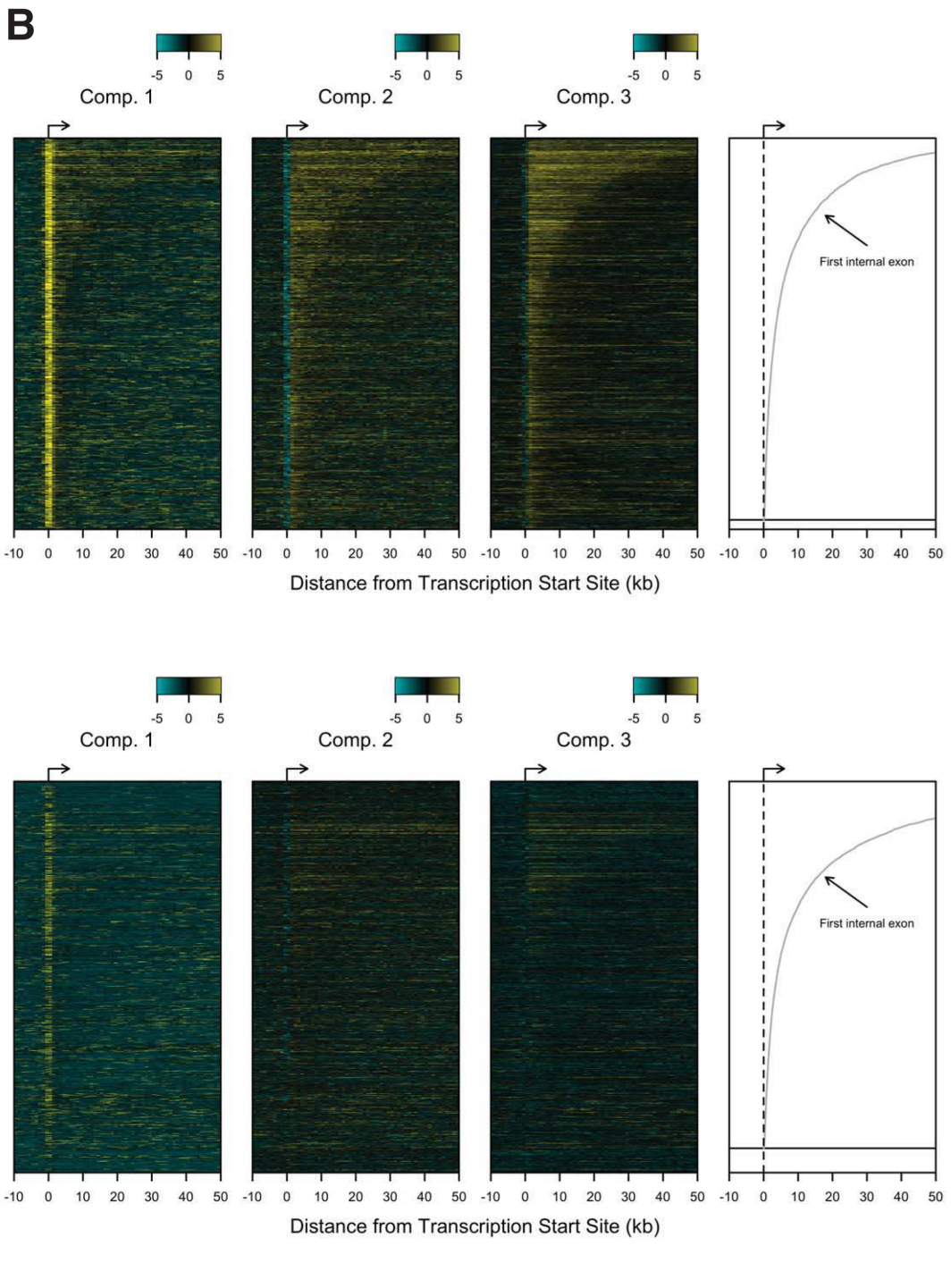
Figure 2.7 PCA of ChIP-seq data

PCA for all ChIP-seq and MNase-seq data 10 kb up- and downstream of transcription start sites. (A) Fully labeled loadings for components one, two, and three, colored as in Figure 2.1. For each projection two axes are plotted and the third is presented as point size, indicated at right. The curly bracket at top left highlights those marks high in component three. The components account for 35.3, 12.5, and 9.5% of the total variance in the dataset, respectively. (B) Heatmaps show the PCA scores, plotted with respect to transcription start sites (bent arrows) with genes sorted by distances to the beginning of the first internal exons (i.e. first 3' splice sites). Annotated RefSeq genes in the top 50% (top row) or bottom 50% (bottom row) of expression (8572 each) are shown. Each gene is displayed as a row and columns are 1-kb bins from 10 kb upstream to 50 kb downstream of transcription start sites (bottom). The value of each bin is shaded by the PCA score, indicated at the top right of each heatmap. At far right the aligned transcription start sites (broken line) and the positions of first internal exons (gray line) are shown as a visual reference. The horizontal black line at far right shows the break between intron-containing (above the line) and intronless (below the line) genes. (C) A scree plot shows the variance accounted for by all 42 principal components, increasing numerically from left-to-right, demonstrating that the first three components (black) explain most of the variance in the dataset: 57.3% total between the first three.

A


H3K79me2
H3K79me3
H3K79me1
H4K20me1
H2BK5me1
H3K4me1
H2Bub
H3K9me1
H3K4me2
H3K23ac
H4K12ac
H3K27me1
H3K14ac
H4K16ac
H3K36me3
H3.3
H3K4me3
H4R3me2
H2AK5ac
H3R2me2
H4K20me3
H3R2me1
H3K9me2
H3K36me1
H3K27me2
Nucs
H3K27me3
H3K9me3
Comp. 3 (9.5%)
Comp. 1 (35.3%)
Comp. 2
-1
0
1
2
H4K8ac
H4K5ac
H2BK20ac
H3K4ac
H2BK12ac
H4K91ac
H3K27ac
H2BK120ac
H2AK9ac
H3K18ac
H3K9ac
H3K36ac
H2BK5ac
H2AZ


H3R2me1
H3K36me1
H3K27me1
Nucs
H3K9me3
H2BK5me1
H3K27me2
H4K20me1
H3K79me1
H3R2me2
H3K4me1
H3K36me3
H3K79me2
H3K9me1
H3K27me3
H4R3me2
H3K79me3
H2Bub
H2AK5ac
H4K20me3
H3K23ac
H3K4me2
H3K14ac
H3.3
H4K12ac
H3K9me2
H4K16ac
H3K4me3
Comp. 2 (12.5%)
Comp. 1 (35.3%)
Comp. 3
-3
-2
-1
0
1
2
H2AK9ac
H2AZ
H4K8ac
H4K5ac
H2BK20ac
H2BK12ac
H3K4ac
H3K27ac
H2BK5ac
H4K91ac
H3K18ac
H3K36ac
H2BK120ac
H3K9ac


B
-5 0 5
Comp. 1
Comp. 2
Comp. 3
First internal exon
-10 0 10 20 30 40 50
Distance from Transcription Start Site (kb)

C

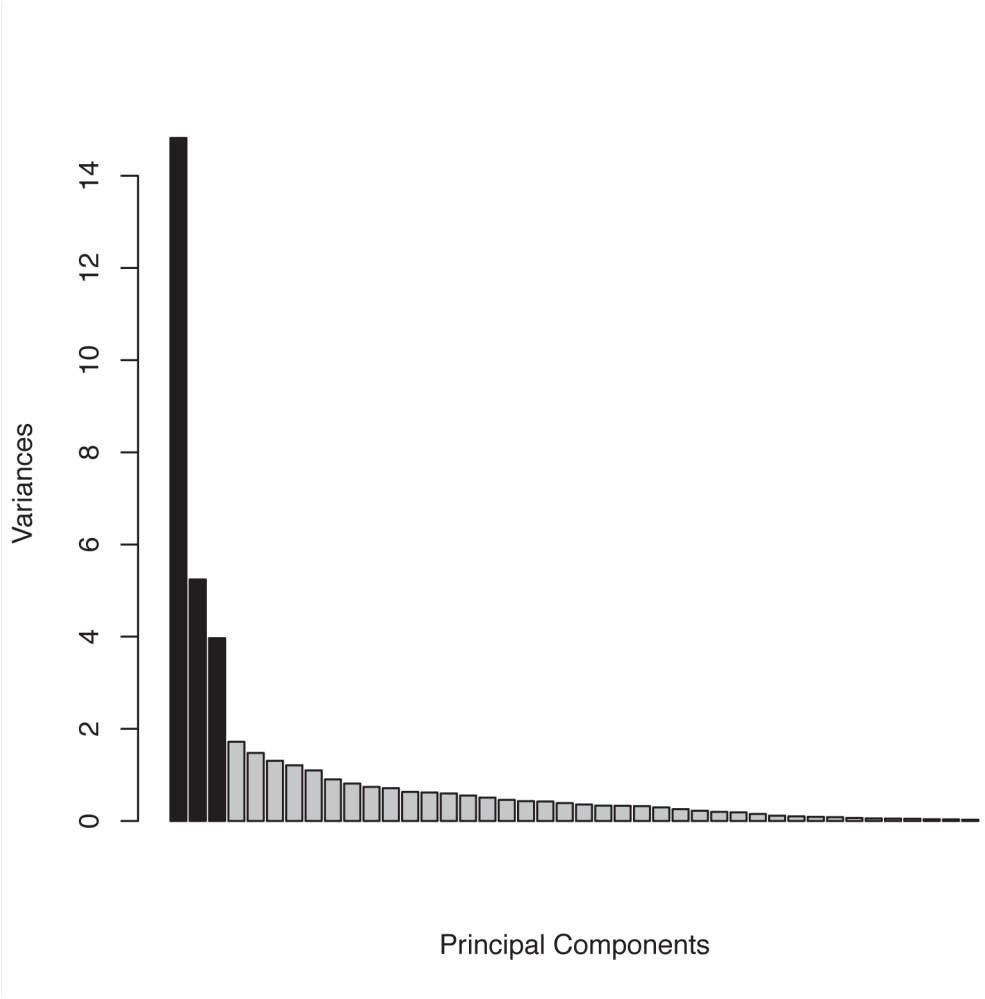
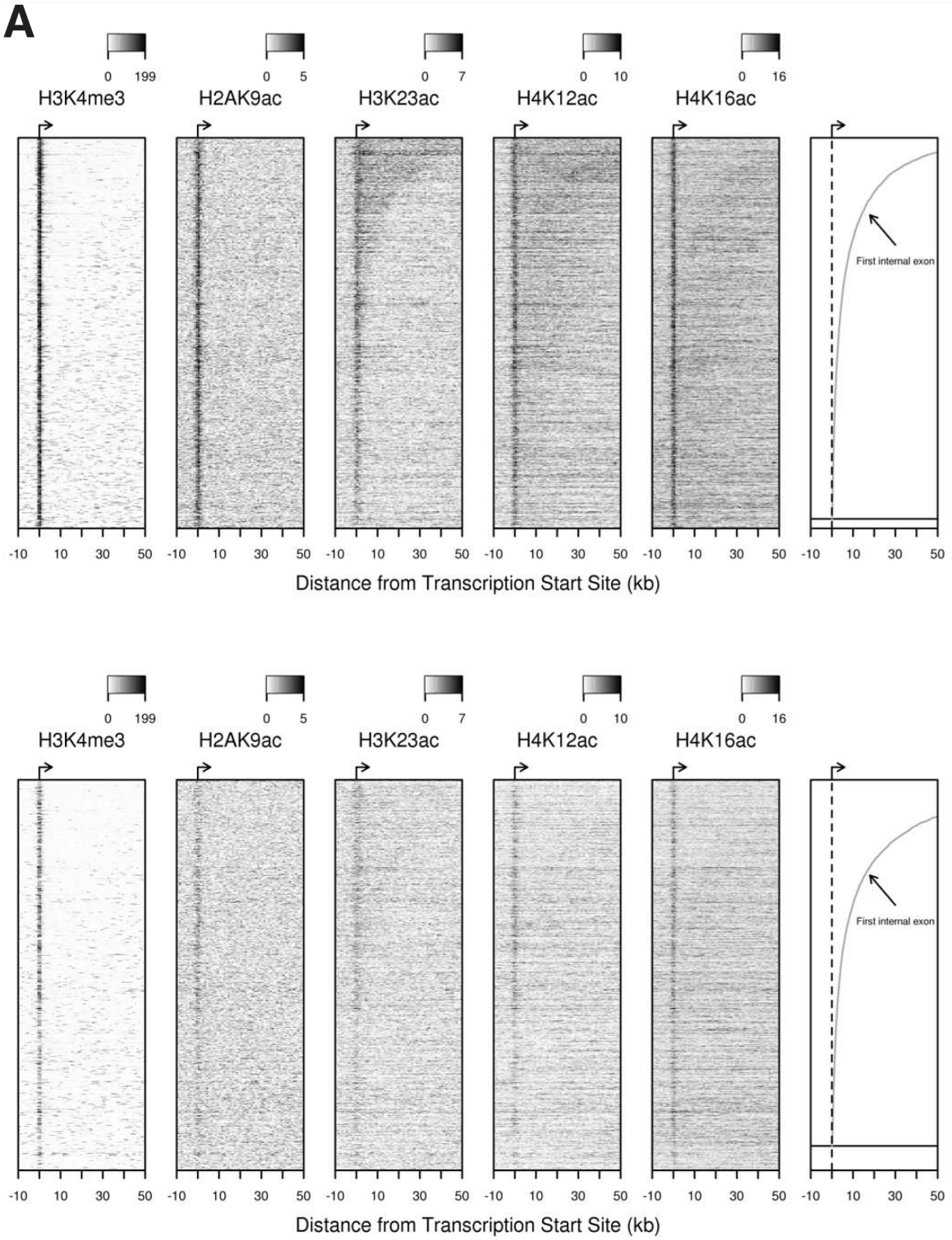
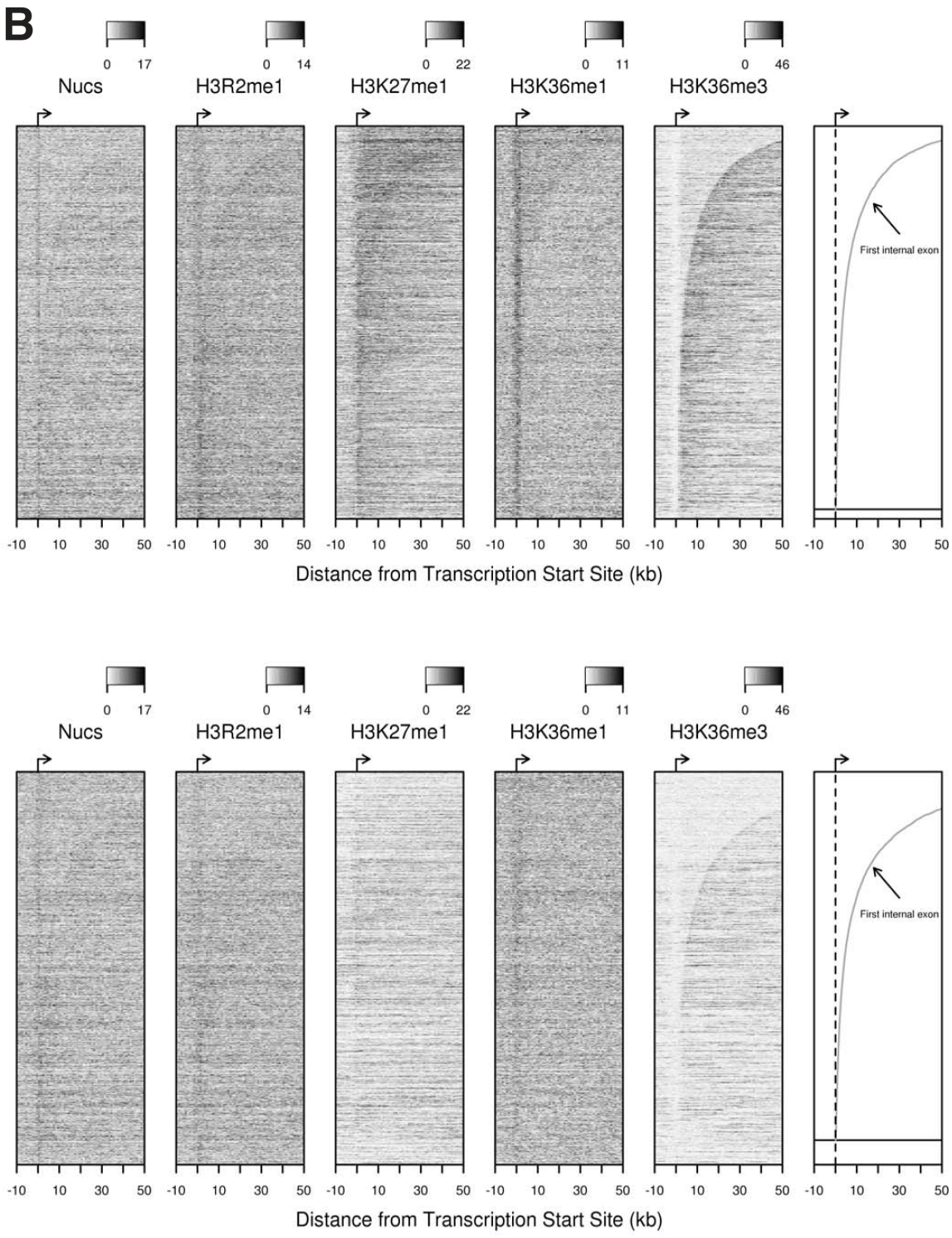
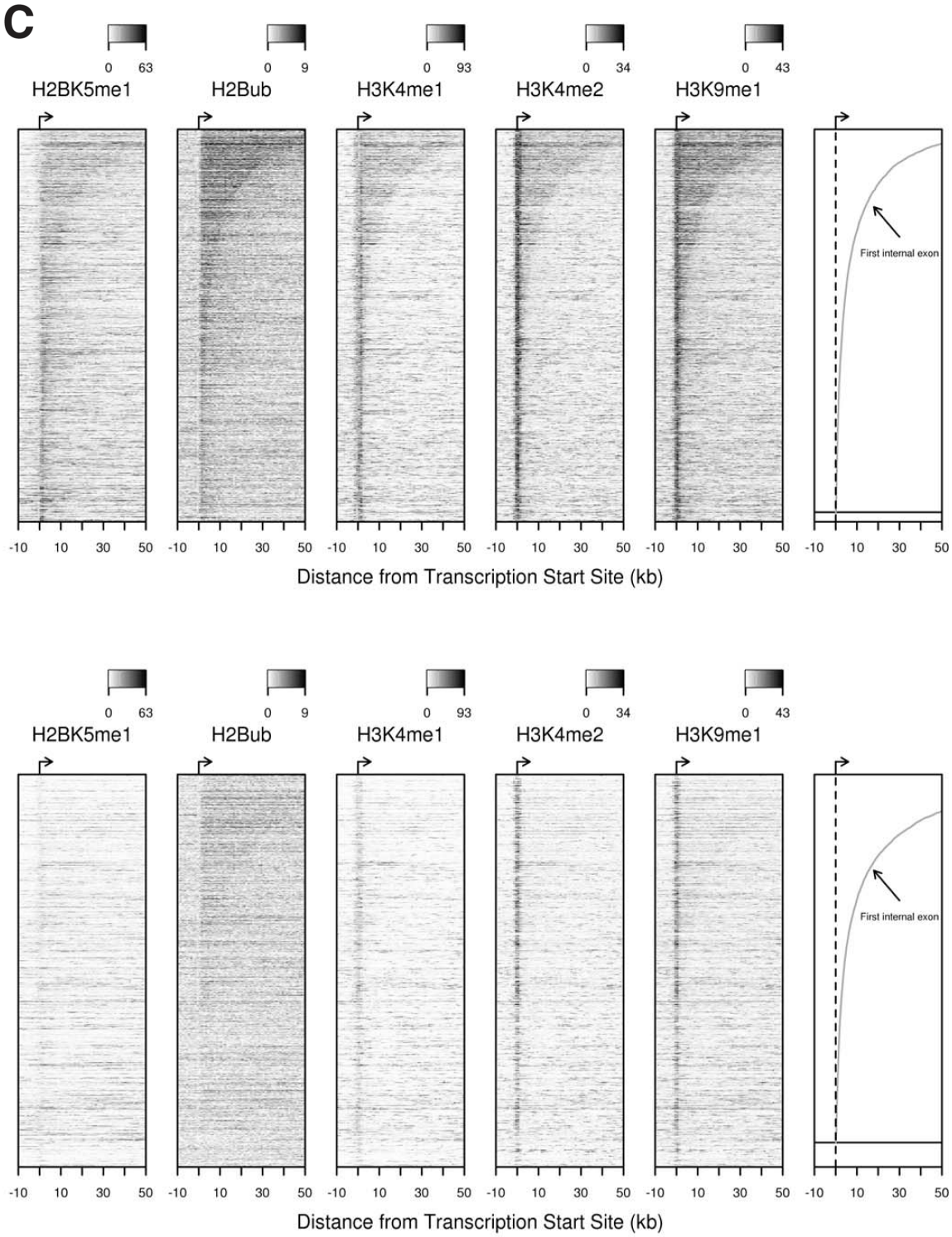


Figure 2.8 Chromatin structure reflects gene architecture (all marks)

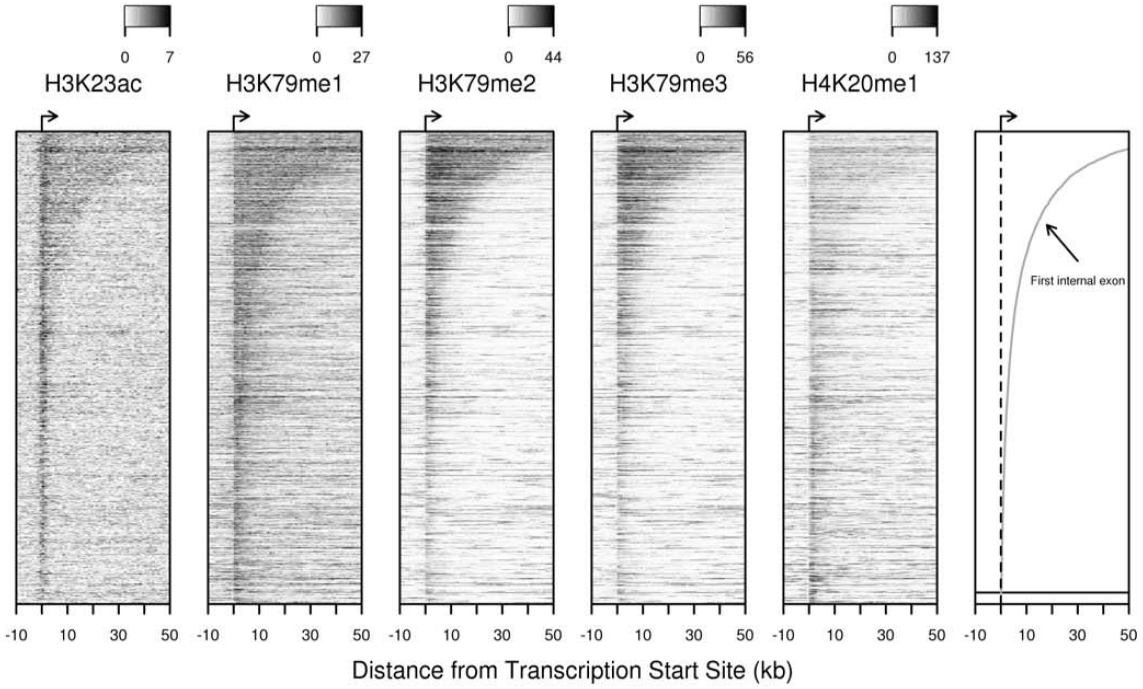
Heatmaps show the density of representative marks, plotted with respect to transcription start sites (bent arrows) with genes sorted by distances to the beginning of the first internal exons (i.e. first 3' splice sites). (A) Promoter marks, including 4 marks not previously described as promoter-enriched: H2AK9ac, H3K23ac, H4K12ac, and H4K16ac. (B) Marks high in component two, including Nucs and H3K36me3. (C) All ten marks high in component three that display intronic enrichment. For all plots, annotated RefSeq genes in the top 50% (top row) or bottom 50% (bottom row) of expression (8572 each) are shown. Each gene is displayed as a row and columns are 1-kb bins from 10 kb upstream to 50 kb downstream of transcription start sites (bottom). The value of each bin is shaded by the number of ChIP- or MNase-seq reads per kb, scaled to cover data between 0 reads (white) and the 99th percentile of each sequencing experiment (black), indicated at the top right of each heatmap. At far right the aligned transcription start sites (broken line) and the positions of first internal exons (gray line) are shown as a visual reference. The horizontal black line at far right shows the break between intron-containing (above the line) and intronless (below the line) genes.


A
0 199
H3K4me3
0 5
H2AK9ac
0 7
H3K23ac
0 10
H4K12ac
0 16
H4K16ac
First internal exon
-10 10 30 50
Distance from Transcription Start Site (kb)
0 199
H3K4me3
0 5
H2AK9ac
0 7
H3K23ac
0 10
H4K12ac
0 16
H4K16ac
First internal exon
-10 10 30 50
Distance from Transcription Start Site (kb)


B
0 17
0 14
0 22
0 11
0 46
Nucs
H3R2me1
H3K27me1
H3K36me1
H3K36me3
First internal exon
-10 10 30 50
Distance from Transcription Start Site (kb)


C
0 63
0 9
0 93
0 34
0 43
H2BK5me1
H2Bub
H3K4me1
H3K4me2
H3K9me1
First internal exon
-10 10 30 50
Distance from Transcription Start Site (kb)

C (cont'd)


0 7
0 27
0 44
0 56
0 137
H3K23ac
H3K79me1
H3K79me2
H3K79me3
H4K20me1
First internal exon
-10 10 30 50
Distance from Transcription Start Site (kb)

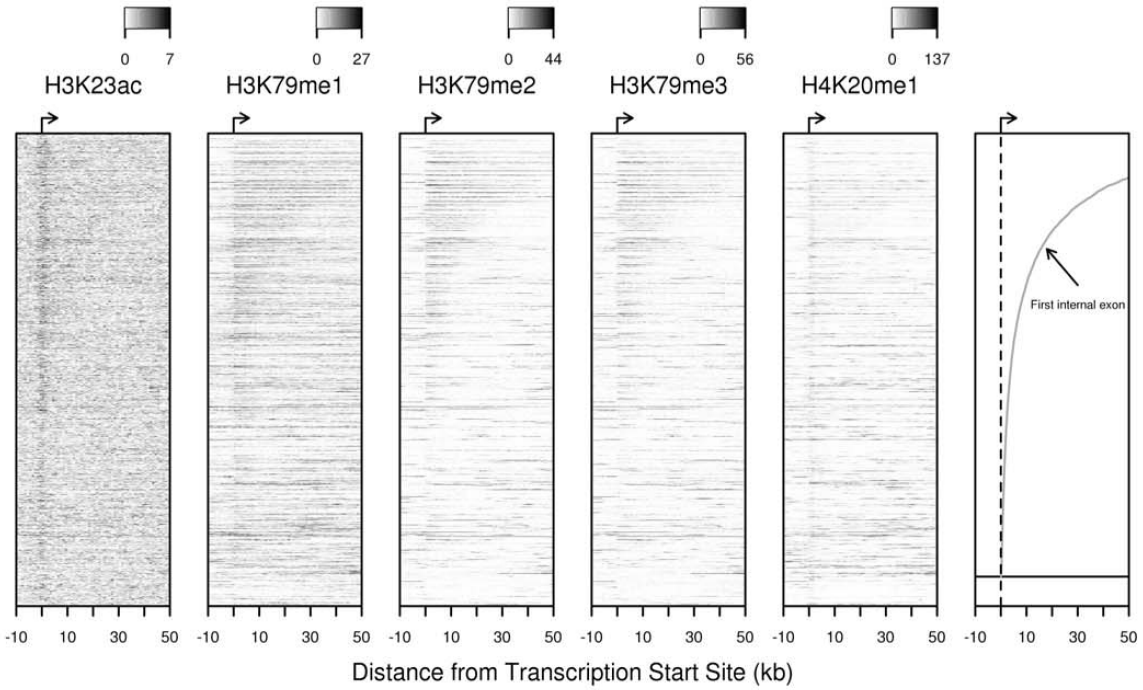

0 7
0 27
0 44
0 56
0 137
H3K23ac
H3K79me1
H3K79me2
H3K79me3
H4K20me1
First internal exon
-10 10 30 50
Distance from Transcription Start Site (kb)

Figure 2.9 ChIP-qPCR validation of ChIP-seq profiles at the STIM1 gene

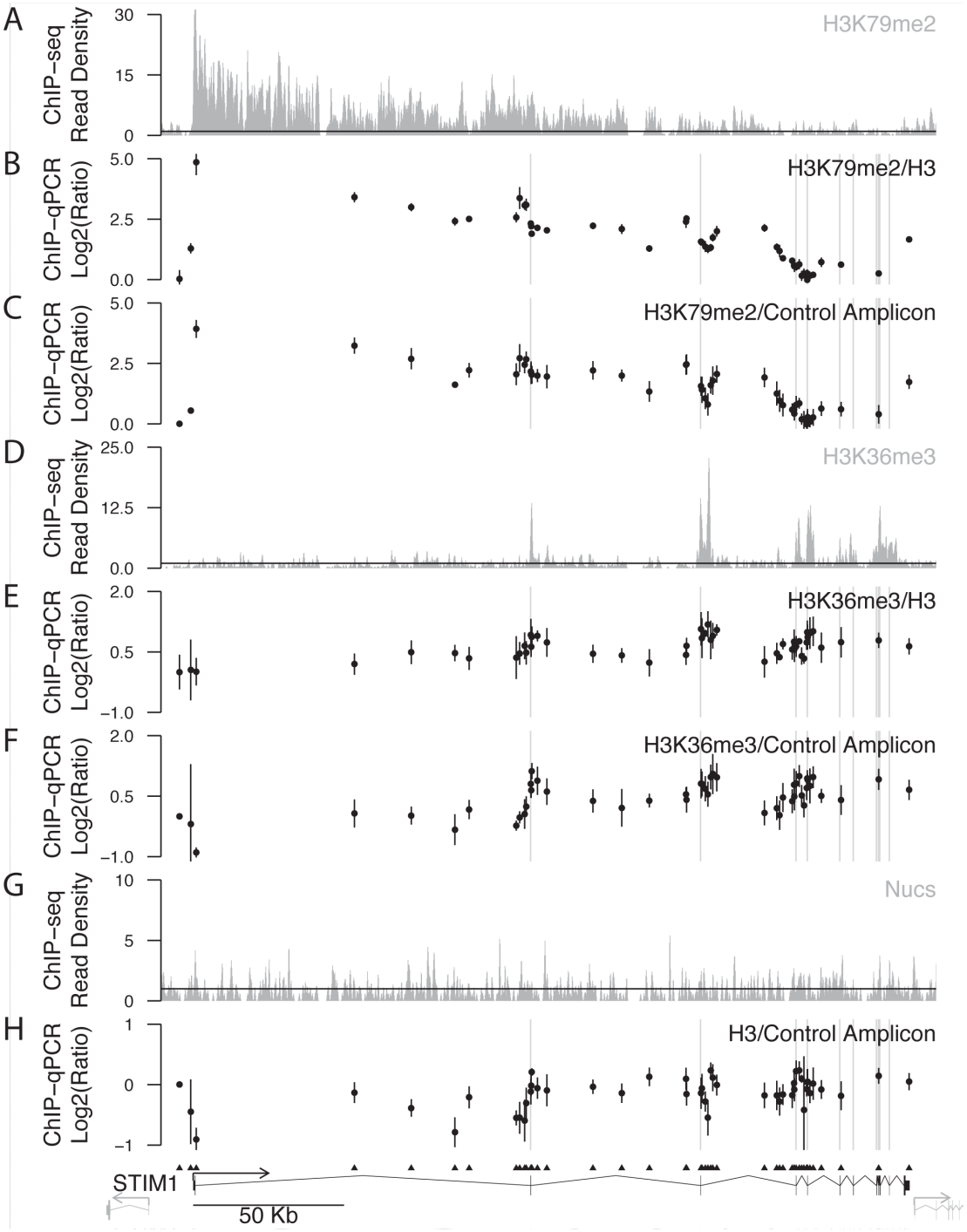


Figure 2.10 PCA of ChIP-seq data is modestly affected by nucleosome density

Modest effect of normalization to MNase-seq (Nucs) for PCA. (A-C) The loadings for principal components 1 (A), 2 (B) , and 3 (C) using ChIP-seq data without (x axis) or with (y axis) normalization to Nucs are shown. Points are colored as in Figs. 1 and S1 and the Pearson correlations (r) are shown at the top right of each comparision, demonstrating that similar components are described without and with normalization to nucleosomes. (D) A scree plot as in Figure 2.7 shows the variance accounted for by all 41 principal components with normalization to Nucs, increasing numerically from left-to-right, demonstrating that, similar to PCA without normalization, the first three components (black) explain most of the variance in the dataset: 59.1% total between the first three.

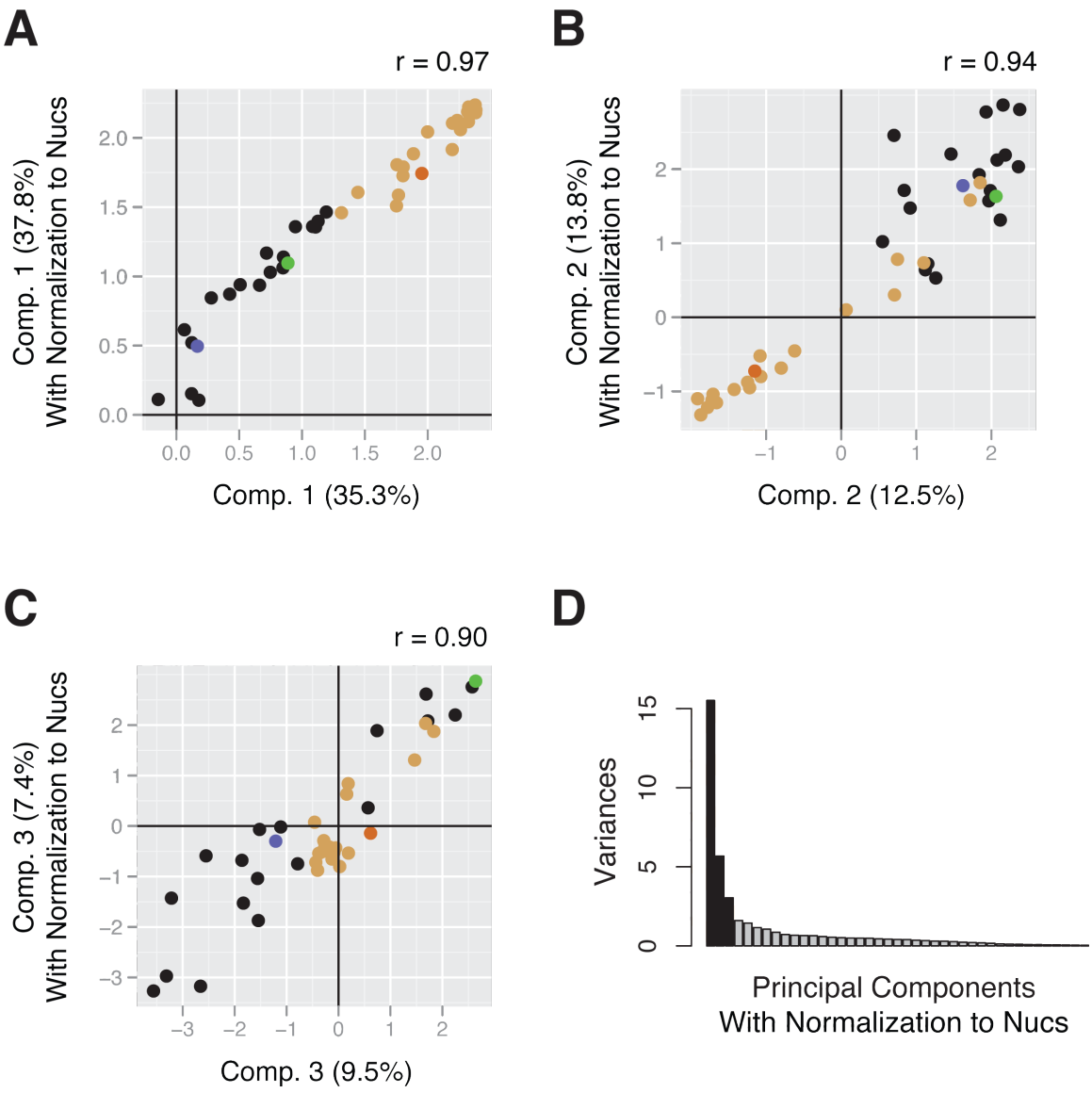

A
r = 0.97
Comp. 1 (37.8%)
With Normalization to Nucs
Comp. 1 (35.3%)
B
r = 0.94
Comp. 2 (13.8%)
With Normalization to Nucs
Comp. 2 (12.5%)
C
r = 0.90
Comp. 3 (7.4%)
With Normalization to Nucs
Comp. 3 (9.5%)
D
Variances
Principal Components
With Normalization to Nucs

Figure 2.11 Intronic and exonic marks do not merely reflect nucleosome density

Heatmaps of nucleosome-normalized marks. Heatmaps were generated just as in Figure 2.2, but each ChIP-seq dataset was first divided by MNase-seq data. Thus, each 1-kb bin now represents the ratio of a mark over nucleosomes. Values without MNase-seq reads in the bin (missing values) are colored white. Notably, the pattern is very similar to the heatmaps without normalization (Figure 2.2), though they are noisier.

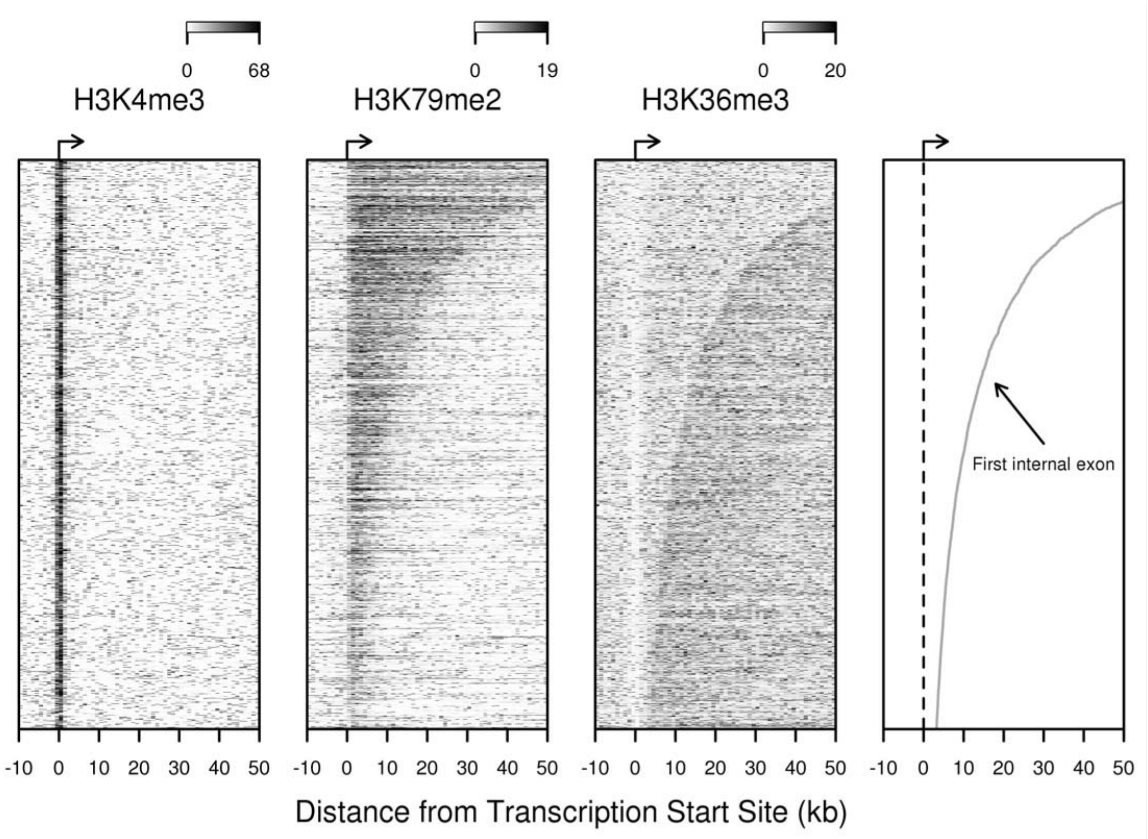

0 68
H3K4me3
0 19
H3K79me2
0 20
H3K36me3
First internal exon
-10 0 10 20 30 40 50
-10 0 10 20 30 40 50
-10 0 10 20 30 40 50
-10 0 10 20 30 40 50
Distance from Transcription Start Site (kb)

Figure 2.12 Histone modifications are more strongly influenced by* bone-fide *exons than exon-like composition regions (ECRs)

Comparison of histone modification spatial dynamics at exon-like composition regions exons. The relative density of ChIP-seq reads for each modification at ECRs and matched *bona fide* exons. The relative density of ChIP-seq reads for each modification were calculated surrounding 12,687 unique intronic ECRs (gray) and surrounding the nearest exon of each of those ECRs (black). Any ECR in which the nearest exon was the first exon (i.e. near the promoter) was omitted to remove promoter-proximal effects. Modification profiles were also normalized to the density of MNase-seq (Nucs) reads, thus controlling for enrichment of nucleosomes at both exons and ECRs. The average size of exons/ECRs is shown as a vertical gray bar. Shown are the intronic marks in the third principal component, exonic H3K36me3, and H3K27me1, which serves as a control (lower right).

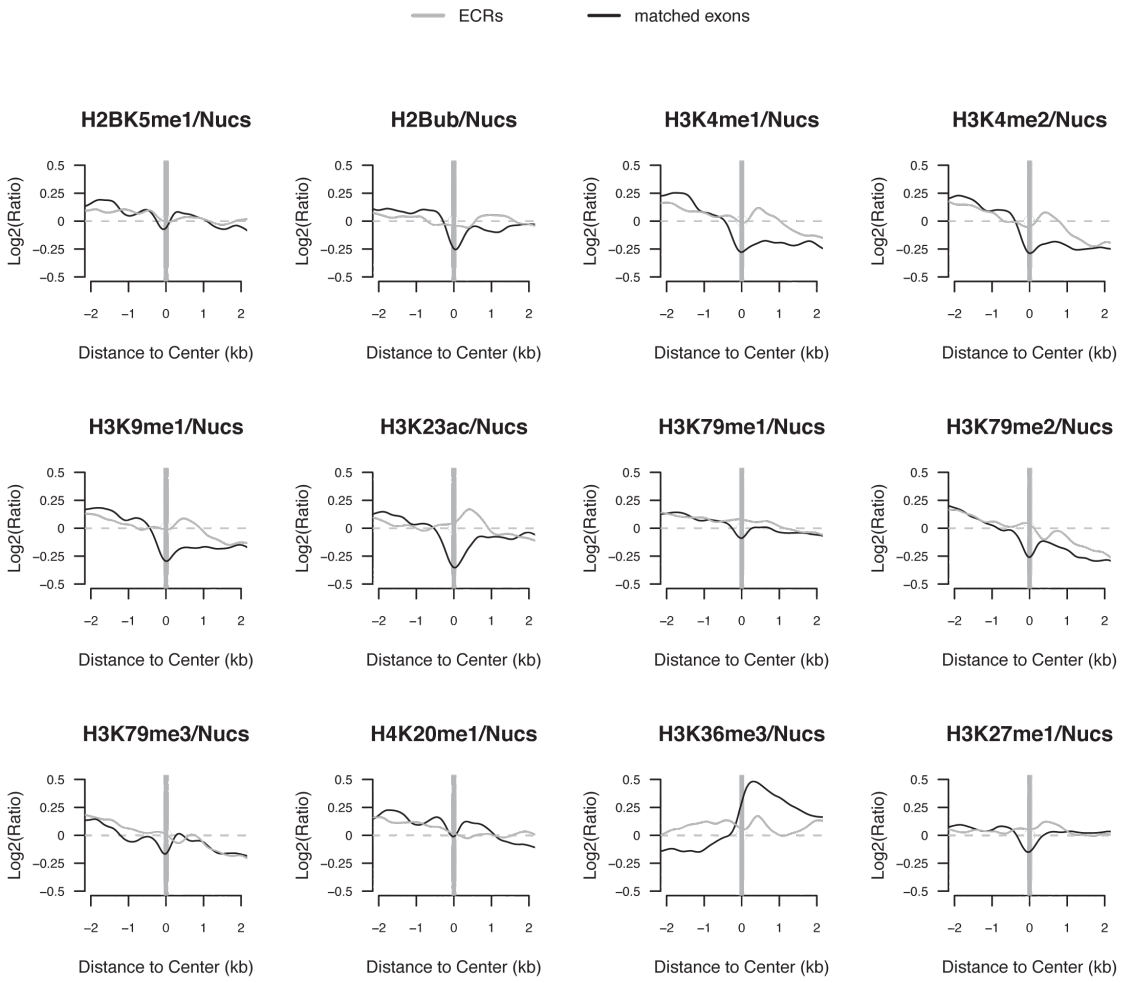

ECRs
matched exons
H2BK5me1/Nucs
H2Bub/Nucs
H3K4me1/Nucs
H3K4me2/Nucs
H3K9me1/Nucs
H3K23ac/Nucs
H3K79me1/Nucs
H3K79me2/Nucs
H3K79me3/Nucs
H4K20me1/Nucs
H3K36me3/Nucs
H3K27me1/Nucs
Log2(Ratio)
Distance to Center (kb)

Figure 2.13 YPEL5 RT-PCR and ChIP-qPCR profiles in response to caffeine induction

YPEL5 alternative splicing, expression, and corresponding chromatin marks during caffeine induction. (A-C) Characterization of YPEL5 exon 2 inclusion in response to caffeine. (A) YPEL5 exon 2 inclusion after 16 hours exposure to a range of caffeine concentrations at 0, 0.02, 0.2, 2, 4, 6, 8, 10, 12, 14, 16, and 18 mM. (B) Kinetics of YPEL5 exon 2 inclusion after 1-7 hours exposure to 18 mM caffeine. YPEL5 isoform ratios indicate the mean gel densitometry measurements of two biological replicates. (C) Mean log2 abundance +/- 1 SD of YPEL5/RPLP0 after 7 hours exposure to 0, 6, and 18 mM caffeine, calculated from qPCR measurements of (n=3) SW620 biological replicates. (D-F) Com- parison of ChIP- and MNase-seq with ChIP-qPCR data for histone modification and nucleosome occupancy profiles. Sequence read densities of published T cell data (top panels, gray) for (D) H3K79me2, (E) H3K36me3, and (F) nucleosomes in which uniform distribution of sequence reads would be equal to one (black line). Percent IP recovery profiles (bottom panels) of (D) H3K79me2, (E) H3K36me3, and (F) H3 from mock treated SW620 cells. Points indicate ChIP-qPCR measurements for each of two biological replicates. Vertical gray lines provide a visual reference for the 3' splice site of each internal exon. Arrowheads indicate the positions of qPCR amplicons along the gene (bottom). The alternatively spliced cassette exon 2 is indicated in red. The gene schematic and markers are drawn to scale according to their genomic coordinates.

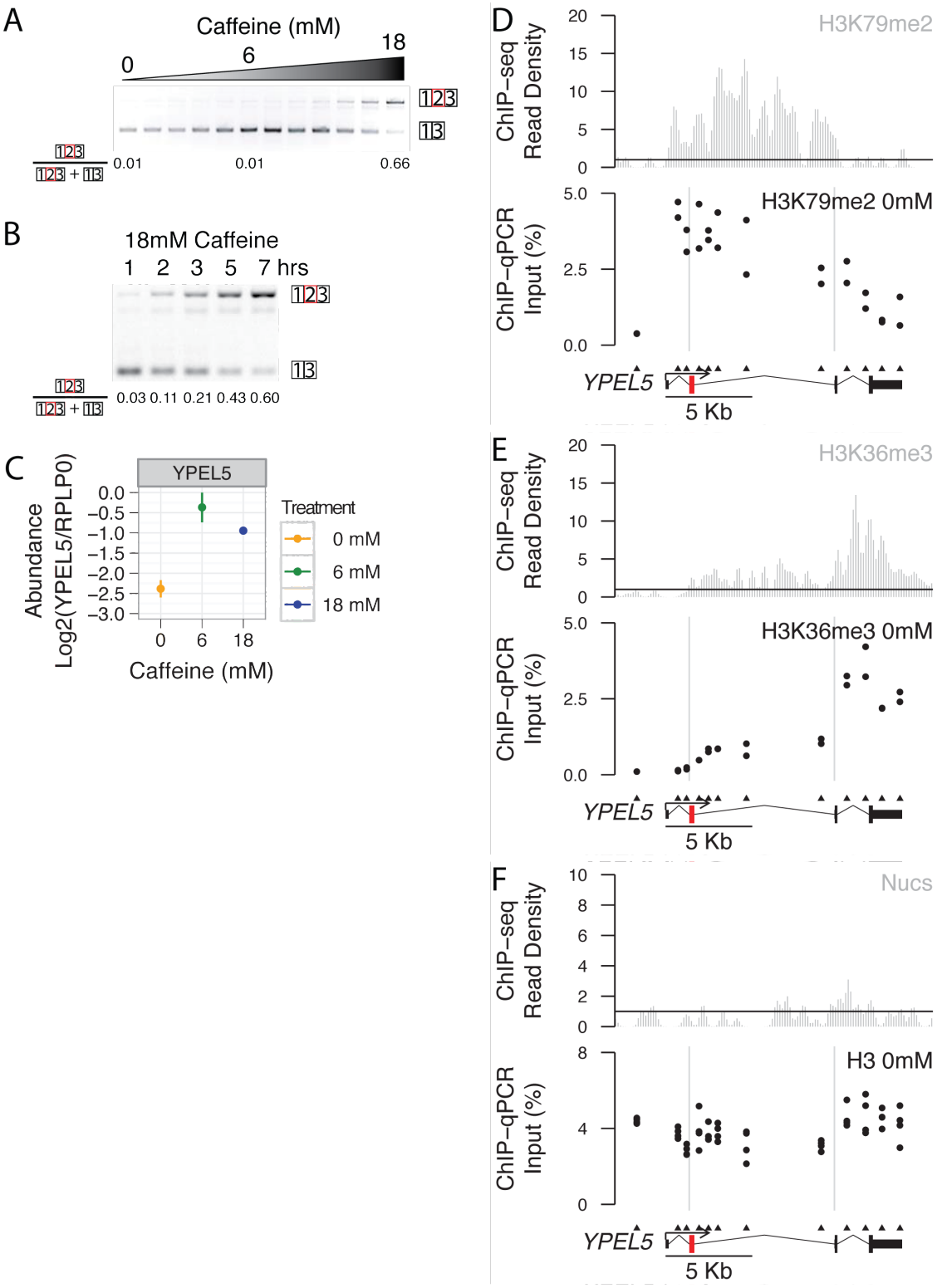

A
Caffeine (mM)
0
6
18
123
13
123 / 123 + 13
0.01
0.01
0.66
B
18mM Caffeine
1 2 3 5 7 hrs
123
13
123 / 123 + 13
0.03 0.11 0.21 0.43 0.60
C
YPEL5
Abundance
Log2(YPEL5/RPLP0)
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
0 6 18
Caffeine (mM)
Treatment
0 mM
6 mM
18 mM
D
H3K79me2
ChIP−seq Read Density
20 15 10 5 0
ChIP−qPCR Input (%)
H3K79me2 0mM
5.0 2.5 0.0
YPEL5
5 Kb
E
H3K36me3
ChIP−seq Read Density
20 15 10 5 0
ChIP−qPCR Input (%)
H3K36me3 0mM
5.0 2.5 0.0
YPEL5
5 Kb
F
Nucs
ChIP−seq Read Density
10 8 6 4 2 0
ChIP−qPCR Input (%)
H3 0mM
8 4 0
YPEL5
5 Kb

Figure 2.14 CD45 ChIP-qPCR profiles in response to changes in hnRNPLL expression

Comparison of ChIP- and MNase-seq with ChIP-qPCR data for histone modification and nucleosome occupancy profiles at the CD45 gene. In all panels colored points and error bars refer to matched BJAB B cell lines expressing a short hairpin targeting knockdown of either RFP as a control (blue) or hnRNPLL (green) or to matched BL41 B cell lines overexpressing hnRNPLL (orange) or mock-treated as a control (red). Sequence read densities of published T cell data (top panels, gray) for (A,D) H3K79me2, (B,E) H3K36me3, and (C,F) nucleosomes in which uniform distribution of sequence reads would be equal to one (black line). Points indicate mean percent IP recovery +/- 1 SD (bottom panels) of (A,D) H3K79me2, (B,E) H3K36me3, and (C,F) H3, calculated from ChIP-qPCR measurents of matched BJAB (n=5) and BL41 (n=6) biological replicates. The alternatively spliced exons are indicated in red and numbered. The gene schematic and markers are drawn to scale. Note that only a portion of the gene is shown.

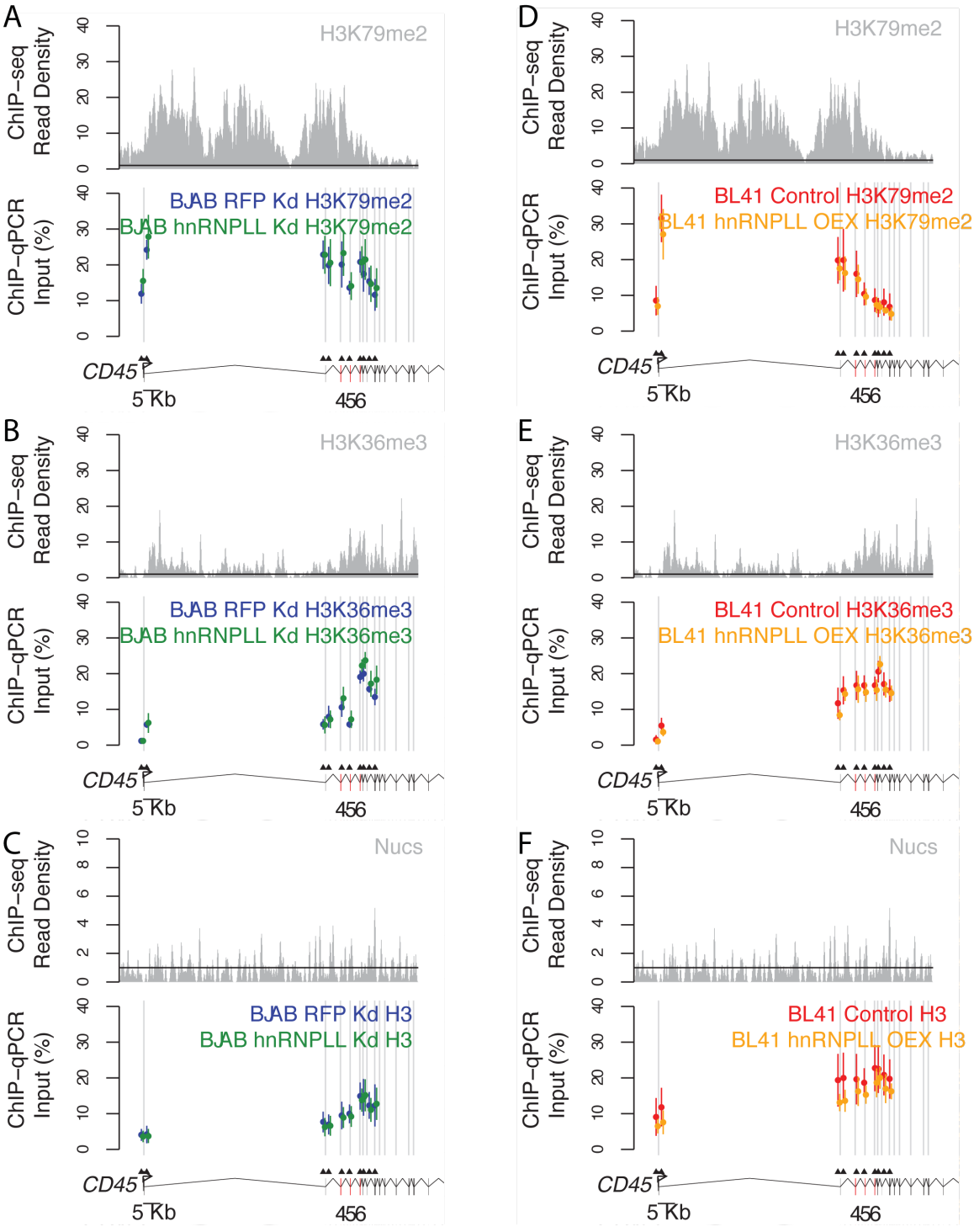

A
H3K79me2
ChIP–seq Read Density
ChIP–qPCR Input (%)
BJAB RFP Kd H3K79me2
BJAB hnRNPLL Kd H3K79me2
CD45
5 Kb
456
D
H3K79me2
BL41 Control H3K79me2
BL41 hnRNPLL OEX H3K79me2
B
H3K36me3
BJAB RFP Kd H3K36me3
BJAB hnRNPLL Kd H3K36me3
E
H3K36me3
BL41 Control H3K36me3
BL41 hnRNPLL OEX H3K36me3
C
Nucs
BJAB RFP Kd H3
BJAB hnRNPLL Kd H3
F
Nucs
BL41 Control H3
BL41 hnRNPLL OEX H3

Figure 2.15 Representative gene models illustrating that H3K36me3 marks are enriched near individual human exons at actively transcribed genes.

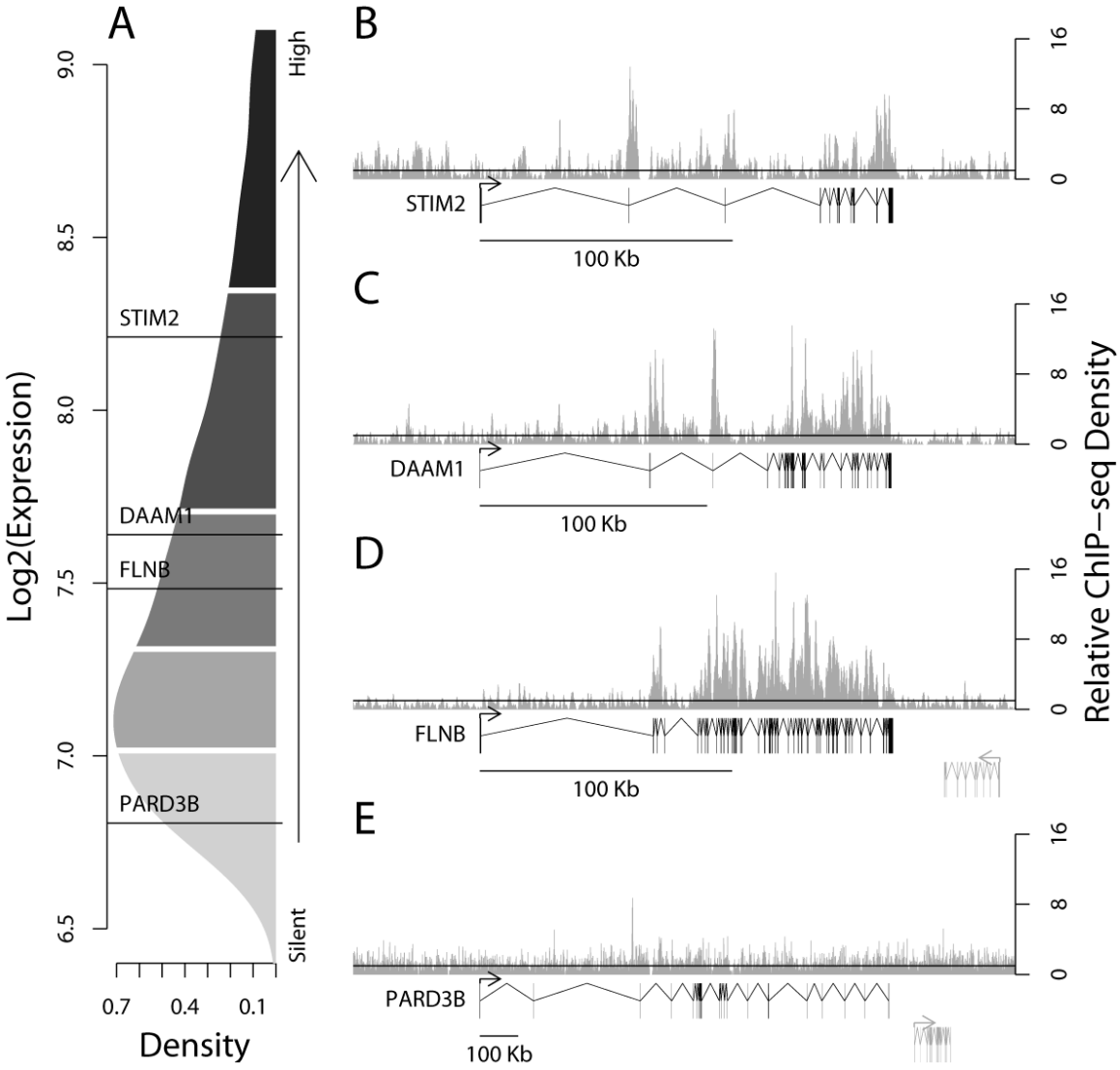


Table 2.1 Primers used in this study

STIM1_P1C12	chr11	4048240	4048304	4048272		64	GAGGGAGAAATGTGTAGCTG	GAAGAATGACCCCATAACAA	ChIP-qPCR
STIM1_P1A09	chr11	4060712	4060764	4060738		52	ACACACTCTTTGGCACCTT	TACATCATCCAGGGAAGAGC	ChIP-qPCR
STIM1_P1A11	chr11	4070792	4070847	4070819.5		55	TTTGCTATGGCTGTAAAGGT	ATAAAGGTATGGCAGGGAGT	ChIP-qPCR
CD45_P1C07	chr1	196875753	196875868	196875810.5		115	TCCTGGGTGGTGGGTAAATA	AGGCACTGGAAGTCCAACTA	ChIP-qPCR
CD45_P1C08	chr1	196929050	196929074	196929012		124	CCGTCCAAATGAAGAAAACAA	GCCATGCTAAACGCTGTAAA	ChIP-qPCR
CD45_P1C09	chr1	196932781	196932893	196932837		112	AAGTGCAGAAATTGCAGGAAA	TCCTTGGAGATCTTGCTTCC	ChIP-qPCR
CD45_P1C11	chr1	196939170	196939262	196939216		92	GGTGATGTGCTCTCACAAGG	TTCCATCACCCTCAGGATTC	ChIP-qPCR
CD45_P1C12	chr1	196939409	196939509	196939459		100	GGAAAAGAACAAGCCCATCA	TATCTCCACGCCCTCTCACT	ChIP-qPCR
CD45_P1D01	chr1	196940833	196940927	196940880		94	GGCTGGCTCAAAGCAGATAC	GAAGAACCGCTTTGCTTCAA	ChIP-qPCR
CD45_P1D02	chr1	196942759	196942844	196942801.5		85	TGTGGTGTTCTCCAGTGGTC	TCACCCACTTGTTTCCAAGG	ChIP-qPCR
CD45_P1D03	chr1	196874203	196874300	196874251.5		97	GTTTTTGGAAGCAGGGTTTG	GTTTTGCCCCAAATCTTTCA	ChIP-qPCR
CD45_P1D04	chr1	196927370	196927465	196927417.5		95	TCTCCTTTAGGCCTCCCTTT	GCTAGCTTGGTCCCAGAAGA	ChIP-qPCR
CD45_P1D06	chr1	196935056	196935152	196935104		96	CAATAAAGCCACCTGGAAACA	AGTGACAGGCACCTGAGTCC	ChIP-qPCR
CD45_P1D11	chr1	196938183	196938282	196938232.5		99	CCAGTTTCCCCATTGACAAC	CTGAGGTGTTCGCTGTGATG	ChIP-qPCR
CD45_P1D12	chr1	196942496	196942613	196942554.5		117	TGCAAACATCACTGTGGATT	ATGCACCTCATTGTTTGTGC	ChIP-qPCR
YPEL5_C	chr2	30221558	30221624	30221591		66	TCTGACATGAAAGGATGGCTA	TTCCTGTTTTTCAACCTGCTT	ChIP-qPCR
YPEL5_H	chr2	30223930	30223990	30223960		60	CGGAACTTGACCGAGTTGTT	CCTGCTCAGGACGGTTCAC	ChIP-qPCR
YPEL5_J	chr2	30224430	30224492	30224461		62	CTTGCAACCCCTTTCCTGTA	GCTGATCCGGACACTCATTT	ChIP-qPCR
YPEL5_O	chr2	30225147	30225221	30225184		74	TGAAAGGAAATAAGGCCAAAAA	TCGTGGACTGTTTTTCAGAGG	ChIP-qPCR
YPEL5_Q	chr2	30225693	30225755	30225724		62	CCTGATTCCACAACAGGACA	CCTCTTGCAGTACGTGACCA	ChIP-qPCR
YPEL5_R	chr2	30226242	30226304	30226273		62	CTGAGGCGATGGTTTCATTT	CTGACGGAAAAACACCCAGT	ChIP-qPCR
YPEL5_U	chr2	30227882	30227949	30227915.5		67	GCCCCTCACTGTTTCAGGTA	CCAAGCTTGACTGTTTAACTGG	ChIP-qPCR
YPEL5_AD	chr2	30232200	30232261	30232230.5		61	TTCTGTTCTCATTGCTTTAGCC	TGATCCTACACTGCCTCCAA	ChIP-qPCR
YPEL5_AH	chr2	30233685	30233750	30233717.5		65	TCTGTTGCCTCAGGAAATGTT	TCCTTCCATAGTCTGTGCTGAA	ChIP-qPCR
YPEL5_AL	chr2	30234764	30234823	30234793.5		59	ACGGTAGTGCATGATTGCAT	GCAGGCAGACATCTGGAAAT	ChIP-qPCR
YPEL5_AM	chr2	30235712	30235768	30235740		56	CCGGTGTGCCATATTCTTGT	AAGGAAGAATGGGCTCTGGT	ChIP-qPCR
YPEL5_AO	chr2	30236750	30236815	30236782.5		65	TCCTTTCTCCACCATGGACT	GCCAGTATGGAAATGAATGGA	ChIP-qPCR
CD45_alt.splicing	NA	NA	NA		256;397;595;739	GCAAAGCCCAACACCTTCCCCCACTG		ATGCACCTCATTGTTTGTGC	RT-PCR
CD45_total	NA	NA	NA			155	CTACTACTCCATCTAAGCCAACATGTG	ATGCACCTCATTGTTTGTGC	RT-qPCR
hnRNPLL	NA	NA	NA			117	CGGCCAACTCGTCTAAATGT	TTCTCCCAAAATGGCTTGTC	RT-qPCR
RPLP0	NA	NA	NA			130	GCTGCTGCCCGTGCTGGTG	TGGTGCCCCTGGAGATTTTAGTGG	RT-qPCR
YPEL5_total	NA	NA	NA			66	AGTTCAAGATCGGGTCATGC	TTGCAGTTTTTGCAGCTCAC	RT-qPCR
YPEL5_alt.splicing	NA	NA	NA	NA	180;477		CATACCAGCTGAAGAGCGACAAGC	AGTGGAGATGAGTTCTGAGCGGTT	RT-PCR

Chapter 3: Diverse forms of RPS9 splicing are part of an evolving autoregulatory circuit

3.1 Abstract

Ribosomal proteins are essential to life. While the functions of ribosomal protein-encoding genes (RPGs) are highly conserved, the evolution of their regulatory mechanisms is remarkably dynamic. In *Saccharomyces cerevisiae*, RPGs are often present as two intron-containing copies, which we show is likely the result of selection against intron loss from RPGs after whole-genome duplication. To investigate the role of introns in the regulation of RPG expression, we constructed 14 *S. cerevisiae* strains with precise deletions of RPG introns. We found that several yeast introns function to repress rather than to increase steady-state mRNA levels. Among these, *RPS9A* and *RPS9B* introns were required for cross-regulation of the two paralogous gene copies, which is consistent with the duplication of an autoregulatory circuit. To test for similar intron function in animals, we performed an experimental test and comparative analyses for autoregulation among distantly related animal *RpS9* orthologs. Overexpression of an exogenous *RpS9* copy in *Drosophila melanogaster* S2 cells induced alternative splicing and degradation of the endogenous copy by nonsense-mediated decay (NMD). Also, analysis of expressed sequence tag data from distantly related animals, including *Homo sapiens* and *Ciona intestinalis*, revealed diverse alternatively spliced RPS9 isoforms predicted to elicit NMD. We propose that multiple forms of splicing regulation among RPS9 orthologs from various eukaryotes operate analogously to translational repression of the alpha operon by S4, the distant prokaryotic ortholog. Thus, *RPS9* orthologs

appear to have independently evolved variations on a fundamental autoregulatory circuit.

3.2 Introduction

The evolution and function of spliceosomal introns are among the largest unsolved mysteries of eukaryotic genomes. Pronounced differences in intron evolution between lineages and between introns within the same lineage provide insight into 1) the selective and mutational forces governing intron evolution and 2) the potential roles of introns in gene function. Here we study the case of ribosomal protein genes (RPGs) in the model yeast *Saccharomyces cerevisiae*. RPGs are highly over-represented among intron-containing genes (69% of RPGs contain introns compared to ~5% of non-RPGs), which has been suggested to reflect ongoing selection for introns that provide one or more functions in gene expression (Bon et al., 2003; Quinn M Mitrovich, Tuch, Guthrie, & Johnson, 2007). However, two major facets of this hypothesis --- the action of selection (intron evolution), and the source of this selection (intron function) --- remain unknown. First, biased intron loss has not been specifically tested within hemiascomycetous yeasts (*S. cerevisiae* and relatives). And second, the effect of intron loss on RPG expression remains uncertain.

RPG expression is remarkable both in terms of synthesis rate and control (Warner, 1999); thus, RPG introns may function to promote these aspects of gene expression. One proposal predicts that RPG introns function to promote high levels of expression. Consistent with this view, intron-containing genes, including RPGs, produce some of the highest transcript and protein abundances in *S. cerevisiae* (Juneau, Miranda,

Hillenmeyer, Nislow, & Davis, 2006). However, the requirement for introns to enhance RPG expression has not been directly tested.

In addition to the above, two other proposals predict that RPG introns function by providing an opportunity for splicing regulation. One possibility is that introns provide rapid regulation in response to environmental stress, as suggested by splicing inhibition of RPG pre-mRNAs in response to amino acid starvation (Pleiss, Whitworth, Bergkessel, & Guthrie, 2007a). Another possibility is that introns provide an opportunity to fine-tune gene expression through autoregulation. For example, negative feedback control of *RPL30* and *RPS14B* expression is achieved through the binding of their respective protein products to RNA structures within their own unspliced transcripts, thereby regulating splicing (Dabeva & Warner, 1993; Fewell & Woolford, 1999). Interestingly, nearly all the ribosomal proteins of *Escherichia coli* are regulated by key ribosomal proteins in an analogous manner; for example, bacterial S4 directly binds its own RNA operon to repress the translation of itself and three other RPGs (D. Dean & Nomura, 1980; Nomura, Gourse, & Baughman, 1984). Given that the majority of *S. cerevisiae* RPGs contain introns, intron-dependent autoregulation may be more common than previously appreciated.

We report the first direct tests of both the action and source of selection on RPG introns. First, we used comparative genomics to show that RPG introns have been preferentially retained following whole genome duplication (WGD), indicating ongoing selection for retention of RPG introns. Second, we generated *S. cerevisiae* strains harboring precise deletions of 14 RPG introns to distinguish between selective hypotheses. We found that RPG introns generally reduce gene expression, suggesting that RPG introns allow for splicing regulation rather than promoting high levels of expression. In particular, we identified intron-dependent cross-regulation between the *RPS9A* and *RPS9B* genes, which both encode ribosomal protein S9 (S9). Finally, overexpression of *RpS9* in *D. melanogaster* S2 cells, and analysis of available EST sequences, suggest that autoregulation of *RpS9* orthologs may involve different forms of splicing regulation between species, but also appears to be widespread across disparate lineages.

3.3 Results

3.3.1 Yeast ribosomal protein genes have resisted recent intron loss

Introns are over-represented in the RPGs of both *C. albicans* and *S. cerevisiae* (Bon et al., 2003; Quinn M Mitrovich et al., 2007). While this shared over-representation may reflect selection pressure to maintain RPG introns prior to the divergence of these two species from a common ancestor, it may also reflect the action of selection in more recent history, since their divergence from a common ancestor. This distinction is important, since selection pressure to maintain RPG introns in more recent history is

more likely to be relevant to the biology of *S. cerevisiae*. To determine if RPGs resisted intron loss compared to other genes since the divergence of *C. albicans* and *S. cerevisiae* (~200 - 800 million years ago; Hedges, Dudley, & Kumar, 2006), we assessed the fates of *S. cerevisiae* introns in paralogs (a.k.a. gene pairs) that were duplicated ~ 100 million years ago by whole-genome duplication (WGD; Wolfe & Shields, 1997). To determine the fates of introns after genome duplication, we took advantage of the well-annotated genome of *S. cerevisiae*, which has been exhaustively searched for introns (Juneau, Palm, Miranda, & Davis, 2007; Z. Zhang, Hesselberth, & Fields, 2007). With these annotations, we identified 121 intron-containing genes among 554 WGD-derived gene pairs obtained from Yeast Gene Order Browser (Byrne & Wolfe, 2005). Assuming that intron loss has largely dominated intron evolution in hemiascomycetous yeast species (Stajich et al., 2007), we inferred intron loss if one of the WGD-derived gene copies had fewer introns than the other. Using this criterion, we calculated the number of apparent intron losses in RPG pairs compared to all other gene pairs. Strikingly, this simple accounting revealed that 16 of 23 non-RPG pairs have a gene with fewer introns than its copy, whereas none of the 46 RPG pairs did (Figure 3.6A). Nonetheless, this analysis ignores intron losses that occured independently in both gene copies and assumes that intron gain did not occur.

To better assess whether WGD-derived RPG pairs have been biased for either intron gain or loss (including losses in both gene copies), we reconstructed the hypothetical intron distribution of the pre-WGD ancestor that existed prior to the WGD event. For each of the 554 *S. cerevisiae* duplicated gene pairs, we assigned the presence or

absence of an intron in the hypothetical pre-WGD ancestral ortholog based on introns annotations and predictions from the genomes of the pre-WGD (so-called protoploid) species (*C. albicans, Lachancea waltii*, *L. thermotolerans*, *L. kluyveri*, *Eremothecium gossypii, Kluyveromyces lactis,* and *Zygosaccharomyces rouxii)* and the genomes of the post-WGD species (*Vanderwaltozyma polyspora, Naumovia castellii*, *C. glabrata*, and *S. bayanus*)**.** A complete list of intron predictions and annotations can be found in Table 3.1. Our analysis revealed 73 intron-containing genes that were likely present in the pre-WGD ancestor from which the duplicated gene pairs in *S. cerevisieae* were descended (Figure 3.1A). Based on this hypothetical intron distribution of the pre-WGD ancestor, we inferred the number of *S. cerevisiae* WGD-derived gene pairs that gained or lost an intron for each post-WGD gene pair (Figure 3.1B). From this improved analysis, we identified 5 *S. cerevisiae* non-RPG pairs that appear to have independently lost introns from both gene copies after gene duplication. This was in addition to 14 non-RPG pairs in which one of two introns were lost (Figure 3.1B, right and middle columns, respectively). Once again, we inferred no intron losses in *S. cerevisiae* RPG pairs (Figure 3.1B, left column). Thus, RPG introns appear to have been biased against loss in the lineage leading to *S. cerevisiae* during the last ~100 million years.

Next, we asked whether intron gains contributed to the bias for introns in *S. cerevisiae* RPGs. For a given *S. cerevisiae* gene, we inferred that an intron was gained if introns were absent in both the pre-WGD ancestor and the majority of post-WGD orthologous gene pairs. Using this criterion, we did not infer intron gains in any of the *S. cerevisiae* RPGs. On the other hand, two introns in non-RPGs (i.e. *USV1* and *BMH2*) have possibly

been gained in the *S. cerevisiae* lineage (Table 2.1); however, both of these introns are located in the 5' UTR, which are not well annotated in other species, and it is therefore difficult to be confident. Taken together, the bias for introns in *S. cerevisiae* RPG pairs appears to have been dominated not by intron gains in RPGs, but by intron losses in non-RPGs.

3.3.2 Introns repress ribosomal protein gene expression

Having found a bias against RPG intron loss, we sought to determine if RPG introns have a function in gene expression. To mimic the effect of RPG intron loss, we created *S. cerevisiae* mutant intron deletion strains (henceforth denoted as Δi). Each Δi mutant was created with a precise deletion of a single RPG intron, such that only an intronless copy of the gene remained at the endogenous locus (Figure 3.7).

Because RPGs are among the most highly expressed genes in the genome, we tested the model that introns are required in *cis* for high levels of gene expression by assessing the expression profiles of 14 Δi mutants compared to a wild-type strain. We also considered the possibility that intron deletions may affect other genes, in particular, the WGD-derived gene copies of RPG pairs. To measure changes in expression of the gene from which an intron was deleted (in addition to 124 RPG and 911 non-RPG features) we used custom splicing-sensitive microarrays designed to detect pre-, mature, and total mRNA species (Pleiss, Whitworth, Bergkessel, & Guthrie, 2007b). Intron deletions, on average, increased the expression of the mutated gene (up to 3.7-fold of wt), whereas only the intron deletion of *RPS8A* resulted in decreased expression (0.87-fold of wt). The gene

most increased by intron deletion was *RPS14B*, which is known to require an RNA stem-loop within its intron to repress its own expression (Fewell & Woolford, 1999). These results suggest that yeast introns are generally not required to increase RPG expression.

Our microarray experiments also revealed instances in which intron deletions affected both the expression of the mutated gene and its WGD-derived gene copy. In the case of the *rps9bΔi* mutant, intron deletion was associated with both a relative increase (1.4-fold wt) in *rps9bΔi* (Figure 3.9; red point) expression as well as a relative decrease (0.35-fold of wt) in *RPS9A* expression (Figure 3.9; blue point). Conversely, in the *rps9aΔi* mutant, *RPS9A* expression was increased (2.4-fold of wt; red point) while *RPS9B* expression was modestly decreased (0.9-fold of wt; Figure 3.9; blue point). Thus, *RPS9A* and *RPS9B* genes require introns to repress their own expression, and further, derepression of one gene copy results in repression of the other gene copy.

We hypothesized that both *RPS9A* and *RPS9B* genes cross-regulate by a common intron-dependent mechanism, because these genes are highly similar to each other. To determine if repression of *RPS9A* expression in the *rps9bΔi* mutant required the *RPS9A* intron (and vice versa), we created a double *rps9a/bΔi* mutant and tested the effect on mRNA abundance by RT-qPCR (Figure 3.2B). In addition to validating the microarray results for single intron deletions, the *rps9a/bΔi* double mutant showed increased abundance of both the mutated *RPS9A* and *RPS9B* genes compared to a wild-type strain, which were similar in magnitude to the single mutants (Figure 3.2B). RT-qPCR also confirmed that the relative effects of intron deletion were different between the two

gene copies as observed by microarray (Figure 3.2B and 3.9). These results suggest that *RPS9A* and *RPS9B* introns are required for cross-regulation by repressing gene expression, and also, that these genes are repressed in different relative amounts.

3.3.3 Drosophila RpS9 *autoregulates through alternative splicing and NMD*

Reminiscent of the cross-regulation between *S. cerevisiae RPS9A* and *RPS9B* genes, several metazoan RPGs have been shown to autoregulate through alternative splicing coupled to NMD (so-called "Regulated Unproductive Splicing and Translation" or RUST): a process in which the synthesis of productively spliced mRNA is repressed in favor of unproductive mRNA isoforms encoding premature termination codons (PTC+; Q M Mitrovich & Anderson, 2000; Cuccurese, Russo, Russo, & Pietropaolo, 2005; Malygin, Parakhnevitch, Ivanov, Eperon, & Karpova, 2007) reviewed in (Lareau, Brooks, et al., 2007). While this process is conserved between distantly related eukaryotes, there is no known overlap between the genes regulated by RUST in yeast and metazoans to facilitate mechanistic comparisons. Intriguingly, an alternatively spliced *RpS9* PTC+ mRNA isoform was recently identified in *Drosophila melanogaster* (Hansen et al., 2009). Thus, we considered the possibility that other *RPS9* orthologs autoregulate in a manner analogous to *RPS9A* and *RPS9B* cross-regulation.

We hypothesized that *D. melanogaster RpS9* expression is regulated in response to excess protein production by alternative splicing coupled to NMD. Therefore, we predicted that increased *RpS9* expression would result in increased abundance of the PTC+ mRNA isoform. To test this hypothesis, we measured the effect of exogenous *RpS9*

overexpression and NMD inhibition (Figure 3.3A) on alternative splicing of *RpS9* messages using RT-qPCR primer sets specific to endogenous *RpS9* mRNA isoforms (Figure 3.3B). We first verified that the previously identified *RpS9* PTC+ isoform in S2 cells was degraded by NMD through RT-PCR amplification of *RpS9* transcripts from S2 cells incubated with either of two dsRNAs targeting *Upf1* (Figure 3.10); RT-qPCR measurements using primer pairs specific to the *RpS9* PTC+ message indicated a > 8-fold increase after *Upf1* knock-down relative to a non-targeting control knock-down (Figure 3.3C, compare top left and top right panels). To then test the effect of increased *RpS9* expression on the abundance of the PTC+ mRNA isoform, we exogenously overexpressed a cDNA copy of *RpS9*. In S2 cells overexpressing *RpS9*, we detected an increase in the abundance of the PTC-containing mRNA isoform (Figure 3.3B, top panels, compare red and blue points) and a decrease in the total *RpS9* expression as compared to the empty vector control (Figure 3.3B, bottom left panel, compare red and blue points). As expected, we observed a UPF1-dependent decrease in total endogenous *RpS9* abundance in response to increased *RpS9* expression (Figure 3.3B, compare bottom left and right panels, blue points). Taken together, these results suggests that Drosophila *RpS9* autoregulates by RUST, in which excess expression shifts the balance of alternative splicing from the synthesis of productively spliced messages towards the synthesis of unproductive *RpS9* PTC+ messages that are selectively degraded by NMD.

3.3.4 Diverse forms of RPS9 alternative splicing are associated with structured and conserved RNA sequences

We hypothesized that *RpS9* autoregulation had an important function, and therefore, would be conserved in other animals. Further, we hypothesized that conserved RNA structures were involved in the cross-regulation of *RPS9A* and *RPS9B* in *S. cerevisiae* and the autoregulation of *RpS9* in *D. melanogaster*, because *E. coli* S4 (the bacterial ortholog), requires an RNA structure to autoregulate by translational repression. Therefore, we predicted that *RpS9* orthologs would be associated with alternatively spliced mRNA isoforms, conserved RNA structures, and PTCs. To identify such messages, we summarized expressed sequence tags (ESTs) data from diverse animals. Indeed, EST coverage extends outside exons and into introns, which support the existence of rare unspliced or alternatively spliced transcripts (< 5% maximum coverage; Figure 3.4, gray bars). To identify ESTs that specifically support alternative splice site usage or cassette exon inclusion, we mapped putative EST exon-exon junctions that spanned both 5' GT and 3' AG splice sites (Figure 3.4, blue and red bars, respectively). With the exception of *Petromyzon marinus*, ESTs from various vertebrates (e.g. *Homo sapiens*, *Rattus norvegicus*, *Xenopus tropicalis*, *Danio rerio*, and *Oryzias latipes*) reveal cassette exons that introduce PTCs from the last canonical intron (Figure 3.4 and Figure 3.10). *P. marinus* and *D. melanogaster* ESTs, on the other hand, reveal alternative 5’ splice sites that also introduce PTCs from a homologous intron (Figure 3.4 and Figure 3.10). Most intriguingly, *Ciona intestinalis* ESTs also support alternative 5’ splice site usage, but in a non-homologous intron compared to those of other animals (Figure 3.4). Thus, our surveys of animal ESTs suggest that animal *RpS9* orthologs are often alternatively spliced to utilize RUST. Further, the conservation of alternatively spliced cassette exons within

the last intron among distantly related vertebrates (e.g. ~400 million years between humans and fish; Hedges et al., 2006), suggest that these isoforms are functional.

Also consistent with function, PTC positions in *RpS9* orthologs were associated with high nucleotide conservation (Figure 3.4). To determine if *RpS9* orthologs were also associated with thermodynamically stable and structurally conserved RNA structures, we screened the gene bodies of *RpS9* orthologs for statistically significant RNA structures using RNAz (Washietl, Hofacker, & Stadler, 2005) on alignments obtained from the UCSC genome browser (Kent et al., 2002). In order to examine both intronic and exonic sequences, we obtained sets of nucleotide alignments from closely related groups of organisms: mammals, drosophilids, teleosts, and hemiascomycetous yeasts. Scanning *RpS9* alignments in 400 bp windows, we identified predicted RNA structures ($P > 0.9$), specifically within the last intron of mammalian, drosophilid, and teleost *RpS9* orthologs, each overlapping with PTC positions (Figure 3.4, green lines and red octagons). Similarly, sequence alignments of *RpS9* orthologs from hemiascomycetous yeasts also revealed predicted RNA structures specifically within the single yeast intron, which if unspliced, would introduce a PTC (Figure 3.4). Due to the lack of *RPS9* sequences similar to *C. intestinalis*, we could not test for conserved elements and predicted RNA structures corresponding to the PTC in its third intron. In any case, these data indicate the potential for autoregulation among distantly related *RpS9* orthologs through the use of different forms of alternative splicing, perhaps through structured RNA elements.

3.4 Discussion

Intron evolution is complex and mysterious. Different lineages have experienced fundamentally different histories of intron gain and loss, leading to dramatic changes in intron numbers that span orders of magnitude. Selection has likely played a role in shaping intron evolution and may act differently depending on the organism and the gene that harbors a given intron (Jeffares, Mourier, & Penny, 2006); however, the functions of introns are still debated and remain difficult to predict.

The complex evolutionary history of introns immediately raises three fundamental questions. First, why do introns persist? If conservation reflects function, then the maintenance of large numbers of introns in many lineages suggests widespread intron function. On the other hand, introns may largely persist merely because mutations that allow precise removal are rare. Therefore, the inference of intron function from intron evolution requires careful consideration of both mutational and selective forces. Second, what functions of introns promote their selection and persistence? The most widely recognized function of introns is their role in increasing protein diversity through alternative splicing; however, such intron function appears to be largely restricted to animals, and thus, cannot explain intron persistence in many other lineages. Third, are intron functions general across species, or have they acquired quite different functions in different organisms (e.g., the lineage-specific functions of alternative splicing in animals)? Our study sheds light onto these questions, which we discuss in detail below.

Here, we used a combination of phylogenetic analysis and directed mutagenesis to identify functional introns in *S. cerevisiae* with relevance to regulated alternative splicing in animals. We detailed the rates of intron loss among hemiascomycous yeast genes after whole-genome duplication (WGD) to show that RPGs have resisted intron loss compared to other genes within recent evolutionary time, possibly to maintain regulatory function (Figure 3.1). To determine the effect that RPG intron loss would have on gene expression, we constructed precise intron deletions in *S. cerevisiae*, which revealed gene-specific repression of RPG expression (Figure 3.2A). In the case of the duplicated *RPS9A* and *RPS9B* genes, introns were required for cross-regulation between paralogs (Figure 3.2B). Because *E. coli* S4, the bacterial ortholog of eukaryotic *RPS9*, is also known to regulate its own expression, we tested for autoregulation in other distantly related organisms. Indeed, we found that *D. melanogaster RpS9* overexpression resulted in the accumulation of unproductively spliced messages that were degraded by NMD (Figure 3). Lastly, comparative genomic analysis of animal *RpS9* orthologs, revealed diverse alternative splicing patterns, including alternative 5' splice sites and cassette exon inclusion, predicted to elicit NMD (Figure 3.4). Taken together, our experimental and computational approaches revealed surprisingly diverse forms of intron-dependent regulation among *RpS9* orthologs, which appear to be elaborations on a fundamental autoregulatory circuit first identified in *E. coli*.

3.4.1 Biased intron loss may reflect selection for functional introns

The genes of *S. cerevisiae*, and hemiascomycetous yeasts in general, contain very few introns compared to other eukaryotes (Jeffares et al., 2006), which is generally

attributed to uncommonly high rates of intron loss within this lineage (Stajich et al., 2007). Previous observations that *S. cerevisiae* introns are biased for RPGs (Bon et al., 2003; Quinn M Mitrovich et al., 2007; Ares, Grate, & Pauling, 1999; Spingola, Grate, Haussler, & Ares, 1999) and other highly expressed genes (Juneau et al., 2006) have been cited as evidence that many *S. cerevisiae* RPG introns have one or more functions. Intriguingly, similar biases are also observed in the intron-poor genomes of *Encephalitozoon cuniculi* (Katinka et al., 2001; R. C. H. Lee, Gill, Roy, & Fast, 2010) and the nucleomorph of *Guillardia theta* (Douglas et al., 2001), suggesting that the bias against RPG intron loss is not limited to yeasts. By measuring the rates of intron loss among recently duplicated genes, we confirm that an ongoing bias against RPG intron loss is apparent in the lineage leading to *S. cerevisiae* (Figure 3.1). Thus, the few remaining introns in *S. cerevisiae* may reflect biases in 1) the mechanisms of intron loss and/or 2) selection to keep important introns. In addition to previously proposed functions of RPG introns (see below), several lines of evidence suggest that the conservation of RPG introns is not merely a function of mutation rates. Reverse transcription-mediated intron loss is expected to preferentially remove 3' end biased introns from highly expressed genes (G R Fink, 1987). First, intron biases for RPGs run counter to the expectation for intron loss among highly expressed genes, since these transcripts would be more likely to be reverse transcribed (as discussed in R. C. H. Lee et al., 2010). Second, the majority of *S. cerevisiae* intron losses observed here are not 3' end biased; in fact, several introns were lost from the 5' UTR (e.g. *GBP1*, *NHP6A* and *ARF1*). Lastly, at least 21 RPG introns that are present in both the *Lachantea* and

Saccharomyces clades appear to have been lost from *Z. rouxii*, indicating that species-specific RPG intron losses can occur, but have not in the lineage leading to *S. cerevisiae* (Figure 3.6). Biased intron loss, therefore, may reflect species-specific selective pressure to retain functional introns.

3.4.2 Intron function in the absence of alternative spicing

The simple gene architectures of *S. cerevisiae* provide limited opportunity for the generation of multiple protein isoforms through alternative splicing (although a few instances have been described; Grund et al., 2008; Juneau, Nislow, & Davis, 2009). *In lieu* of alternative splicing, *S. cerevisiae* RPG introns have been proposed to confer other functions, such as transcriptional enhancement (Juneau et al., 2006) and splicing regulation in response to stress (Pleiss, Whitworth, Bergkessel, & Guthrie, 2007b) or to maintain homeostasis (Dabeva & Warner, 1993; Fewell & Woolford, 1999). Unlike intronless copies of some mammalian genes (Brinster, Allen, Behringer, Gelinas, & Palmiter, 1988), RPG intron deletion decreased the expression of only one of 14 genes we assessed by microarray (Figure 3.2A). Rather, intron deletions resulted in increased expression from *RPS14B* and both gene copies encoding ribosomal protein S9. Similar in concept to the classically described autoregulation of *RPL30* and *RPS14B* genes (Dabeva & Warner, 1993; Fewell & Woolford, 1999), we suggest that the *RPS9A* and *RPS9B* gene copies cross-regulate in an intron-dependent manner to regulate total ribosomal protein S9 expression. While this is the first demonstrated case of intron-dependent cross-regulation in *S. cerevisiae*, it is possible that additional examples or auto- or cross-regulation remain undiscovered given the large number of intron-containing RPG

transcripts. Like many *S. cerevisiae* intron deletion mutants, *RPS9A* and *RPS9B* intron deletions are viable and lack substantial growth defects (Parenteau et al., 2008 and data not shown); however, these introns (Figure 3.4) and others (data not shown) harbor conserved nucleotides in regions outside of splice sites, strongly implying that mutations within these introns have been detrimental to fitness during recent natural history. Therefore, experimental tests for a variety of functions beyond growth may be required to reveal non-essential, yet important, intron functions.

Both Rpl30p and Rps14bp directly bind their own transcripts through RNA structures present in their introns (Dabeva & Warner, 1993; Fewell & Woolford, 1999). The aforementioned conserved elements within the introns of *RPS9A* and *RPS9B* are composed of two remarkably long stretches of 25- and 31-bp of identical sequence (Vincent & Liebman, 1992). While gene conversion between gene copies might facilitate these similarities, the pre-WGD orthologs from *C. albicans* and other protoploid yeasts harbor conserved nearly identical sequences, which suggests the action of negative (purifying) selection (Figure 3.8). RNAz predictions at these conserved regions suggest they form stable RNA structures (Figure 3.4). Intuitively, such RNA structures would be likely targets for ribosomal protein S9 binding (Vincent & Liebman, 1992). Nonetheless, *in vitro* binding assays are needed to determine if ribosomal protein S9 directly regulates its own expression in *S. cerevisiae* and other eukaryotes. If auto- and cross-regulation were indeed mediated by ribosomal protein S9 binding, then comparative

biochemical studies using proteins and RNA sequences from different species could provide mechanistic detail to describe how the different forms of regulation have evolved (see below).

3.4.3 Evolution of the RPS9 autoregulatory circuit

In *S. cerevisiae*, we found that the *RPS9A* and *RPS9B* genes (which encode nearly identical S9 proteins; 99% identity) reciprocally cross-regulate (Figure 3.2B). Given that intron conservation extends to protoploid yeasts (Figure 3.8), we speculate that this form of regulation likely evolved from an ancient autoregulatory circuit and continues to fine-tune S9 levels produced from the two gene copies. The propensity for RNA-binding proteins to utilize autoregulation at the level of splicing has long been noted (Mattox, Ryner, & Baker, 1992), and in the case of RNA-binding proteins, remarkably common (Lareau, Inada, Green, Wengrod, & Brenner, 2007; Ni et al., 2007; Saltzman et al., 2008). Presumably, autoregulation of S9 production would benefit the cell by reducing waste (Warner, 1999) and by preventing potentially harmful interactions with low-affinity targets (Von Hippel et al., 1982).

Coss-regulation, such as between the *RPS9A* and *RPS9B* genes, has also been observed between multiple sets of paralogous splicing regulators, including hnRNPL and hnRNPLL (Rossbach et al., 2009), as well as PTB, nPTB and ROD1 (Spellman, Llorian, & Smith, 2007). We speculate that these genes exemplify a straightforward principle of gene duplication and evolution: upon gene duplication, autoregulation would inherently become reciprocal cross-regulation, but could evolve asymmetric cross-regulation as the

paralogs diverge in abundance and/or protein function (Figure 3.5A). In the case of *RPS9A* and *RPS9B*, we observed that the deletion of the *RPS9B* intron resulted in a dramatic decrease in relative *RPS9A* expression, while deletion of the *RPS9A* intron resulted in a modest decrease in relative *RPS9B* expression. These data can be explained by a type of feedback inhibition in which the messages of the two gene copies are synthesized in unequal amounts, but repressed in equimolar amounts. Indeed, *RPS9B* is expressed at substantially higher levels than *RPS9A* (Pnueli & Arava, 2007), which is consistent with our observation that *RPS9A* is more greatly affected by intron-deletion than is *RPS9B*. Thus, *RPS9B* appears to effectively regulate *RPS9A* expression. In theory, such asymmetric cross-regulation may allow differential expression of functionally distinct ribosomal proteins to produce a "ribosome code" (Komili, Farny, Roth, & Silver, 2007). What distinct functions, if any, *RPS9A* and *RPS9B* genes provide remains to be seen. In any case, functional assays are needed to determine if *S. cerevisiae* utilizes differential expression of *RPS9A* and *RPS9B* genes to exploit functional differences in the proteins they encode.

Why are there so many forms of regulation among *RpS9* orthologs? One possibility is that particular aspects of these forms are ancient and conserved, while others have evolved independently in different lineages. For example, the genetic circuits that specify the development of diverse animal forms (e.g. eyes and limbs) exemplify deep homology, where recent evolutionary innovations overlay a shared "genetic toolkit" (Shubin, Tabin, & Carroll, 2009). For example, the compound eyes of arthropods and camera eyes of vertebrates are two forms that were once thought to have each evolved

independently; however, it is now recognized that they rely on key genetic circuits and transcription factors such as PAX6, which underlie eye formation in diverse animals. By analogy, genetic circuits themselves (in this case, autoregulation) may share a common "biochemical toolkit" comprised of highly conserved biochemical processes (e.g. RNA:protein interactions), while independently evolving elaborations on these basic circuits. Thus, translational inhibition of the alpha-operon by S4-binding may represent but one of many possible forms of regulation accessible to the highly conserved S4 RNA-binding domain proteins found throughout cellular life. Assuming that animal and yeast messages are directly regulated by ribosomal protein S9 binding (see above), then alternative splicing in animals and regulated splicing in *S. cerevisiae* are different elaborations on this autoregulatory circuit, perhaps mediated by changes in RNA structure within introns (Figure 3.5B). In any case, the different forms of alternative splicing within homologous *D. melanogaster* and metazoan introns demonstrate that splicing patterns can diverge yet maintain the potential for autoregulation (Figure 3.4). Moreover, alternative splicing within non-homologous introns of *RpS9* orthologs (compare *S. cerevisiae* and *C. intestinalis* to metazoans) are reminiscent of the independently evolved PTCs of SR protein paralogs (Lareau, Inada, et al., 2007). Thus, we propose that the highly conserved function of ribosomal protein S9 (and RNA-binding proteins in general) is one part of a biochemical toolkit that is frequently used and reused, as the fundamental autoregulatory circuit is maintained, elaborated and reinvented.

3.5 Methods

3.5.1 Intron gain and loss analysis

To estimate the propensity for intron loss among RPGs and non-RPGs, we compared annotated *S. cerevisiae* intron-containing genes and WGD-derived gene pairs. *S. cerevisiae* intron annotations were obtained from the *Saccharomyces* Genome Database (http://www.yeastgenome.org/) on 7/20/2011. WGD-derived gene pairs (as inferred from genomic synteny; a.k.a. "Ohnologs") were obtained from the Yeast Gene Order Browser (http://wolfe.gen.tcd.ie/ygob/; Byrne & Wolfe, 2005). Because introns are commonly identified by gaps in BLAST-based homology searches, intron-containing genes with short first exons are commonly may not be accurately annotated. To identify annotated introns upstream of an annotated gene, custom [R] scripts were used to scan 800 bp upstream and 100 bp downstream of the ORF start site with a regular expression that recognizes >90% of *S. cerevisiae* introns by indentifying the most common splice sites to minimize false positive matches. The regular expression matches sequences meeting the following criteria in order: 1) any one of the 4 most common of 5' SSs, 2) an S1 length of at least 30, 3) any one of the 5 most common branchsites, and 4) an S2 length between 1 and 50, and 5) any one of the 3 most common 3' SS trinucleotides, which was formalized as:

"(gtatgt|gtacgt|gtaagt|gtatga).{30,}?(tactaac|gactaac|aactaac|tgctaac|cactaac).{1,50}?[tca]ag". The pre-WGD ancestor was inferred to contain an intron if the majority of available outgroup pre-WGD species orthologs (*C. albicans, L. waltii, L. thermotolerans, L. kluyveri, E. gossypii, K. lactis*) and 1) the *Z. rouxii* ortholog had an intron or 2) the majority of post-WGD species intron (*S. cerevisiae, S. bayanus, C. glabrata, N. castellii* and *V. polyspora*) gene pairs had at least one intron. In this manner we distinguished

independent intron gains and losses in *Z. rouxii* from intron gains and losses immediately after the WGD event.

3.5.2 S. cerevisiae strain construction

Intron deletions were generated by the method depicted in Figure 3.7. Briefly, a PCR product amplified from the plasmid pJPS1232 (generously provided by J. Staley, University of Chicago), which contains the CORE construct (Storici, Durham, Gordenin, & Resnick, 2003) fused to the I-SceI endonuclease site, using gene-specific primers containing exon 1 and exon 2 sequences that allow integration and subsequent intron deletion via homologous recombination. Transformed diploid cells (yAP047) were incubated for 4 hrs at 30°C in the presence of 2% galactose to induce I-SceI endonuclease expression and precise deletion of the CORE cassette. Sporulated haploid cells were confirmed to harbor intron deletions by PCR. To ensure that a precise intron deletion was obtained without any additional mutations, the region surrounding the newly created exon-exon junction (at least 100 bp) was PCR amplified and sequenced. Gene specific primers used for mutagenesis and genotype confirmation are detailed in Table 3.2.

3.5.3 D. melanogaster *S2 cell RNAi and transfection*

Routine passaging of S2 cell cultures and RNAi depletion was performed as described (S. L. Rogers & Rogers, 2008) with the following modifications. Briefly, 1 ug/ml of dsRNA was incubated with 3.5E5 cells in 350 ul of media in 24-well plates. After 48 hrs incubation with dsRNA, cells were transfected with 0.2 ug plasmid with Effectene (Quigen) according to manufactures instructions. Cells were harvested with 1 mL TRIzol

(Invitrogen) for analysis by RT-qPCR (see below). Primers used to generate PCR products used for dsRNA synthesis (Promega RiboMAX) are described in Table 3.2.

3.5.4 RNA isolation and cDNA synthesis

To analyze the expression of genes in *S. cerevisiae* intron deletion strains, 15 ml cultures of mutant and wild-type yeast were grown in parallel at 30°C in rich medium supplemented with 2% glucose to an optical density between A_{600} = 0.5 and 0.7. For microarray hybridization, RNA was isolated by acid-phenol extraction and converted to cDNA as described (Pleiss, Whitworth, Bergkessel, & Guthrie, 2007b). A similar protocol was performed for qPCR applications with the following modifications. After RNA isolation, 2 µg of DNase-treated RNA was random primed in a 40 µl reaction containing 1 µg dN9 primer, 50 mM TrisHCl (pH 8.4), 75 mM KCl, 3 mM MgCl2, 10 mM DTT, 0.5 mM dNTPs, and 5 ng murine Moloney leukemia virus (M-MLV) RT. Primers were hybridized at 60°C for 7 min prior to the addition of enzyme, and then incubated with enzyme at 42°C for at least 2 h. Prepared cDNA was diluted at least 10-fold before use in qPCR. Similarly, to analyze D. melanogaster S2 cell, 725 ul cultures of S2 cells (UCSF cell culture facility) were grown in 24-well plates at 25°C in Schneider's Drosophila Medium (Gibco) supplemented with 10% fetal bovine serum (UCSF) to a count of ~5E6 cells. RNA was extracted with 1 ml TriZol (Invitrogen) according to manufactures instructions. After RNA isolation, cDNA was prepared with SuperScript III (Invitogen) and random priming according to manufactures instructions. Prepared cDNA was diluted at least 10-fold before use in qPCR.

3.5.5 Microarray analysis

Splicing-sensitive microarrays were constructed and performed as described (Pleiss, Whitworth, Bergkessel, & Guthrie, 2007b). In each experiment, a wild-type strain derived from the same parent as the intron deletion mutant strain was used as a reference. Data was analyzed using the [R] Bioconductor package marray() (Gentleman et al., 2004).

3.5.6 Quantitative PCR

Quantitative PCR primers (Table 3.2) were designed with Primer3 (Rozen & Skaletsky, 2000) and *S. cerevisiae* or *D. melanogaster* genomic sequence obtained from the UCSC genome browser (SacCer1 or dm3, respectively; Cherry et al., 1997; Adams et al., 2000). Serial dilutions of DNA ranging from 100 to 0.16 ng of the genomic DNA were used to obtain calibration curves, measure primer efficiencies, and ensure that quantification was in a linear dynamic range. Primer sets yielding multiple amplification products or calibration curves with R-squared values of < 0.96 were excluded. For each qPCR sample, diluted cDNA was amplified in 25 µl volume reactions containing 250 µM dNTPs, 1X $(NH_4)_2SO_4$ buffer (Fermentas), 0.5 µM primer, 1.5 mM MgCl2, 1.25 Units Dynazyme II (Finnzymes), and Sybr Green I fluorescent dye (Sigma). Fluorescence was measured on a BioRad Opticon machine using standard cycling conditions (3 min at 95°C, 40 cycles of 15 s at 95°C, 30 s at 55°C, and 15 s at 72°C). Biological replicate qPCR values were determined as the median of technical replicates. For each of 3 biological replicates, target gene values (e.g. RPS9) were divided by reference gene values (e.g. SCR1) before

log transformation. Plots were generated using the [R] package ggplot2() (Wickham, 2009).

3.5.7 Assessment of alternative splicing among animal RPS9 orthologs

To assess EST coverage and splicing, we obtained genomic coordinates corresponding to GenBank ESTs from the UCSC genome browser 'spliced EST' track, which span at least one canonical intron of at least 32 bases (Benson, Karsch-Mizrachi, Lipman, Ostell, & Sayers, 2011). Custom [R] scripts were used to calculate EST coverage per genomic nucleotide position, and identify all exon-exon junctions that span putative GT/AG splice site (2011). The following genome assemblies were used in the analysis: *Branchiostoma floridae* (braFlo1); *Ciona intestinalis* (Ci2); *Danio rerio* (danRer7); *Drosophila melanogaster* (dm3), *Oryzias latipes* (oryLat2); *Petromyzon marinus* (petMar1); *Xenopus tropicalis* (xenTro2); *Rattus norvegicus* (rn4); *Mus musculus* (mm9); *Homo sapiens* (hg19).

3.6 Acknowledgements

We thank Jon Staley the very generous gift of *S. cerevisiae* strains, procedures, and advice required for the generation of intron deletion strains. We thank Jeff Farrell and Hansong Ma for their excellent advice concerning S2 cell culture and dsRNA knockdowns. We also thank Jason Huff, Imke Listerman, Scott Roy, and the Guthrie group for their helpful discussions in the preparation of this manuscript and throughout the course of this work.

3.7 Figures and Tables

Figure 3.1 Biased intron loss in hemiascomycetous yeasts after the recent whole-genome duplication event

A) The dendogram (top) illustrates the assumed topology of the phylogenetic relationships used to infer intron-containing genes present in the pre-WGD ancestor prior to the WGD event (based on [55]; not to scale); the estimated time of the WGD event is indicated (gray circle and scale bar) [11]. A heatmap (bottom) images the number of introns in *S. cerevisiae* gene pairs and their orthologs (rows) by species (columns). Genes containing an intron in pre-WGD species (brown tiles) were used to infer the intron-containing genes present in the pre-WGD ancestor (see Methods). Among the 95 *S. cerevisiae* gene pairs derived from an intron-containing gene in the pre-WGD ancestor, those with an intron in both gene copies (dark blue-green) were inferred to have no intron losses. *S. cerevisiae* gene pairs with an intron in only one of two gene copies (light blue-green) or no intron in either gene copy (white) were inferred to have had one or two intron losses, respectively. Missing genes (red) are indicated and were not included in the analysis. RPGs (green) and other functional gene classes (purple) are indicated (right-most column). See Figure 3.6C and Table 3.1 for intron predictions and annotations. *Ca = C. albicans, Lw = L. waltii, Lt = L. thermotolerans, Lk = L. kluyveri, Eg = E. gossypii, Kl = K. lactis, Zr = Z. rouxii, Sc = S. cerevisiae*.

B) A histogram counts the number of inferred intron losses for each *S. cerevisiae* gene pair that descended from an intron-containing pre-WGD ortholog. Intron losses from RPGs (green) are compared to other functional gene classes (purple). Asterisks indicates statistical significance values $p < 0.01$ (**), and 0.001 (***); exact binomial test.

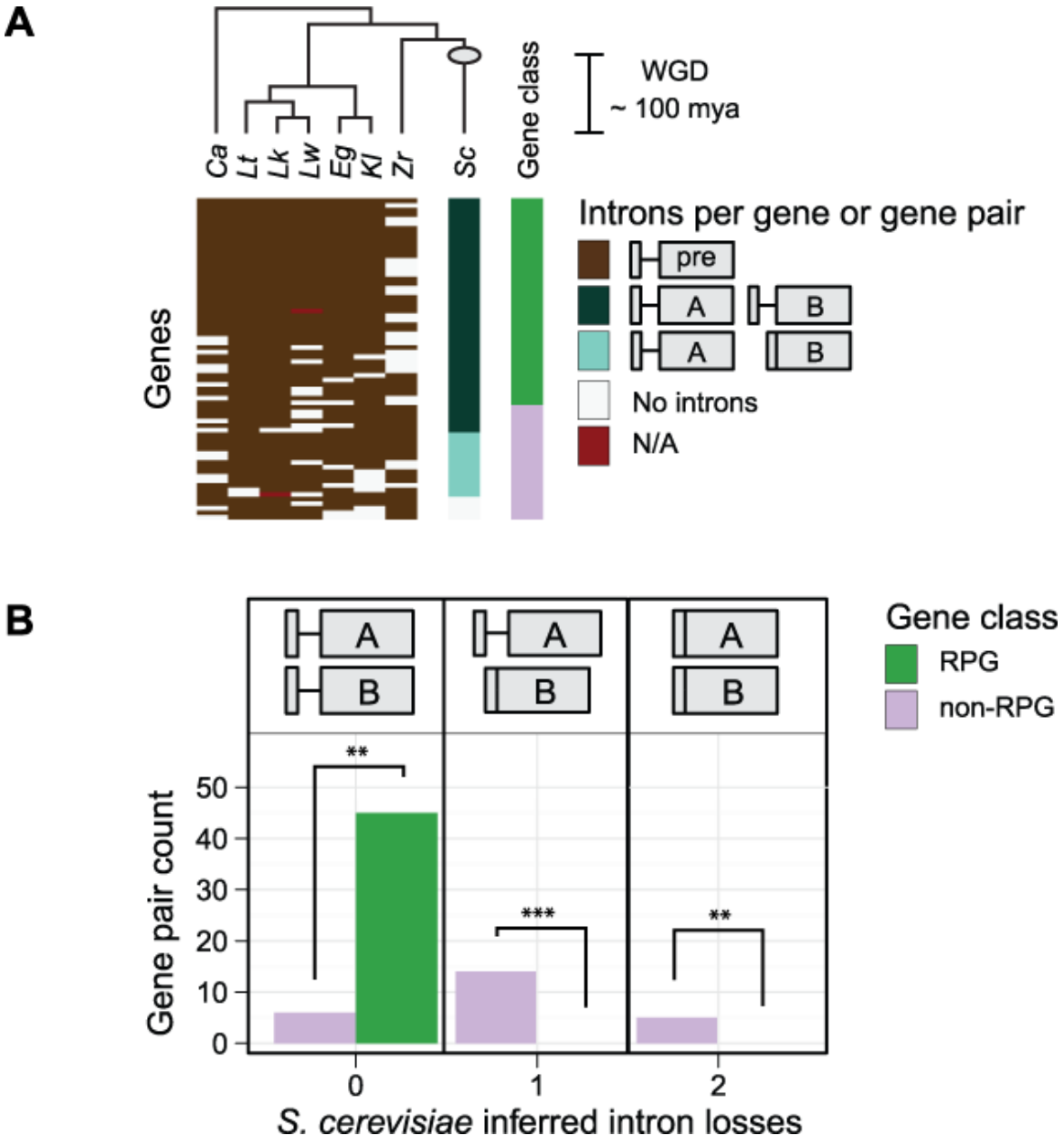

A
Ca
Lt
Lk
Lw
Eg
Kl
Zr
Sc
Gene class
WGD
~ 100 mya
Genes
Introns per gene or gene pair
pre
A
B
A
B
No introns
N/A
B
A
B
A
B
A
B
Gene class
RPG
non-RPG
**
***
**
Gene pair count
50
40
30
20
10
0
0
1
2
S. cerevisiae inferred intron losses

Figure 3.2 RPG intron deletions cause gene-specific increases in steady-state expression

A) Boxplots summarizing splicing-sensitive microarray expression data of 14 RPG intron deletion strains (x-axis) compared to a common wild-type strain. For each comparison, a red dot indicates the exon feature for the gene in which an intron has been deleted. Microarray data is expressed as normalized $\log_2$ transformed probe intensity from exon, intron, and junction features. Whiskers represent 1.5 times the interquartile range. For clarity, outliers are not shown (see Figure 3.9).

B) RT-qPCR determination of *RPS9A* (red circles) and *RPS9B* (blue triangles) expression changes between wt and intron deletion strains (panels). *RPS9A* and *RPS9B* values were divided by SCR1 values to obtain ratios controlled for variations in cDNA quantity. Log_2 transformed ratios are plotted relative to wt (based on the mean of three biological replicates). Each of three biological replicates is shown as a point and the mean as a dash.

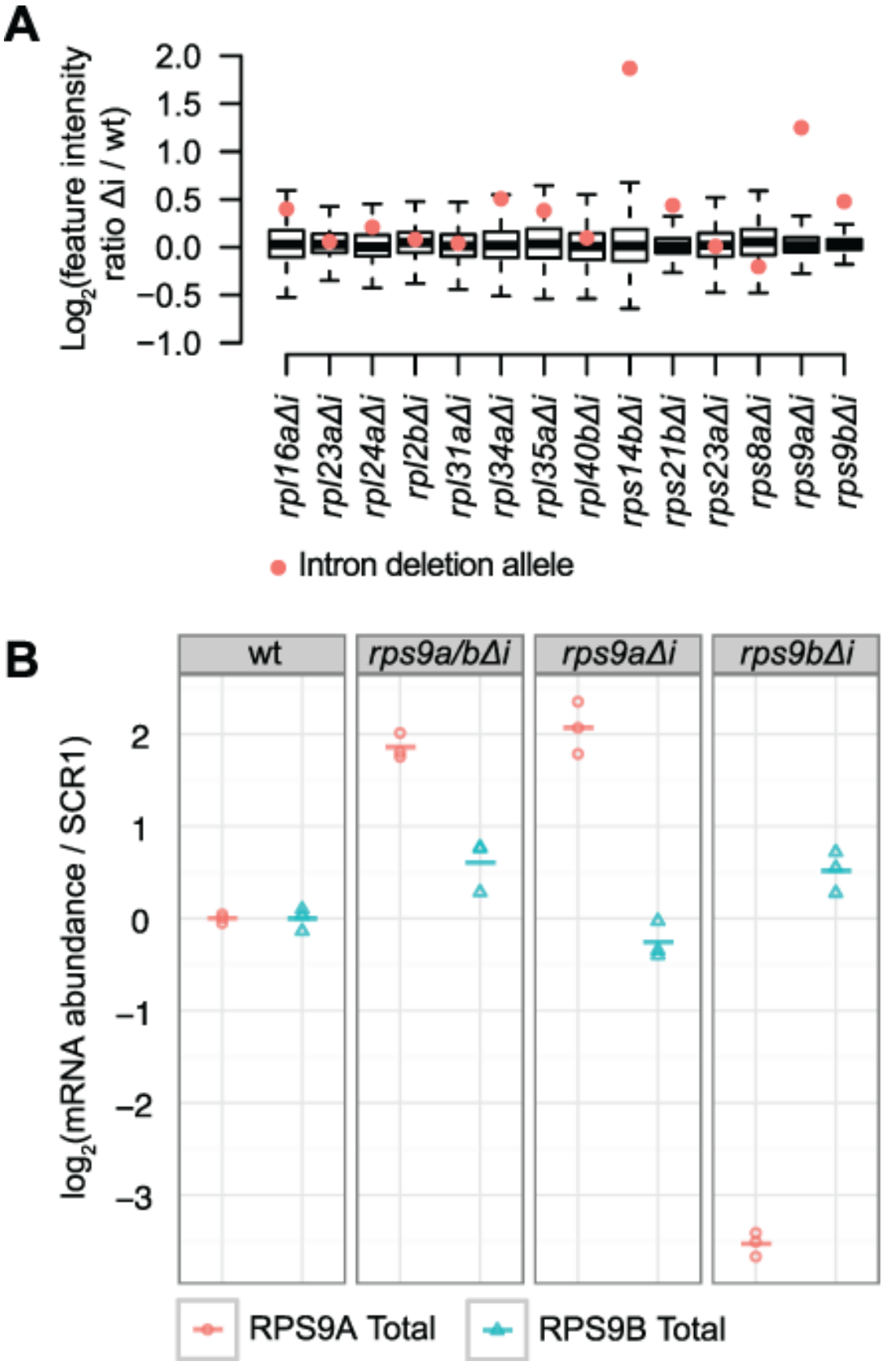

A
Log2(feature intensity ratio Δi / wt)
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
rpl16aΔi
rpl23aΔi
rpl24aΔi
rpl2bΔi
rpl31aΔi
rpl34aΔi
rpl35aΔi
rpl40bΔi
rps14bΔi
rps21bΔi
rps23aΔi
rps8aΔi
rps9aΔi
rps9bΔi
Intron deletion allele
B
wt
rps9a/bΔi
rps9aΔi
rps9bΔi
log2(mRNA abundance / SCR1)
2
1
0
−1
−2
−3
RPS9A Total
RPS9B Total

Figure 3.3 D. melanogaster RPS9 is autoregulated by alternative splicing coupled to NMD

A) Experimental scheme used to test for *RpS9* autoregulation and the role of NMD in the expression of alternatively spliced *RpS9* mRNA isoforms.

B) Schematic of *RpS9* mRNA isoforms. RT-qPCR primers (arrows) were designed to specifically amplify total endogenous *RpS9* expression or the PTC+ mRNA isoform, but not the overexpressed *RpS9* cDNA.

C) RT-qPCR determination of *RpS9* PTC+ mRNA isoform abundance (top panels) and total endogenous *RPS9* mRNA abundance (bottom panels) in S2 cells transfected with a plasmid constitutively expressing an *RPS9* cDNA (red circles) or an empty vector control (blue circles). The affect of *UPF1* knock-down (via incubation with dsRNA) on each *RPS9* mRNA isoform (right panels) is compared to a non-specific dsRNA control (left panels). *RPS9* mRNA isoform abundance values were divided by *GAPDH1* mRNA abundance values to obtain ratios internally controlled for variations in cDNA quantity. Log_2 transformed ratios for each of three biological replicates is shown as a point and the mean as a dash.

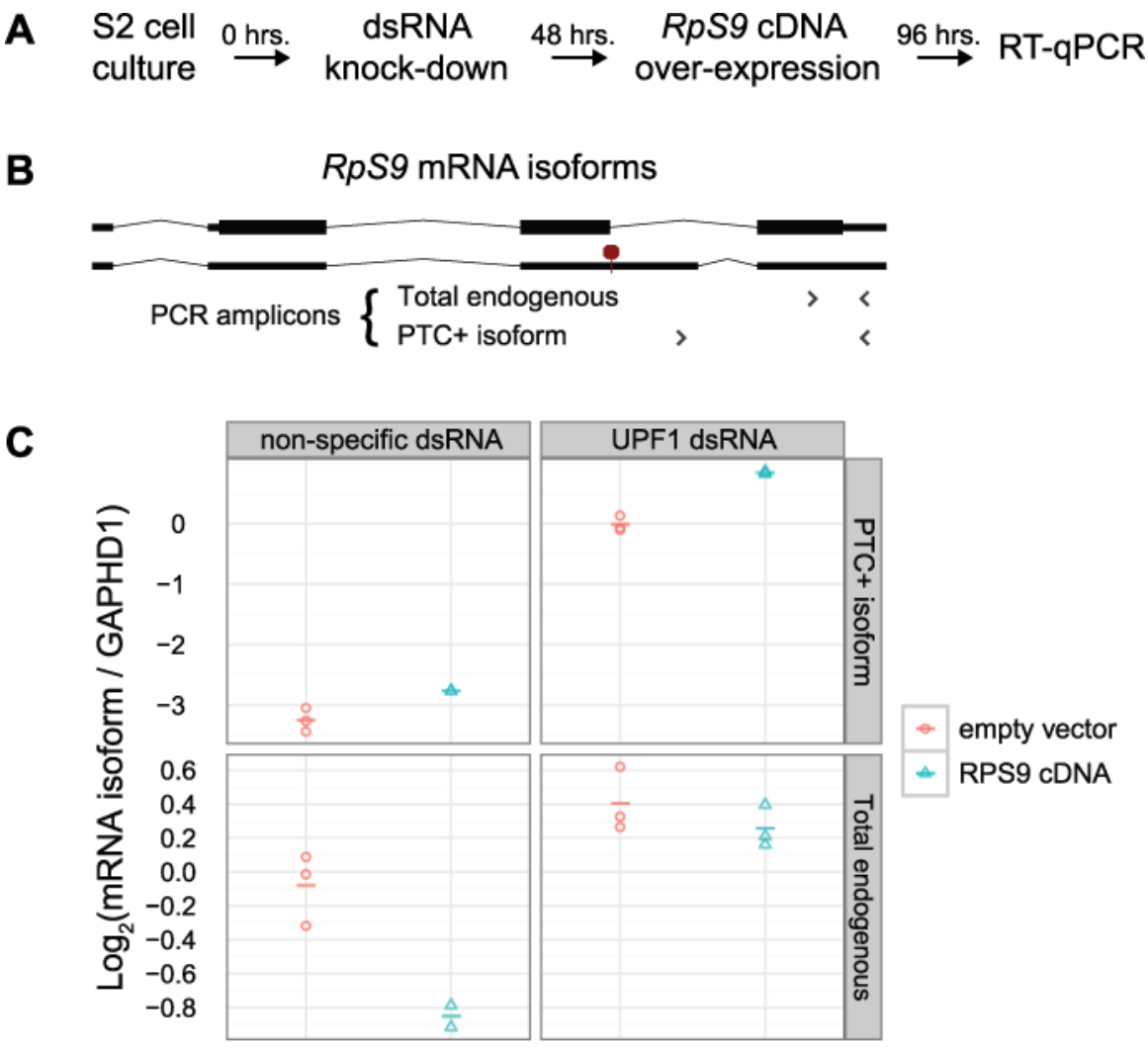

A
S2 cell culture
0 hrs.
dsRNA knock-down
48 hrs.
RpS9 cDNA over-expression
96 hrs.
RT-qPCR
B
RpS9 mRNA isoforms
PCR amplicons
Total endogenous
PTC+ isoform
C
non-specific dsRNA
UPF1 dsRNA
PTC+ isoform
Total endogenous
Log2(mRNA isoform / GAPHD1)
0
−1
−2
−3
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
empty vector
RPS9 cDNA

Figure 3.4 Diverse alternatively spliced RPS9 isoforms encode PTC+ exons associated with high nucleotide conservation and predicted RNA structures

Summaries of ESTs, predicted RNA structures, and sequence conservation from animal *RPS9* orthologs (*H. sapiens*, *X. tropicalis*, *O. latipes*, *D. melanogaster*, *C. intestinalis*, and *S. cerevisiae*) are presented along a dendogram illustrating their phylogenetic relationships (not to scale). For each species, histograms summarize EST coverage (gray bars) and inferred splice junctions with both 5' GT (blue bars) and 3' AG splice sites (red bars). Dashed lines separate the lower 5% and upper 95% histogram values; EST coverage is labeled on the y-axis. Two gene models (below each histogram) are plotted to scale (black line; 1 kb) representing either the major isoform (top gene model) or a spliced PTC+ EST (bottom gene model) for each species (an "unspliced" pre-mRNA is modeled for *S. cerevisiae* in lieu of an EST). The major isoform sequence is annotated as coding (thick black lines) or UTR (thin black lines) and interrupted by GT-AG introns (angled black lines). The first PTC (red line and octagon) in the representative PTC+ EST (thin lines) is indicated. Below the two gene models, PhastCons scores (black bars), and sequence windows identified by RNAz with statistically significant (P > 0.9) RNA structure predictions (green lines), respectively (*X. tropicalis* not shown; *C. intestinalis* not applicable). PhastCons scores and RNAz predictions were based on MultiZ alignments obtained from the UCSC genome browser where available (see Methods).

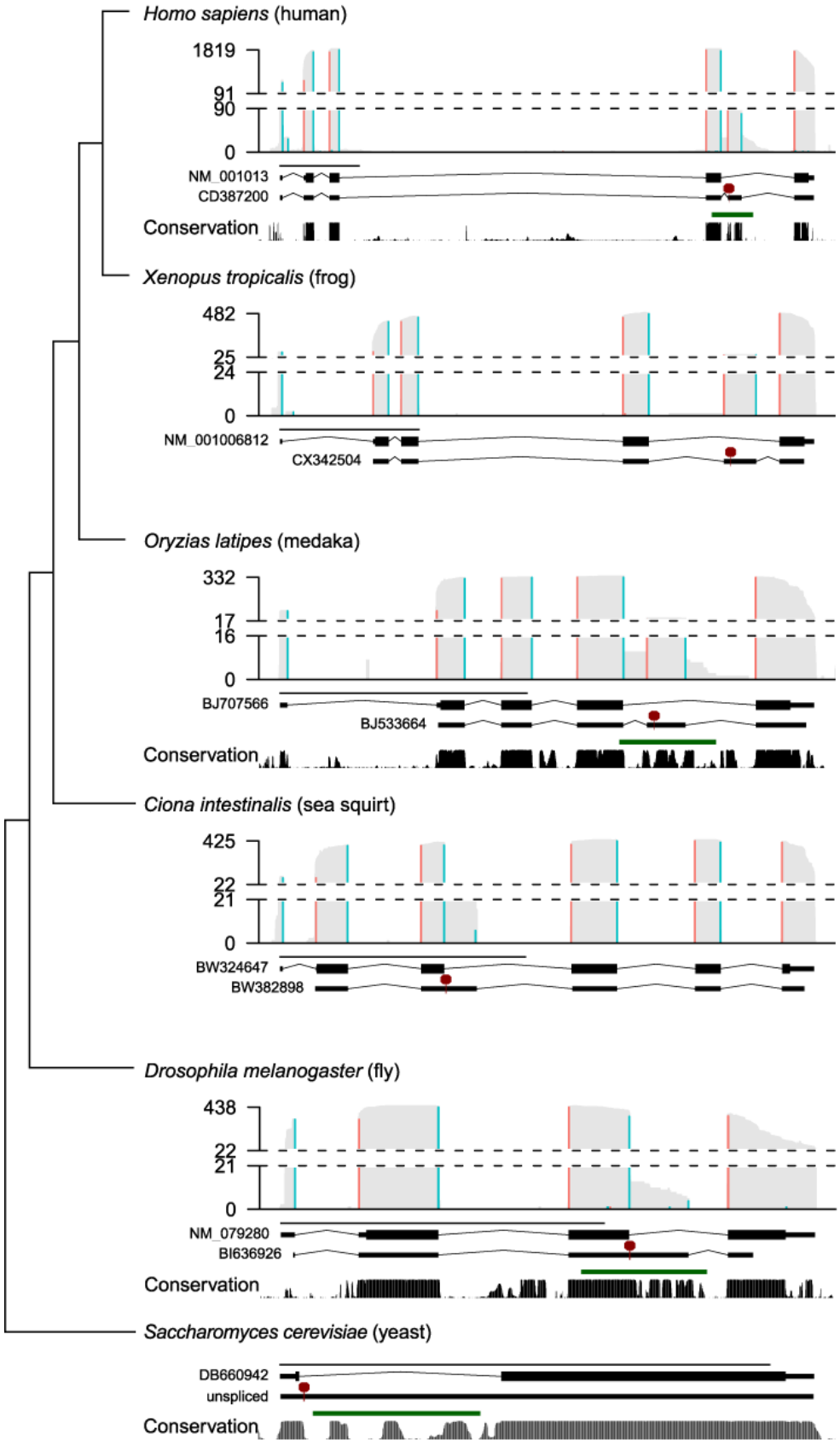

Homo sapiens (human)
1819
91
90
0
NM_001013
CD387200
Conservation
Xenopus tropicalis (frog)
482
25
24
0
NM_001006812
CX342504
Oryzias latipes (medaka)
332
17
16
0
BJ707566
BJ533664
Conservation
Ciona intestinalis (sea squirt)
425
22
21
0
BW324647
BW382898
Drosophila melanogaster (fly)
438
22
21
0
NM_079280
BI636926
Conservation
Saccharomyces cerevisiae (yeast)
DB660942
unspliced
Conservation

Figure 3.5 Evolutionary models

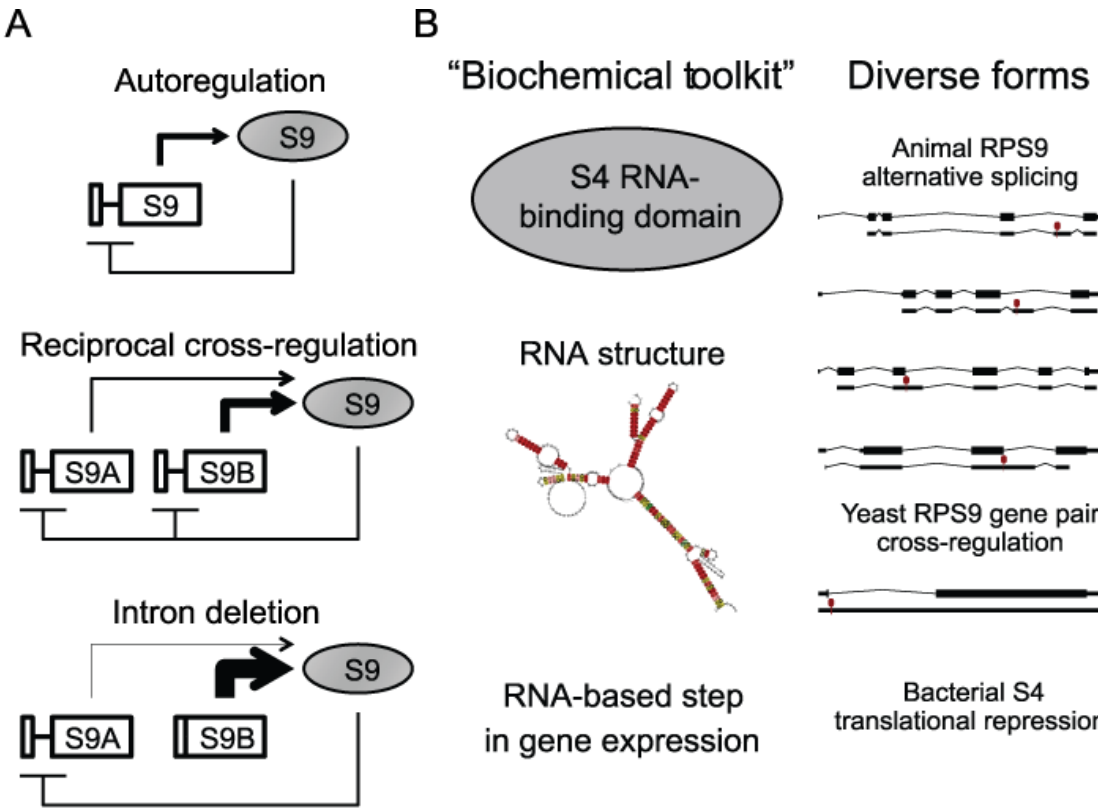


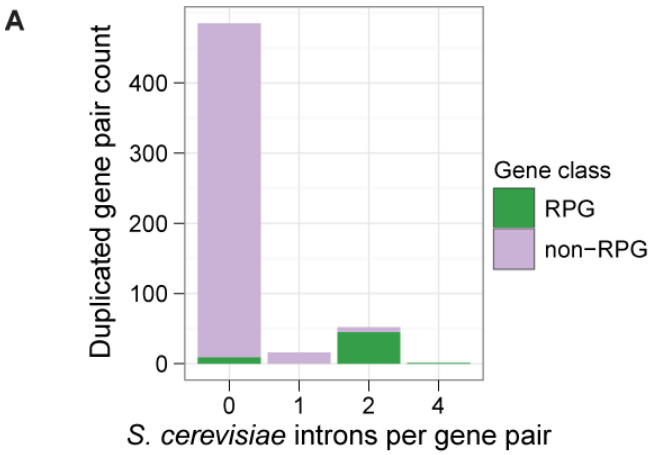
Figure 3.6 Heatmap of annotated and predicted introns in hemiascomycetous yeast orthologs of S. cerevisiae duplicated gene pairs

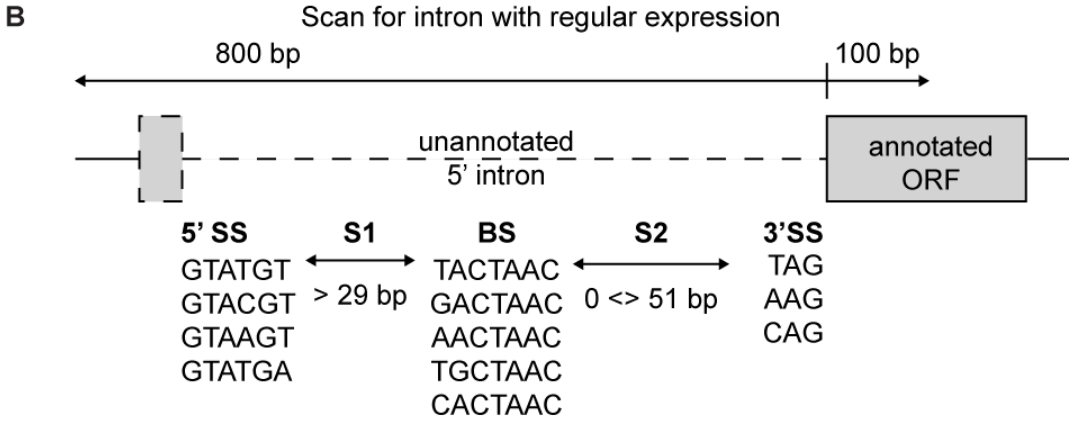
A) Histogram showing illustrating a simple accounting of introns per *S. cerevisiae* gene pair comparing RPGs (green) to other functional gene classes (purple).

B) Diagram illustrating the intron prediction method.

C) The dendogram (top) illustrates the assumed topology of the phylogenetic relationships used to infer intron-containing genes present in the pre-WGD ancestor prior to the WGD event (based on [55]; not to scale); the estimated time of the WGD event is indicated (gray circle and scale bar) [11]. Introns were inferred to be present in the pre-WGD ancestor if 1) present (annotated or predicted) in the majority of outgroup pre-WGD species and *Z. rouxii* or 2) if absent in *Z. rouxii*, but present the majority of both outgroup pre-WGD and post-WGD species. *S. cerevisiae* gene pairs that were derived from an intron-containing pre-WGD ancestor are shown in a heatmap with their orthologs (rows) from other species (columns) to highlight inferred intron losses. Introns present in pre-WGD and post-WGD genes are indicated (brown and shades of blue-green). Post-WGD gene pairs with an intron in both genes (dark blue-green), only one of two genes (light blue-green), or no introns in either gene (white) are shown for each post-WGD species. Post-WGD genes that are present in only one intron-containing copy due to gene loss are indicated (gray). Missing genes (red) or genes containing multiple introns (pink) are indicated, but were not included in the analysis. Genes encoding ribosomal proteins (light blue) and other proteins (light red) are indicated (right-most column). *Ca = C. albicans, Lw = L. waltii, Lt = L. thermotolerans, Lk = L. kluyveri, Eg = E.*

gossypii, Kl = K. lactis, Zr = Z. rouxii, Sc = S. cerevisiae, Sb = S. bayanus, Cg = C. glabrata, Nc = N. castellii and Vp = V. polyspora.


A
Duplicated gene pair count
400
300
200
100
0
0
1
2
4
S. cerevisiae introns per gene pair
Gene class
RPG
non−RPG


B
Scan for intron with regular expression
800 bp
100 bp
unannotated
5' intron
annotated
ORF
5' SS
S1
BS
S2
3'SS
GTATGT
GTACGT
GTAAGT
GTATGA
> 29 bp
TACTAAC
GACTAAC
AACTAAC
TGCTAAC
CACTAAC
0 <> 51 bp
TAG
AAG
CAG

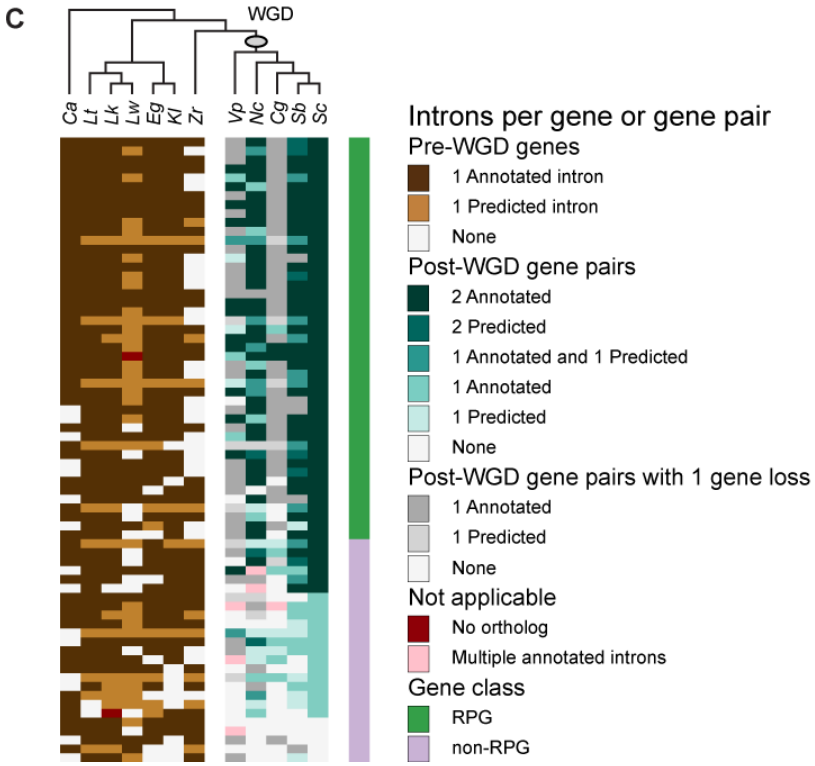

C
WGD
Ca
Lt
Lk
Lw
Eg
Kl
Zr
Vp
Nc
Cg
Sb
Sc
Introns per gene or gene pair
Pre-WGD genes
1 Annotated intron
1 Predicted intron
None
Post-WGD gene pairs
2 Annotated
2 Predicted
1 Annotated and 1 Predicted
1 Annotated
1 Predicted
None
Post-WGD gene pairs with 1 gene loss
1 Annotated
1 Predicted
None
Not applicable
No ortholog
Multiple annotated introns
Gene class
RPG
non-RPG

Figure 3.7 Illustration of the intron deletion method

Diagram illustrating the strain construction protocol used to delete introns.

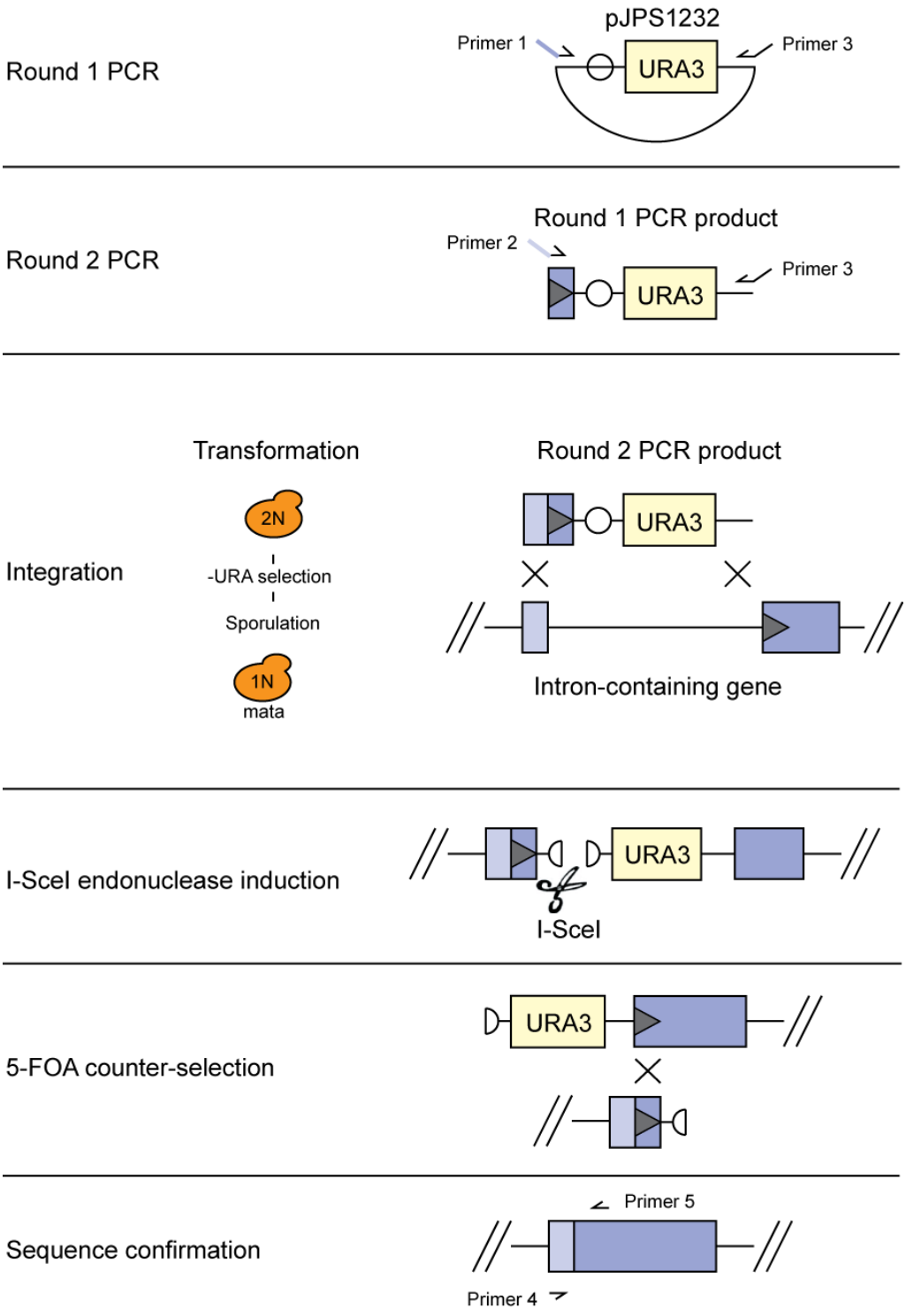
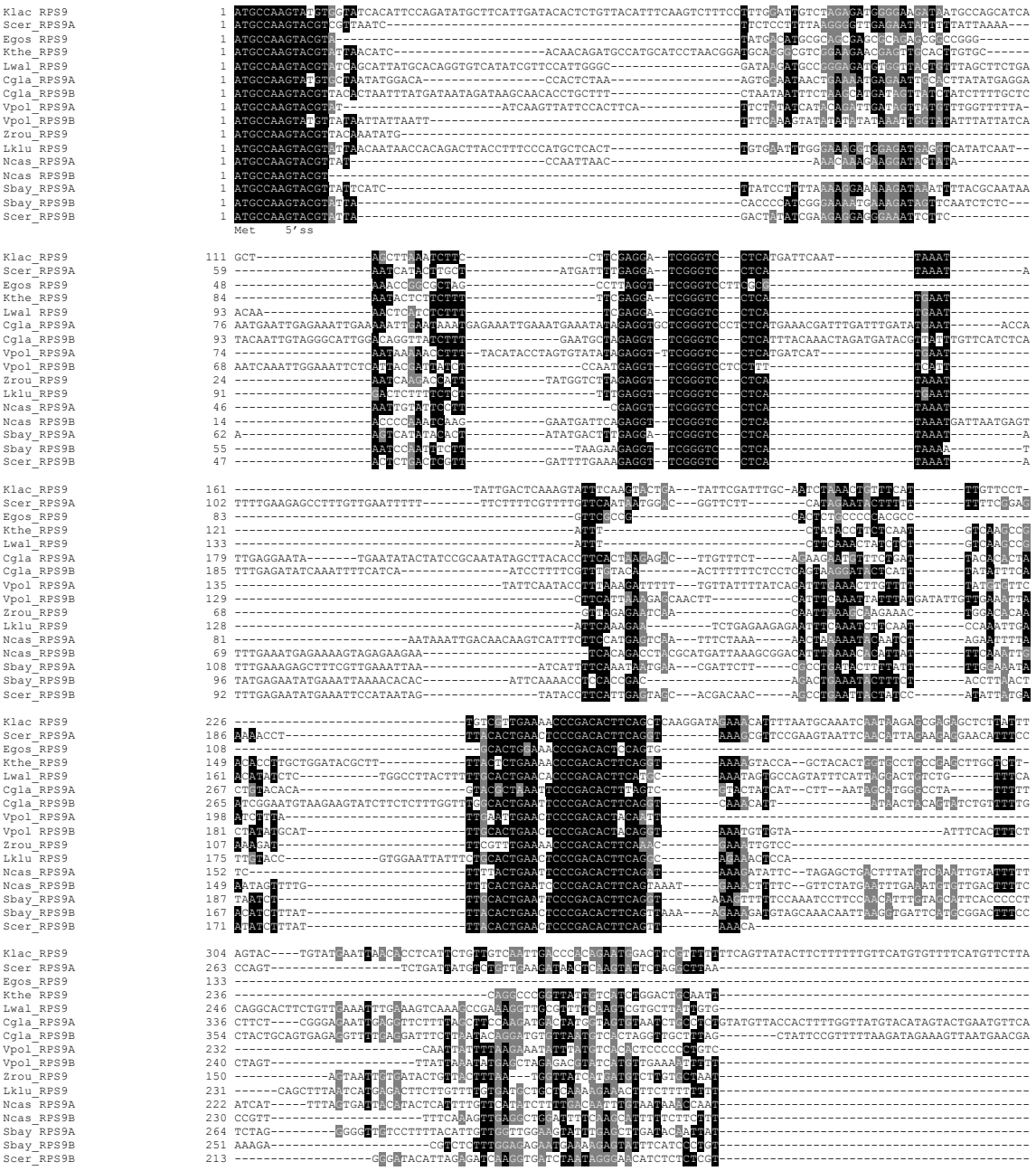


Figure 3.8 Conserved intronic regions among yeast RPS9 orthologs

Nucleotide alignment of genes encoding ribosomal protein S9 from *L. waltii, L. thermotolerans, L. kluyveri, E. gossypii, K. lactis, Z. rouxii, S. cerevisiae, S. bayanus, C. glabrata, N. castellii* and *V. polyspora.* Note that all genes contain an intron with regions of identical sequences (black shading).




Klac_RPS9
Scer_RPS9A
Egos_RPS9
Kthe_RPS9
Lwal_RPS9
Cgla_RPS9A
Cgla_RPS9B
Vpol_RPS9A
Vpol_RPS9B
Zrou_RPS9
Lklu_RPS9
Ncas_RPS9A
Ncas_RPS9B
Sbay_RPS9A
Sbay_RPS9B
Scer_RPS9B
branchsite
3' ss

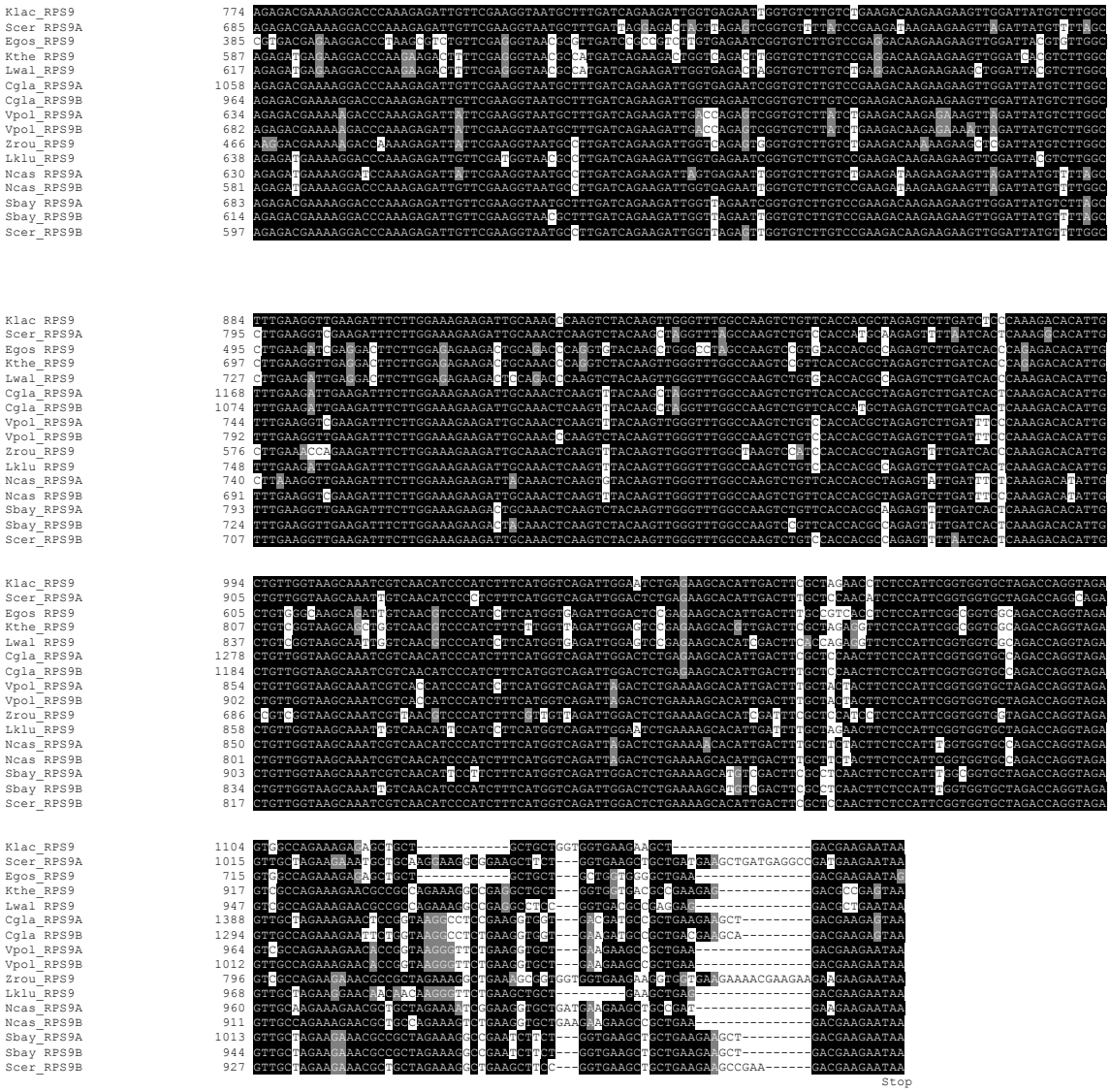

Klac_RPS9
Scer_RPS9A
Egos_RPS9
Kthe_RPS9
Lwal_RPS9
Cgla_RPS9A
Cgla_RPS9B
Vpol_RPS9A
Vpol_RPS9B
Zrou_RPS9
Lklu_RPS9
Ncas_RPS9A
Ncas_RPS9B
Sbay_RPS9A
Sbay_RPS9B
Scer_RPS9B
Stop

Figure 3.9 Detailed splicing-sensitive microarray data of intron deletion mutants

Boxplots summarizing splicing-sensitive microarray expression data of 14 RPG intron deletion strains (x-axis) compared to a common wild-type strain. For each comparison, exon features for the gene in which an intron has been deleted (red point) and its gene copy (blue point) are indicated. Microarray data is expressed as normalized $\log_2$ transformed probe intensity from exon, intron, and junction features. Whiskers represent 1.5 times the interquartile range. Outliers are not shown (black points).

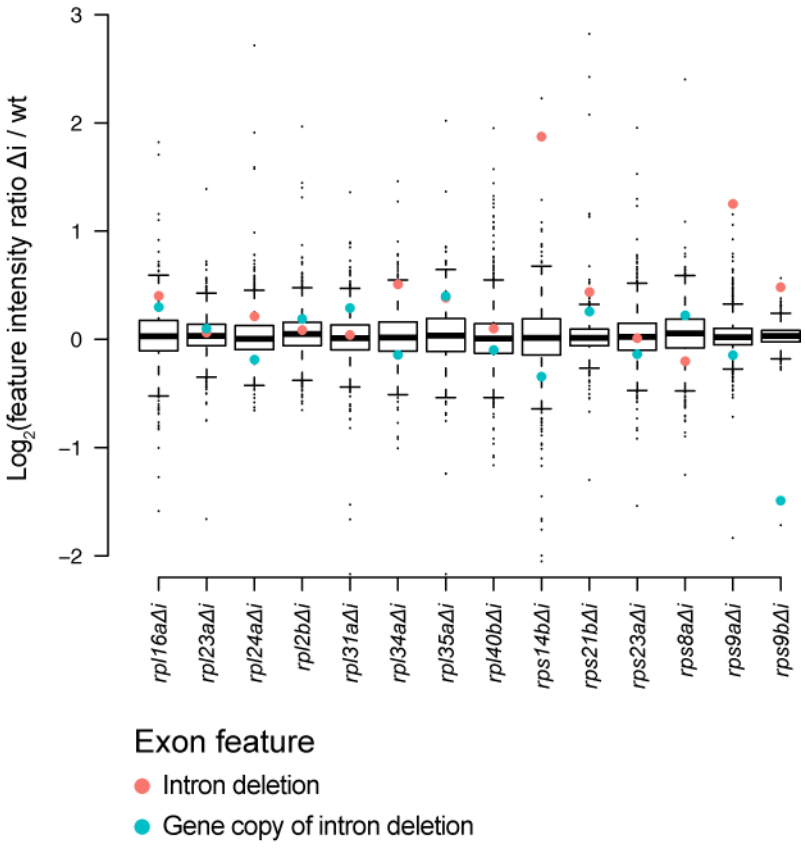


Figure 3.10 Comparison of RPS9 spliced isoforms from 10 animals

EST summaries of RpS9 orthologs from 10 animal species illustrated as in Figure 3.4.

Genes are plotted to scale (black line; 1 kb).

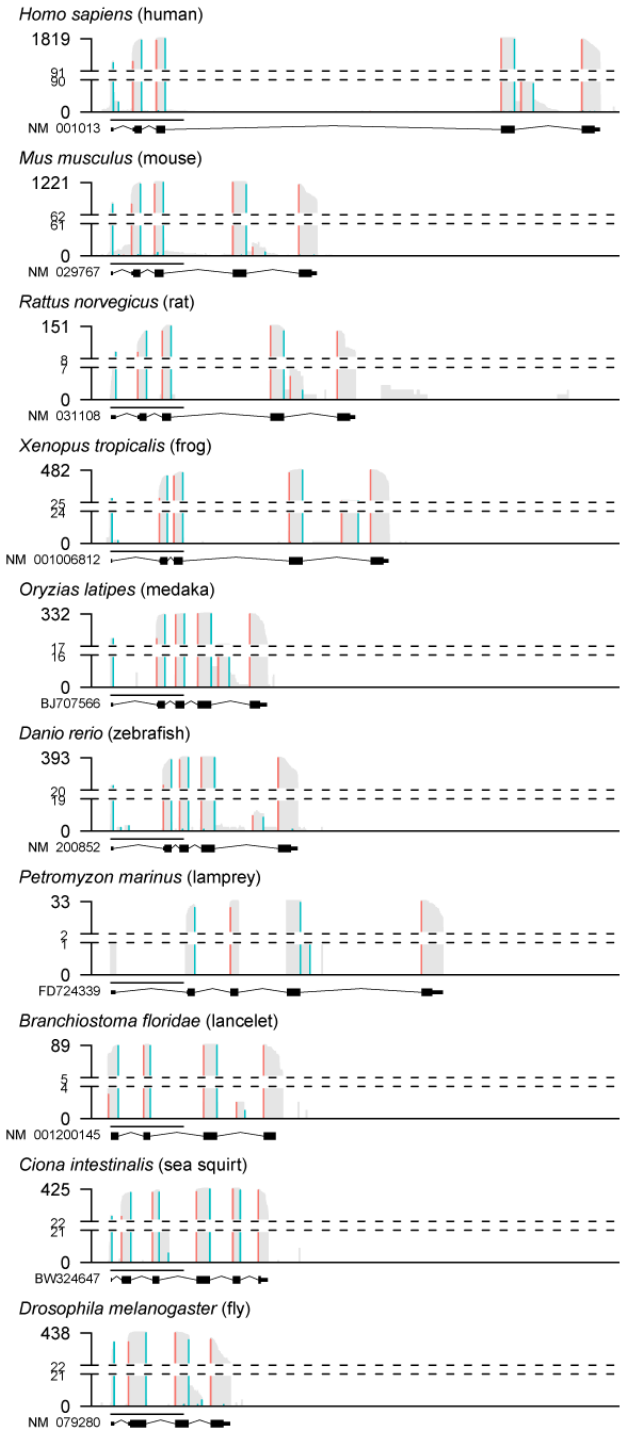
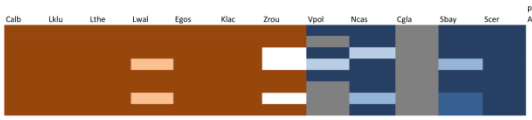


Table 3.1 Intron annotations and predictions

Code	WGD	Intron evidence	Intron number	Gene number	
	pre-WGD	none		0	1
	pre-WGD	1 annotation		1	1
	pre-WGD	1 prediction		1	1
	pre-WGD	1 annotation	multiple per gene		1
	post-WGD	none		0	2
	post-WGD	prediction		1	2
	post-WGD	annotation		1	2
	post-WGD	1 prediction and 1 annotation		2	2
	post-WGD	2 predictions		2	2
	post-WGD	2 annotation		2	2
	post-WGD	annotation	multiple per gene		2
	post-WGD	none		0	1
	post-WGD	1 predictions		1	1
	post-WGD	1 annotation		1	1
	No ortholog	NA	NA	NA	

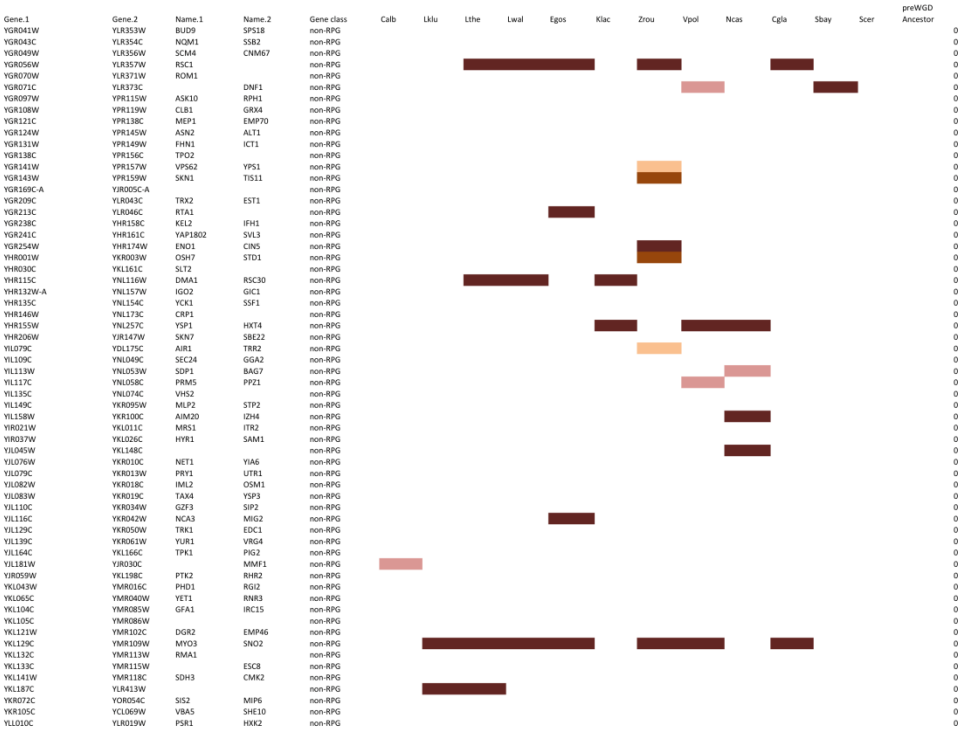
Gene.1	Gene.2	Name.1	Name.2	Gene class	Calb	Klu	Lthe	Lwal	Egos	Klac	Zrou	Vpol	Ncas	Cgla	Sbay	Scer	preWGD Ancestor
Scer_YGOB_SDC25	YLR310C	NA	PUS1	non-RPG													1
Scer_YGOB_YDR134C	YLR110C	NA	FKS1	non-RPG													1
YAL015C	YOL043C	NTG1		non-RPG													1
YAL037W	YOR342C		SKI7	non-RPG													1
YAL051W	YOR363C	OAF1	YPT32	non-RPG													1
YAL056W	YOR371C	GPB2	HRB1	non-RPG													1
YAL062W	YOR375C	GDH3	NHP6A	non-RPG													1
YAR002C-A	YGL002W	ERP1	RIM11	non-RPG													1
YBL005W	YGL013C	PDR3	SEC14	non-RPG													1
YBL009W	YGL021W	ALK2	AFE2	non-RPG													1
YBL039C	YJR103W	URA7	SCS2	non-RPG													1
YBL043W	YJR115W	ECM13	SRC1	non-RPG													1
YBL061C	YER096W	SKT5	PST1	non-RPG													1
YBL067C	YER098W	UBP13		non-RPG													1
YBL068W	YER099C	PIB54	ARF1	non-RPG													1
YBL069W	YER101C	AST1	SNC2	non-RPG													1
YBL079W	YER105C	NUP170		non-RPG													1
YBL101C	YPR030W	ECM21	IMD3	non-RPG													1
YBL106C	YPR032W	SRO77	COX5B	non-RPG													1
YBR001C	YDR001C	NTH2	CPT1	non-RPG													1
YBR014C	YDL010W	GRX7	UBC5	non-RPG													1
YBR068C	YDR046C	BAP2	AIM11	non-RPG													1
YBR104W	YPR058W	YMC2	TUB1	non-RPG													1
YBR117C	YPR074C	TKL2	YSC84	non-RPG													1
YBR145W	YOL086C	ADH5	SUN4	non-RPG													1
YBR150C	YOL089C	TBS1	RPL14A	RPG													1
YBR177C	YPL095C	EHT1	RPL17A	RPG													1
YBR183W	YPL087W	YPC1	RPL27A	RPG													1
YBR214W	YGL056C	SDS24	RPS26A	RPG													1
YBR270C	YJL058C	BIT2	RPL40B	RPG													1
YBR278W	YJL065C	DPB3	RPL43A	RPG													1
YBR284W	YJL070C		RPS10A	RPG													1
YCL036W	YDR514C	GFD2	RPL13B	RPG													1
YCL037C	YDR515W	SRO9	RPS24B	RPG													1
YCR024C-A	YEL017C-A	PMP1	RPL2B	RPG													1
YCR026C	YEL016C	NPP1	RPL24B	RPG													1
YCR048W	YNR019W	ARE1	RPS23B	RPG													1
YCR052W	YNR023W	RSC6	RPS27A	RPG													1
YCR073W-A	YNR034W	SOL2	RPL16B	RPG													1
YCR089W	YNR044W	FIG2	RPS7A	RPG													1
YDL019C	YAR042W	OSH2	RPS4A	RPG													1
YDL036C	YOL066C	PUS9	RPL19A	RPG													1
YDL042C	YOL068C	SIR2	RPS8B	RPG													1
YDL048C	YLR375W	STP4	RPL23B	RPG													1
YDL070W	YLR399C	BDF2	RPS11A	RPG													1
YDL085W	YMR145C	NDE2	RPS6A	RPG													1
YDL155W	YLR210W	CLB3	RPS9A	RPG													1
YDL179W	YDL127W	PCL9	RPL21B	RPG													1
YDL181W	YDL130W-A	INH1	RPS14B	RPG													1
YDL188C	YDL134C	PPH22	RPS29A	RPG													1
YDL194W	YDL138W	SNF3	RPL31B	RPG													1
YDL204W	YDR233C	RTN2	RPS16A	RPG													1
YDL222C	YNL194C	FMP45	RPL35B	RPG													1
YDL224C	YNL197C	WHI4	RPS17A	RPG													1
YDL226C	YNL204C	GCS1	RPS18B	RPG													1
YDL229W	YNL209W	SSB1	RPL37A	RPG													1
YDL239C	YNL225C	ADY3	RPL34B	RPG													1
YDR034W-B	YBR056W-A		RPL22A	RPG													1
YDR085C	YER158C	AFR1	RPS25B	RPG													1
YDR093W	YER166W	DNF2	RPL26A	RPG													1
YDR096W	YER169W	GIS1	RPS0B	RPG													1
YDR098C	YER174C	GRX3	RPL42A	RPG													1



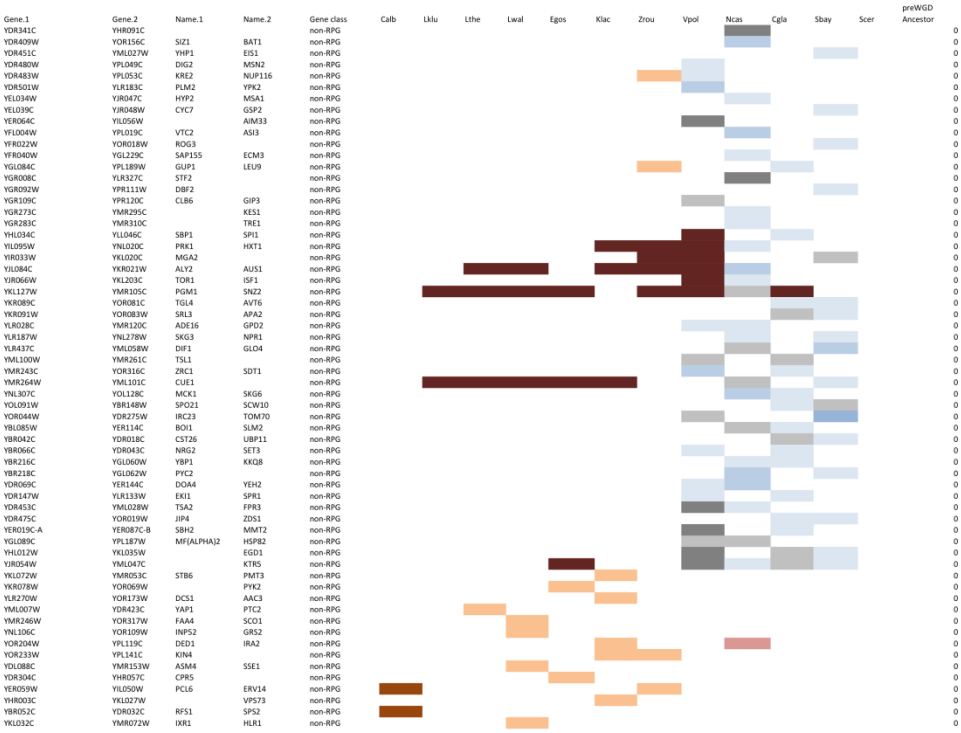
Gene.1	Gene.2	Name.1	Name.2	Gene class	preWGD Ancestor
YDR107C	YLR083C	TMN2	RPS21A	RPG	1
YDR111C	YLR089C	ALT2	RPS30B	RPG	1
YDR125C	YLR090C	ECM18	RPL6A	RPG	1
YDR132C	YLR108C		RPL36B	RPG	1
YDR144C	YLR120C	MKC7	RPL20B	RPG	1
YDR151C	YLR136C	CTH1	RPL18A	RPG	1
YDR178W	YLR164W	SDH4	RPS19A	RPG	1
YDR206W	YLR233C	EBS1	RPL33A	RPG	1
YDR222W	YLR255C		CDC25	non-RPG	0
YDR223W	YLR223C	CRF1	CCW12	non-RPG	0
YDR251W	YPL032C	PAM1	NTG2	non-RPG	0
YDR259C	YOR028C	YAP6		non-RPG	0
YDR277C	YOR047C	MTH1	PIP2	non-RPG	0
YDR300C	YHR033W	PRO1	GPB1	non-RPG	0
YDR303C	YHR056C	RSC3	GDH1	non-RPG	0
YDR309C	YHR061C	GIC2	ERP6	non-RPG	0
YDR312W	YHR066W	SSF2	PDR1	non-RPG	0
YDR326C	YHR080C	YSP2	ALK1	non-RPG	0
YDR342C	YHR092C	HXT7	URA8	non-RPG	0
YDR351W	YHR103W	SBE2		non-RPG	0
YDR353W	YHR106W	TRR1	SHC1	non-RPG	0
YDR358W	YHR108W	GGA1	UBP9	non-RPG	0
YDR389W	YOR134W	SAC7	PRS2	non-RPG	0
YDR436W	YML016C	PPZ2	AST2	non-RPG	0
YDR438W	YML018C	THI74	NUP157	non-RPG	0
YDR463W	YHR006W	STP1	CSR2	non-RPG	0
YDR492W	YOL101C	IZH1	SRO7	non-RPG	0
YDR497C	YOL103W	ITR1	NTH1	non-RPG	0
YDR502C	YLR180W	SAM2	GRX6	non-RPG	0
YDR505C	YLR177W	PSP1	BAP3	non-RPG	0
YEL006W	YIL006W	YEA6	YMC1	non-RPG	0
YEL041W	YJR049C	YEF1	TKL1	non-RPG	0
YEL047C	YJR051W		ADH1	non-RPG	0
YEL060C	YOR003W	PRB1	HAL9	non-RPG	0
YER027C	YGL208W	GAL83	EEB1	non-RPG	0
YER028C	YGL209W	MIG3	YDC1	non-RPG	0
YER035W	YGL222C	EDC2	SDS23	non-RPG	0
YER039C	YGL225W	HVG1	BIT61	non-RPG	0
YER054C	YIL045W	GIP2	DLS1	non-RPG	0
YER057C	YIL051C	HMF1		non-RPG	0
YER062C	YIL053W	HOR2		non-RPG	0
YER067W	YIL057C	RGI1	SLF1	non-RPG	0
YER070W	YIL066C	RNR1	PMP2	non-RPG	0
YFL018C	YPL017C	LPD1	NPP2	non-RPG	0
YFL042C	YLR072W		ARE2	non-RPG	0
YFL048C	YLR080W	EMP47	SNF12	non-RPG	0
YFL060C	YNL334C	SNO3	SOL1	non-RPG	0
YFR012W	YOL019W		AGA1	non-RPG	0
YFR013W	YOL017W	IOC3	SWH1	non-RPG	0
YFR014C	YOL016C	CMK1	RIB2	non-RPG	0
YFR017C	YOL024W	IGD1	HST1	non-RPG	0
YFR023W	YHR015W	PES4	STP3	non-RPG	0
YFR039C	YGL228W		BDF1	non-RPG	0
YFR053C	YGL253W	HXK1	NDE1	non-RPG	0
YGL049C	YGR162W	TIF4632	CLB4	non-RPG	0
YGL082W	YPL191C		PCL2	non-RPG	0
YGL133W	YPL216W	ITC1	STF1	non-RPG	0
YGL134W	YPL219W	PCL10	PPH21	non-RPG	0
YGL139W	YPL221W	FLC3	RGT2	non-RPG	0
YGL162W	YPR009W	SUT1	RTN1	non-RPG	0
YGR010W	YLR328W	NMA2		non-RPG	0
YGR038W	YLR350W	ORM1	WHI3	non-RPG	0



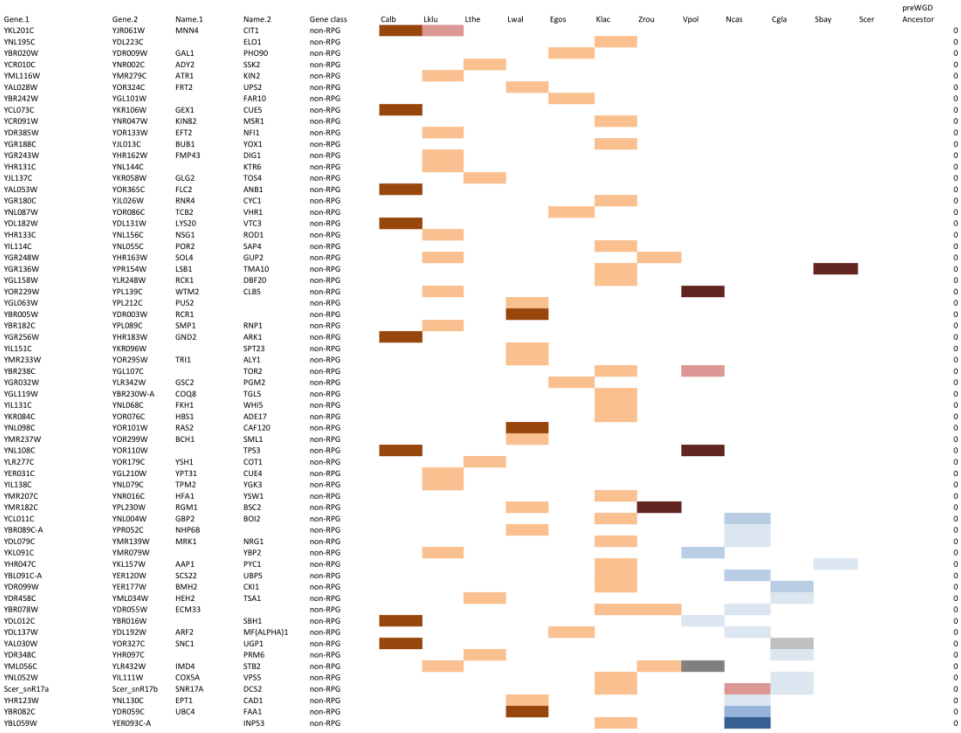
Gene.1	Gene.2	Name.1	Name.2	Gene class	preWGD Ancestor
YML125C	YML087C	PGA3	MRH1	non-RPG	0
YMR119W	YNL008C	ASI1	TEF1	non-RPG	0
YMR206W	YNR014W		SUR1	non-RPG	0
YNL095C	YOR092W		SYH1	non-RPG	0
YNL104C	YOR108W	LEU4	UBX6	non-RPG	0
YNL309W	YOL131W	STB1	GIN4	non-RPG	0
YNR034W-A	YCR075W-A		GNP1	non-RPG	0
YOR227W	YPL137C	HER1	GRX2	non-RPG	0
YOR237W	YPL145C	HES1	EMI2	non-RPG	0
YOR256C	YPL176C	TRE2	EUG1	non-RPG	0
YDR077W	YER150W	SED1	SEC12	non-RPG	0
YDR343C	YHR094C	HXT6	CPR8	non-RPG	0
YGR197C	YJR015W	SNG1		non-RPG	0
YIL013C	YOR011W	PDR11	GPM3	non-RPG	0
YKL093W	YMR081C	MBR1	ROG1	non-RPG	0
YFL059W	YNL333W	SNZ3	ENT2	non-RPG	0
YBL089W	YER119C	AVT5	TDA7	non-RPG	0
YCL050C	YDR530C	APA1	REG2	non-RPG	0
YDL022W	YOL059W	GPD1	RTR1	non-RPG	0
YDL214C	YNL183C	PRR2	ACE2	non-RPG	0
YDR272W	YOR040W	GLO2	ECM22	non-RPG	0
YDR524C-B	YCL048W-A		SKS1	non-RPG	0
YER037W	YGL224C	PHM8	MET31	non-RPG	0
YGR035C	YLR346C		EXO1	non-RPG	0
YGR221C	YHR149C	TOS2	AKR2	non-RPG	0
YGR279C	YMR305C	SCW4	GCY1	non-RPG	0
YHR117W	YNL121C	TOM71	RGA1	non-RPG	0
YIL105C	YNL047C	SLM1	APT1	non-RPG	0
YIL156W	YKR098C	UBP7	PKH2	non-RPG	0
YIL105W	YKR029C	SET4	GEA1	non-RPG	0
YJL165C	YKL168C	HAL5	SSC1	non-RPG	0
YKR075C	YOR062C		CST6	non-RPG	0
YLL012W	YLR020C	YEH1	SER33	non-RPG	0
YLR300W	YOR190W	EXG1	TOS3	non-RPG	0
YLR449W	YML074C	FPR4	MDS3	non-RPG	0
YML109W	YMR273C	ZDS2	GSY2	non-RPG	0
YMR177W	YPL224C	MMT1	AFT2	non-RPG	0
YMR186W	YPL240C	HSC82	CSM4	non-RPG	0
YDR252W	YPL037C	BTT1	CUP9	non-RPG	0
YIL085C	YNL029C	KTR7	HAA1	non-RPG	0
YAL023C	YOR321W	PMT2	PEX30	non-RPG	0
YAL038W	YOR347C	CDC19	MID2	non-RPG	0
YBL030C	YBR085W	PET9	CUR1	non-RPG	0
YBL056W	YER089C	PTC3	TDH2	non-RPG	0
YBR024W	YBR037C	SCO2	SPO12	non-RPG	0
YBR121C	YPR081C	GRS1	PEX18	non-RPG	0
YBR140C	YOL081W	IRA1	LAC1	non-RPG	0
YBR147W	YOL092W	RTC2	PTM1	non-RPG	0
YBR169C	YPL106C	SSE2	YBT1	non-RPG	0
YBR197C	YPL077C		HXT3	non-RPG	0
YBR210W	YGL054C	ERV15	YGP1	non-RPG	0
YBR241C	YGL104C		YAP7	non-RPG	0
YCL048W	YDR522C	SPS22	SAP190	non-RPG	0
YCL051W	YDR528W	LRE1	BCH2	non-RPG	0
YCR005C	YNR001C	CIT2	CAF4	non-RPG	0
YCR034W	YJL196C	FEN1	MRS4	non-RPG	0
YCR037C	YJL198W	PHO87	YSR3	non-RPG	0
YCR073C	YNR031C	SSK22	TIF1	non-RPG	0
YDR122W	YLR096W	KIN1	PIR3	non-RPG	0
YDR185C	YLR168C	UPS3	PIR1	non-RPG	0
YDR200C	YLR238W	VPS64	PUF2	non-RPG	0
YDR273W	YOR042W	DON1	TDA9	non-RPG	0



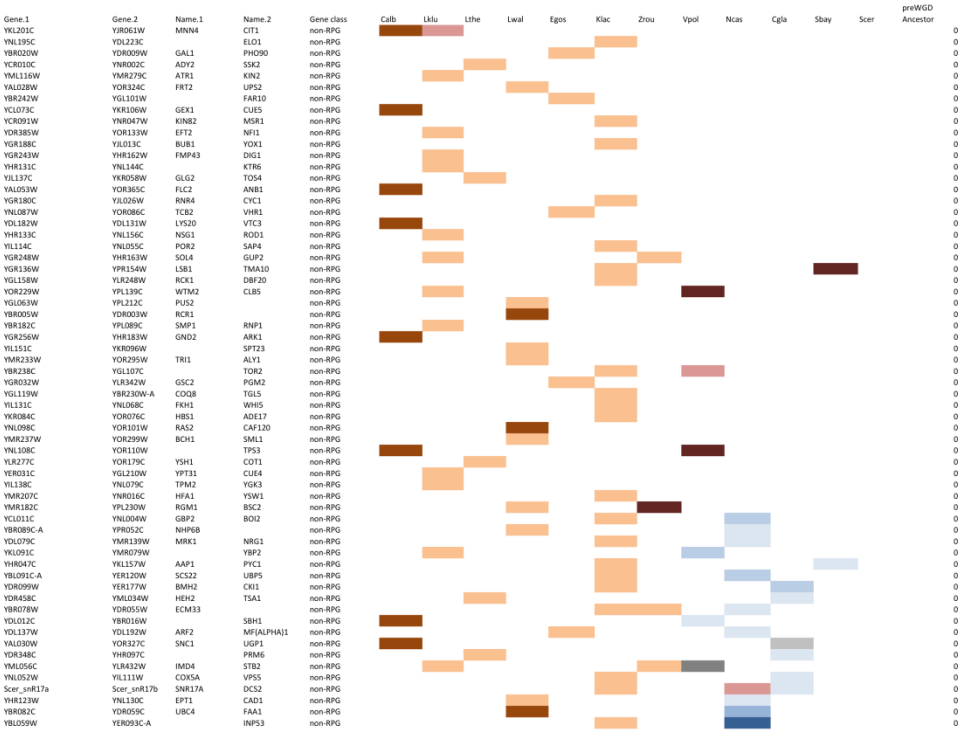
Gene.1	Gene.2	Name.1	Name.2	Gene class	Calb	Lklu	Lthe	Lwal	Egos	Klac	Zrou	Vpol	Ncas	Cgla	Sbay	Scer	preWGD Ancestor
YGR041W	YLR353W	BUD9	SPS18	non-RPG													0
YGR043C	YLR354C	NQM1	SSB2	non-RPG													0
YGR049W	YLR356W	SCM4	CNM67	non-RPG													0
YGR056W	YLR357W	RSC1		non-RPG													0
YGR070W	YLR371W	ROM1		non-RPG													0
YGR071C	YLR373C		DNF1	non-RPG													0
YGR097W	YPR115W	ASK10	RPH1	non-RPG													0
YGR108W	YPR119W	CLB1	GRX4	non-RPG													0
YGR121C	YPR138C	MEP1	EMP70	non-RPG													0
YGR124W	YPR145W	ASN2	ALT1	non-RPG													0
YGR131W	YPR149W	FHN1	ICT1	non-RPG													0
YGR138C	YPR156C	TPO2		non-RPG													0
YGR141W	YPR157W	VPS62	YPS1	non-RPG													0
YGR143W	YPR159W	SKN1	TIS11	non-RPG													0
YGR169C-A	YJR005C-A			non-RPG													0
YGR209C	YLR043C	TRX2	EST1	non-RPG													0
YGR213C	YLR046C	RTA1		non-RPG													0
YGR238C	YHR158C	KEL2	IFH1	non-RPG													0
YGR241C	YHR161C	YAP1802	SVL3	non-RPG													0
YGR254W	YHR174W	ENO1	CIN5	non-RPG													0
YHR001W	YKR003W	OSH7	STD1	non-RPG													0
YHR030C	YKL161C	SLT2		non-RPG													0
YHR115C	YNL116W	DMA1	RSC30	non-RPG													0
YHR132W-A	YNL157W	IGO2	GIC1	non-RPG													0
YHR135C	YNL154C	YCK1	SSF1	non-RPG													0
YHR146W	YNL173C	CRP1		non-RPG													0
YHR155W	YNL257C	YSP1	HXT4	non-RPG													0
YHR206W	YJR147W	SKN7	SBE22	non-RPG													0
YIL079C	YDL175C	AIR1	TRR2	non-RPG													0
YIL109C	YNL049C	SEC24	GGA2	non-RPG													0
YIL113W	YNL053W	SDP1	BAG7	non-RPG													0
YIL117C	YNL058C	PRM5	PPZ1	non-RPG													0
YIL135C	YNL074C	VHS2		non-RPG													0
YIL149C	YKR095W	MLP2	STP2	non-RPG													0
YIL158W	YKR100C	AIM20	IZH4	non-RPG													0
YIR021W	YKL011C	MRS1	ITR2	non-RPG													0
YIR037W	YKL026C	HYR1	SAM1	non-RPG													0
YJL045W	YKL148C			non-RPG													0
YJL076W	YKR010C	NET1	YIA6	non-RPG													0
YJL079C	YKR013W	PRY1	UTR1	non-RPG													0
YJL082W	YKR018C	IML2	OSM1	non-RPG													0
YJL083W	YKR019C	TAX4	YSP3	non-RPG													0
YJL110C	YKR034W	GZF3	SIP2	non-RPG													0
YJL116C	YKR042W	NCA3	MIG2	non-RPG													0
YJL129C	YKR050W	TRK1	EDC1	non-RPG													0
YJL139C	YKR061W	YUR1	VRG4	non-RPG													0
YJL164C	YKL166C	TPK1	PIG2	non-RPG													0
YJL181W	YJR030C		MMF1	non-RPG													0
YJR059W	YKL198C	PTK2	RHR2	non-RPG													0
YKL043W	YMR016C	PHD1	RGI2	non-RPG													0
YKL065C	YMR040W	YET1	RNR3	non-RPG													0
YKL104C	YMR085W	GFA1	IRC15	non-RPG													0
YKL105C	YMR086W			non-RPG													0
YKL121W	YMR102C	DGR2	EMP46	non-RPG													0
YKL129C	YMR109W	MYO3	SNO2	non-RPG													0
YKL132C	YMR113W	RMA1		non-RPG													0
YKL133C	YMR115W		ESC8	non-RPG													0
YKL141W	YMR118C	SDH3	CMK2	non-RPG													0
YKL187C	YLR413W			non-RPG													0
YKR072C	YOR054C	SIS2	MIP6	non-RPG													0
YKR105C	YCL069W	VBA5	SHE10	non-RPG													0
YLL010C	YLR019W	PSR1	HXK2	non-RPG													0



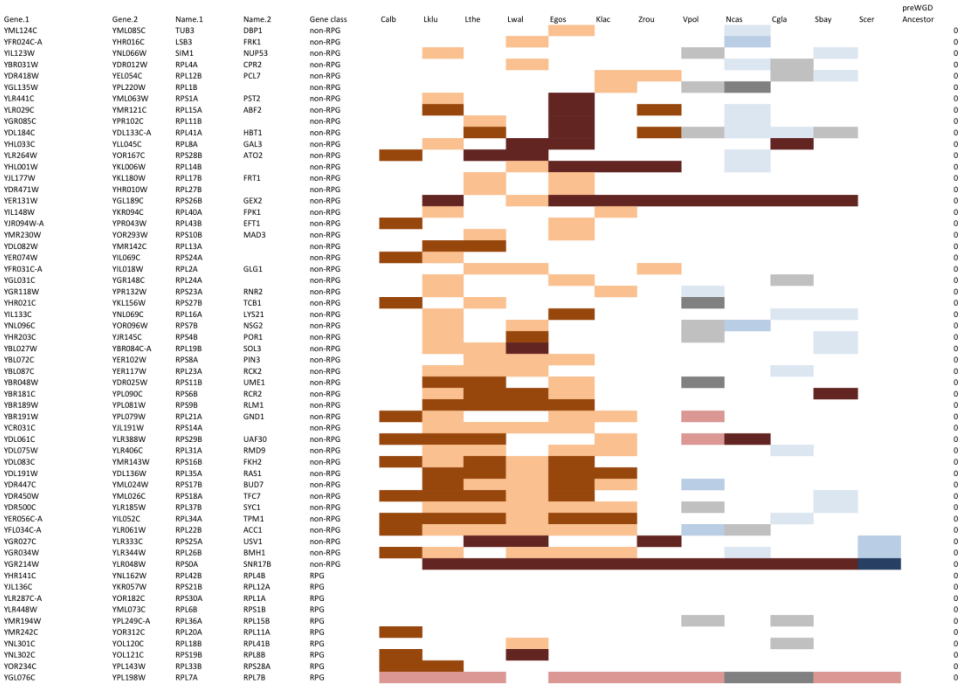
Gene.1	Gene.2	Name.1	Name.2	Gene class	preWGD Ancestor
YDR341C	YHR091C			non-RPG	0
YDR409W	YOR156C	SIZ1	BAT1	non-RPG	0
YDR451C	YML027W	YHP1	EIS1	non-RPG	0
YDR480W	YPL049C	DIG2	MSN2	non-RPG	0
YDR483W	YPL053C	KRE2	NUP116	non-RPG	0
YDR501W	YLR183C	PLM2	YPK2	non-RPG	0
YEL034W	YJR047C	HYP2	MSA1	non-RPG	0
YEL039C	YJR048W	CYC7	GSP2	non-RPG	0
YER064C	YIL056W		AIM33	non-RPG	0
YFL004W	YPL019C	VTC2	ASI3	non-RPG	0
YFR022W	YOR018W	ROG3		non-RPG	0
YFR040W	YGL229C	SAP155	ECM3	non-RPG	0
YGL084C	YPL189W	GUP1	LEU9	non-RPG	0
YGR008C	YLR327C	STF2		non-RPG	0
YGR092W	YPR111W	DBF2		non-RPG	0
YGR109C	YPR120C	CLB6	GIP3	non-RPG	0
YGR273C	YMR295C		KES1	non-RPG	0
YGR283C	YMR310C		TRE1	non-RPG	0
YHL034C	YIL046C	SBP1	SPI1	non-RPG	0
YIL095W	YNL020C	PRK1	HXT1	non-RPG	0
YIR033W	YKL020C	MGA2		non-RPG	0
YJL084C	YKR021W	ALY2	AUS1	non-RPG	0
YJR066W	YKL203C	TOR1	ISF1	non-RPG	0
YKL127W	YMR105C	PGM1	SNZ2	non-RPG	0
YKR085C	YOR081C	TGL4	AVT6	non-RPG	0
YKR091W	YOR083W	SRL3	APA2	non-RPG	0
YLR028C	YMR120C	ADE16	GPD2	non-RPG	0
YLR187W	YNL278W	SKG3	NPR1	non-RPG	0
YLR437C	YML058W	DIF1	GLO4	non-RPG	0
YML100W	YMR261C	TSL1		non-RPG	0
YMR243C	YOR316C	ZRC1	SDT1	non-RPG	0
YMR264W	YML101C	CUE1		non-RPG	0
YNL307C	YOL128C	MCK1	SKG6	non-RPG	0
YOL091W	YBR148W	SPO21	SCW10	non-RPG	0
YOR044W	YDR275W	IRC23	TOM70	non-RPG	0
YBL085W	YER114C	BOI1	SLM2	non-RPG	0
YBR042C	YDR018C	CST26	UBP11	non-RPG	0
YBR066C	YDR043C	NRG2	SET3	non-RPG	0
YBR216C	YGL060W	YBP1	KKQ8	non-RPG	0
YBR218C	YGL062W	PYC2		non-RPG	0
YDR069C	YER144C	DOA4	YEH2	non-RPG	0
YDR147W	YLR133W	EKI1	SPR1	non-RPG	0
YDR453C	YML028W	TSA2	FPR3	non-RPG	0
YDR475C	YOR019W	JIP4	ZDS1	non-RPG	0
YER019C-A	YER087C-B	SBH2	MMT2	non-RPG	0
YGL089C	YPL187W	MF(ALPHA)2	HSP82	non-RPG	0
YHL012W	YKL035W		EGD1	non-RPG	0
YJR054W	YML047C		KTR5	non-RPG	0
YKL072W	YMR053C	STB6	PMT3	non-RPG	0
YKR078W	YOR069W		PYK2	non-RPG	0
YLR270W	YOR173W	DCS1	AAC3	non-RPG	0
YML007W	YDR423C	YAP1	PTC2	non-RPG	0
YMR246W	YOR317W	FAA4	SCO1	non-RPG	0
YNL106C	YOR109W	INP52	GRS2	non-RPG	0
YOR204W	YPL119C	DED1	IRA2	non-RPG	0
YOR233W	YPL141C	KIN4		non-RPG	0
YDL088C	YMR153W	ASM4	SSE1	non-RPG	0
YDR304C	YHR057C	CPR5		non-RPG	0
YER059W	YIL050W	PCL6	ERV14	non-RPG	0
YHR003C	YKL027W		VPS73	non-RPG	0
YBR052C	YDR032C	RFS1	SPS2	non-RPG	0
YKL032C	YMR072W	IXR1	HLR1	non-RPG	0



Gene.1	Gene.2	Name.1	Name.2	Gene class	preWGD Ancestor
YKL201C	YJR061W	MNN4	CIT1	non-RPG	0
YNL195C	YDL223C		ELO1	non-RPG	0
YBR020W	YDR009W	GAL1	PHO90	non-RPG	0
YCR010C	YNR002C	ADY2	SSK2	non-RPG	0
YML116W	YMR279C	ATR1	KIN2	non-RPG	0
YAL028W	YOR324C	FRT2	UPS2	non-RPG	0
YBR242W	YGL101W		FAR10	non-RPG	0
YCL073C	YKR106W	GEX1	CUE5	non-RPG	0
YCR091W	YNR047W	KIN82	MSR1	non-RPG	0
YDR385W	YOR133W	EFT2	NFI1	non-RPG	0
YGR188C	YJL013C	BUB1	YOX1	non-RPG	0
YGR243W	YHR162W	FMP43	DIG1	non-RPG	0
YHR131C	YNL144C		KTR6	non-RPG	0
YIL137C	YKR058W	GLG2	TOS4	non-RPG	0
YAL053W	YOR365C	FLC2	ANB1	non-RPG	0
YGR180C	YJL026W	RNR4	CYC1	non-RPG	0
YNL087W	YOR086C	TCB2	VHR1	non-RPG	0
YDL182W	YDL131W	LYS20	VTC3	non-RPG	0
YHR133C	YNL156C	NSG1	ROD1	non-RPG	0
YIL114C	YNL055C	POR2	SAP4	non-RPG	0
YGR248W	YHR163W	SOL4	GUP2	non-RPG	0
YGR136W	YPR154W	LSB1	TMA10	non-RPG	0
YGL158W	YLR248W	RCK1	DBF20	non-RPG	0
YOR229W	YPL139C	WTM2	CLB5	non-RPG	0
YGL063W	YPL212C	PUS2		non-RPG	0
YBR005W	YDR003W	RCR1		non-RPG	0
YBR182C	YPL089C	SMP1	RNP1	non-RPG	0
YGR256W	YHR183W	GND2	ARK1	non-RPG	0
YIL151C	YKR096W		SPT23	non-RPG	0
YMR233W	YOR295W	TRI1	ALY1	non-RPG	0
YBR238C	YGL107C		TOR2	non-RPG	0
YGR032W	YLR342W	GSC2	PGM2	non-RPG	0
YGL119W	YBR230W-A	COQ8	TGL5	non-RPG	0
YIL131C	YNL068C	FKH1	WHI5	non-RPG	0
YKR084C	YOR076C	HBS1	ADE17	non-RPG	0
YNL098C	YOR101W	RAS2	CAF120	non-RPG	0
YMR237W	YOR299W	BCH1	SML1	non-RPG	0
YNL108C	YOR110W		TPS3	non-RPG	0
YLR277C	YOR179C	YSH1	COT1	non-RPG	0
YER031C	YGL210W	YPT31	CUE4	non-RPG	0
YIL138C	YNL079C	TPM2	YGK3	non-RPG	0
YMR207C	YNR016C	HFA1	YSW1	non-RPG	0
YMR182C	YPL230W	RGM1	BSC2	non-RPG	0
YCL011C	YNL004W	GBP2	BOI2	non-RPG	0
YBR089C-A	YPR052C	NHP6B		non-RPG	0
YDL079C	YMR139W	MRK1	NRG1	non-RPG	0
YKL091C	YMR079W		YBP2	non-RPG	0
YHR047C	YKL157W	AAP1	PYC1	non-RPG	0
YBL091C-A	YER120W	SCS22	UBP5	non-RPG	0
YDR099W	YER177W	BMH2	CKI1	non-RPG	0
YDR458C	YML034W	HEH2	TSA1	non-RPG	0
YBR078W	YDR055W	ECM33		non-RPG	0
YDL012C	YBR016W		SBH1	non-RPG	0
YDL137W	YDL192W	ARF2	MFA[ALPHA]1	non-RPG	0
YAL030W	YOR327C	SNC1	UGP1	non-RPG	0
YDR348C	YHR097C		PRM6	non-RPG	0
YML056C	YLR432W	IMD4	STB2	non-RPG	0
YNL052W	YIL111W	COX5A	VPS5	non-RPG	0
Scer_snR17a	Scer_snR17b	SNR17A	DCS2	non-RPG	0
YHR123W	YNL130C	EPT1	CAD1	non-RPG	0
YBR082C	YDR059C	UBC4	FAA1	non-RPG	0
YBL059W	YER093C-A		INP53	non-RPG	0



Gene.1	Gene.2	Name.1	Name.2	Gene class	preWGD Ancestor
YKL201C	YJR061W	MNN4	CIT1	non-RPG	0
YNL195C	YDL223C		ELO1	non-RPG	0
YBR020W	YDR009W	GAL1	PHO90	non-RPG	0
YCR010C	YNR002C	ADY2	SSK2	non-RPG	0
YML116W	YMR279C	ATR1	KIN2	non-RPG	0
YAL028W	YOR324C	FRT2	UPS2	non-RPG	0
YBR242W	YGL101W		FAR10	non-RPG	0
YCL073C	YKR106W	GEX1	CUE5	non-RPG	0
YCR091W	YNR047W	KIN82	MSR1	non-RPG	0
YDR385W	YOR133W	EFT2	NFI1	non-RPG	0
YGR188C	YJL013C	BUB1	YOX1	non-RPG	0
YGR243W	YHR162W	FMP43	DIG1	non-RPG	0
YHR131C	YNL144C		KTR6	non-RPG	0
YIL137C	YKR058W	GLG2	TOS4	non-RPG	0
YAL053W	YOR365C	FLC2	ANB1	non-RPG	0
YGR180C	YJL026W	RNR4	CYC1	non-RPG	0
YNL087W	YOR086C	TCB2	VHR1	non-RPG	0
YDL182W	YDL131W	LYS20	VTC3	non-RPG	0
YHR133C	YNL156C	NSG1	ROD1	non-RPG	0
YIL114C	YNL055C	POR2	SAP4	non-RPG	0
YGR248W	YHR163W	SOL4	GUP2	non-RPG	0
YGR136W	YPR154W	LSB1	TMA10	non-RPG	0
YGL158W	YLR248W	RCK1	DBF20	non-RPG	0
YOR229W	YPL139C	WTM2	CLB5	non-RPG	0
YGL063W	YPL212C	PUS2		non-RPG	0
YBR005W	YDR003W	RCR1		non-RPG	0
YBR182C	YPL089C	SMP1	RNP1	non-RPG	0
YGR256W	YHR183W	GND2	ARK1	non-RPG	0
YIL151C	YKR096W		SPT23	non-RPG	0
YMR233W	YOR295W	TRI1	ALY1	non-RPG	0
YBR238C	YGL107C		TOR2	non-RPG	0
YGR032W	YLR342W	GSC2	PGM2	non-RPG	0
YGL119W	YBR230W-A	COQ8	TGL5	non-RPG	0
YIL131C	YNL068C	FKH1	WHI5	non-RPG	0
YKR084C	YOR076C	HBS1	ADE17	non-RPG	0
YNL098C	YOR101W	RAS2	CAF120	non-RPG	0
YMR237W	YOR299W	BCH1	SML1	non-RPG	0
YNL108C	YOR110W		TPS3	non-RPG	0
YLR277C	YOR179C	YSH1	COT1	non-RPG	0
YER031C	YGL210W	YPT31	CUE4	non-RPG	0
YIL138C	YNL079C	TPM2	YGK3	non-RPG	0
YMR207C	YNR016C	HFA1	YSW1	non-RPG	0
YMR182C	YPL230W	RGM1	BSC2	non-RPG	0
YCL011C	YNL004W	GBP2	BOI2	non-RPG	0
YBR089C-A	YPR052C	NHP6B		non-RPG	0
YDL079C	YMR139W	MRK1	NRG1	non-RPG	0
YKL091C	YMR079W		YBP2	non-RPG	0
YHR047C	YKL157W	AAP1	PYC1	non-RPG	0
YBL091C-A	YER120W	SCS22	UBP5	non-RPG	0
YDR099W	YER177W	BMH2	CKI1	non-RPG	0
YDR458C	YML034W	HEH2	TSA1	non-RPG	0
YBR078W	YDR055W	ECM33		non-RPG	0
YDL012C	YBR016W		SBH1	non-RPG	0
YDL137W	YDL192W	ARF2	MF(ALPHA)1	non-RPG	0
YAL030W	YOR327C	SNC1	UGP1	non-RPG	0
YDR348C	YHR097C		PRM6	non-RPG	0
YML056C	YLR432W	IMD4	STB2	non-RPG	0
YNL052W	YIL111W	COX5A	VPS5	non-RPG	0
Scer_snR17a	Scer_snR17b	SNR17A	DCS2	non-RPG	0
YHR123W	YNL130C	EPT1	CAD1	non-RPG	0
YBR082C	YDR059C	UBC4	FAA1	non-RPG	0
YBL059W	YER093C-A		INP53	non-RPG	0



Gene.1	Gene.2	Name.1	Name.2	Gene class	Calb	Lklu	Lthe	Lwal	Egos	Klac	Zrou	Vpol	Ncas	Cgla	Sbay	Scer	preWGD Ancestor
YML124C	YML085C	TUB3	DBP1	non-RPG													0
YFR024C-A	YHR016C	LSB3	FRK1	non-RPG													0
YIL123W	YNL066W	SIM1	NUP53	non-RPG													0
YBR031W	YDR012W	RPL4A	CPR2	non-RPG													0
YDR418W	YEL054C	RPL12B	PCL7	non-RPG													0
YGL135W	YPL220W	RPL1B		non-RPG													0
YLR441C	YML063W	RPS1A	PST2	non-RPG													0
YLR029C	YMR121C	RPL15A	ABF2	non-RPG													0
YGR085C	YPR102C	RPL11B		non-RPG													0
YDL184C	YDL133C-A	RPL41A	HBT1	non-RPG													0
YHL033C	YLL045C	RPL8A	GAL3	non-RPG													0
YLR264W	YOR167C	RPS28B	ATO2	non-RPG													0
YHL001W	YKL006W	RPL14B		non-RPG													0
YIL177W	YKL180W	RPL17B	FRT1	non-RPG													0
YDR471W	YHR010W	RPL27B		non-RPG													0
YER131W	YGL189C	RPS26B	GEX2	non-RPG													0
YIL148W	YKR094C	RPL40A	FPK1	non-RPG													0
YJR094W-A	YPR043W	RPL43B	EFT1	non-RPG													0
YMR230W	YOR293W	RPS10B	MAD3	non-RPG													0
YDL082W	YMR142C	RPL13A		non-RPG													0
YER074W	YIL069C	RPS24A		non-RPG													0
YFR031C-A	YIL018W	RPL2A	GLG1	non-RPG													0
YGL031C	YGR148C	RPL24A		non-RPG													0
YGR118W	YPR132W	RPS23A	RNR2	non-RPG													0
YHR021C	YKL156W	RPS27B	TCB1	non-RPG													0
YIL133C	YNL069C	RPL16A	LYS21	non-RPG													0
YNL096C	YOR096W	RPS7B	NSG2	non-RPG													0
YHR203C	YJR145C	RPS4B	POR1	non-RPG													0
YBL027W	YBR084C-A	RPL19B	SOL3	non-RPG													0
YBL072C	YER102W	RPS8A	PIN3	non-RPG													0
YBL087C	YER117W	RPL23A	RCK2	non-RPG													0
YBR048W	YDR025W	RPS11B	UME1	non-RPG													0
YBR181C	YPL090C	RPS6B	RCR2	non-RPG													0
YBR189W	YPL081W	RPS9B	RLM1	non-RPG													0
YBR191W	YPL079W	RPL21A	GND1	non-RPG													0
YCR031C	YJL191W	RPS14A		non-RPG													0
YDL061C	YLR388W	RPS29B	UAF30	non-RPG													0
YDL075W	YLR406C	RPL31A	RMD9	non-RPG													0
YDL083C	YMR143W	RPS16B	FKH2	non-RPG													0
YDL191W	YDL136W	RPL35A	RAS1	non-RPG													0
YDR447C	YML024W	RPS17B	BUD7	non-RPG													0
YDR450W	YML026C	RPS18A	TFC7	non-RPG													0
YDR500C	YLR185W	RPL37B	SYC1	non-RPG													0
YER056C-A	YIL052C	RPL34A	TPM1	non-RPG													0
YFL034C-A	YLR061W	RPL22B	ACC1	non-RPG													0
YGR027C	YLR333C	RPS25A	USV1	non-RPG													0
YGR034W	YLR344W	RPL26B	BMH1	non-RPG													0
YGR214W	YLR048W	RPS0A	SNR17B	non-RPG													0
YHR141C	YNL162W	RPL42B	RPL4B	RPG													0
YIL136C	YKR057W	RPS21B	RPL12A	RPG													0
YLR287C-A	YOR182C	RPS30A	RPL1A	RPG													0
YLR448W	YML073C	RPL6B	RPS1B	RPG													0
YMR194W	YPL249C-A	RPL36A	RPL15B	RPG													0
YMR242C	YOR312C	RPL20A	RPL11A	RPG													0
YNL301C	YOL120C	RPL18B	RPL41B	RPG													0
YNL302C	YOL121C	RPS19B	RPL8B	RPG													0
YOR234C	YPL143W	RPL33B	RPS28A	RPG													0
YGL076C	YPL198W	RPL7A	RPL7B	RPG													0

Table 3.2 Primers used in this study

Organism	Template	Amplicon	Forward Primer	Reverse Primer	Technique	Figure
Sc	pJPS1232	RPL16 Round 1 PCR	aacaaatcggagaaaatgtctgttgaaccagttgttgtcatttgatggTAAGGGTCATTTAGTAGGTCGTTT	gatgctcattcaaaaaaatggattttagtatatactctttaaatTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL16 Round 2 PCR	gtaaggtctatttagtagtctcgttagtctccagttgttgtcataggcGTATTACCCTGTTATCCCTA	gatgctcattcaaaaaaatggattttagtatatactctttaaatTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL23A Round 1 PCR	atatatgtccggtaacggtgctcaaggtactaagtttagaatctcaTTAGGTCTACCAGTCGGTGCCATCA	gaggaatatttttgtagtaaaagcatgaggggagaaaagttagcagatTCCTTACCATTAAGTTTGATC	intron deletion	2A
Sc	pJPS1232	RPL23A Round 2 PCR	ttaagtctctaccagtcggtgcatcatgaactgtgctgacaaacagtGTATTACCCTGTTATCCCTA	gaggaatatttttgtagtaaaagcatgaggggagaaaagttagcagatTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL24A Round 1 PCR	tgctattgccaagttaagcaagcaccaagacaataactttgaggTGATAAAATGAAGGTTGAAATCGAT	gaaaaaaaaaaaatgttagtaagtctatgctttcatagtagccatTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL24A Round 2 PCR	tgataaaatgtgaaggtgaaatcgattctttttcaggtgccaaaatGTATTACCCTGTTATCCCTA	gaaaaaaaaaaaatgttagtaagtctatgctttcatagtagccatTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL2B Round 1 PCR	ttgtcataaactcaccaagaaaaccaaagtattgaaacaatggGTAGAGTTATTCGTAACCAAAGAAA	tgttagtaaatatttctcgaaagttttttcattatcaaaaatatggTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL2B Round 2 PCR	gtagagttatttctgtaaccaaagaaagggtgctgtttctatcttcagTATTACCCTGTTATCCCTA	tgttagtaaatatttctcgaaagttttttcattatcaaaaatatggTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL13A Round 1 PCR	tttgtagaaaagcatagtatgcagcaagcaggaatcgtaacaaatggCCATTTCCAAGAATTTACCAATCTT	aaaaaatgttagtatatacgaaaaaggtgggacgtttcaattagacacctTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL13A Round 2 PCR	ccatttccaagaatttaccaatctttaaagaatcatttcagaaaagcGTATTACCCTGTTATCCCTA	aaaaaatgttagtatatacgaaaaaggtgggacgtttcaattagacacctTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL34A Round 1 PCR	aaccaaagatggcccaacgtgttactttcagagaagaaatccatACAACACCCGTTCTACAAAATCAA	actttctcattgaaaaatatgttagtacgaaatgtttcaaatagGTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL34A Round 2 PCR	acaacacccgtttctaacaaaatcaaggttgttaagacccccaggtgGTATTACCCTGTTATCCCTA	actttctcattgaaaaatatgttagtacgaaatgtttcaaatagGTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL35A Round 1 PCR	ttattgaccgtagaatcaggtacgtgagacgtatatatcgaaatgGCCGGTGTTAAAGCTTACGAACTAA	ccaaatttttatatacgttagtatttatagcactatagaatTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL35A Round 2 PCR	gccggtgttaaagctacgaactatgaaaccaaatccaaggaacaaaGTATTACCCTGTTATCCCTA	ccaaatttttatatacgttagtatttatagcactatagaatTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL40B Round 1 PCR	cccgttcagcaagtagtaaagcccacaaagttcaaatgcaaatCTTCGTTAAGACTTTGACTGGTAAG	agtaagacagcagacgtgcgtgtgctacatcataaagaaagaaaaacttcTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPL40B Round 2 PCR	cttcgttaagacttgactggtaagacatcactttgaaagttgaGTATTACCCTGTTATCCCTA	agtaagacagcagacgtgcgtgctacatcataaagaaagaaaaacttcTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPS14B Round 1 PCR	agactgaaatcaacaactcccaataacaattaagaatggctaacgACCTTGTTCAAGCTCGCGATAACTC	acttttctcattgaaaaatatgttagtacgaaatgtttcaaatagGTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPS14B Round 2 PCR	acctttgttcaagctcgcgataactctcaagtctttggtgttgcaGTATTACCCTGTTATCCCTA	aacagaaacttctgctttttcgattaataaaaaataaaacgaatgagTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPS21B Round 1 PCR	aaagaatatcaataagccaaaatggaaaacgataggggtcaattaGTTTGAATTATATGTTCCAAGAAAGT	acatttgcgctatatctttgtaaaattattctgcacatgaatgaTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPS21B Round 2 PCR	gttgaattatatgttccaagaaagtgttctgctaccaacagaatcGTATTACCCTGTTATCCCTA	acatttgcgctatatctttgtaaaattattctgcacatgaatgaTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPS23A Round 1 PCR	aggtttgaactctgctagaaagctacgtgtccacagagagaaacaaCCGTTGGCGCGAAAACAACTACAAG	aaatgattcgtgtgtaaaaaaacggcctcctgtgacttctgagttTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPS23A Round 2 PCR	ccgtgggcgaaaacaactacaagagagagtgtgggtacgcGTATTACCCTGTTATCCCTA	aaatgattcgtgtgtaaaaaaacggcctcctgtgacttctgagttTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPS8A Round 1 PCR	ttttggtactccacaagaagcgtttataaacacaccaggacaaaaaaTTACAAAATGGGTATTTCTCGTGAT	taaaataaagaagcgtacaaaagtgcgggaattttttttcttatcttTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPS8A Round 2 PCR	ttacaaaatggtatttctcgtgattcttgtcacaaaagatccgcGTATTACCCTGTTATCCCTA	taaaataaagaagcgtacaaaagtgcgggaattttttttcttatcttTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPS9A Round 1 PCR	aaaactaatacagcaacagaaaatacaaaagtatacaacatgcaagAGCCCCAAGAACATATTCCAAGAC	gtgagcgctctcttttttcaaacgaatgcgagggatcaagatcactaTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPS9A Round 2 PCR	gagcccaagaacatatccaagactactaagacaacaatgcaagAGCCCCAAGAACTTACTCTATAAGAC	gtgagcgctctcttttttcaaacgaatgcgagggatcaagatcactaTCCTTACCATTAAGTTGATC	intron deletion	2A
Sc	pJPS1232	RPS9B Round 1 PCR	aagaaagaaagcaacaagacaactagacaacaatgccaaGAGCTCCAAGAACTTACTCTATAAGAC	ttagtcaaaagtttttccgggaacaaaagtctgcgaagggcccaagtTCCTTACCATTAAGTTTGATC	intron deletion	2A
Sc	pJPS1232	RPS9B Round 2 PCR	gagctccaagaacttactctaagactactctacaccaaagagagcGTATTACCCTGTTATCCCTA	ttagtcaaaagtttttccgggaacaaaagtctgcgaagggcccaagtTCCTTACCATTAAGTTTGATC	intron deletion	2A
Sc	cDNA	Total RPS9A	ccaacatctccattcggtgg	aaaatcagcacaatattcttca	qPCR	2B
Sc	cDNA	Total RPS9B	gctagaagaaacgctgctaga	tgaaaaatgatc-gtgattga	qPCR	2B
Dm	cDNA	UPF1_1	GGGCGGGTgcctctcgcaaagtttgttc	GGGCGGGTgttatgatcgcacgcaaagaca	RNAi	3A
Dm	cDNA	UPF1_2	GGGCGGGTtgatcctacggaaactcac	GGGCGGGTccgagcttcttgttagc	RNAi	3A
Dm	cDNA	HNRNPLL	GGGCGGGTgatcactcggccaggaaata	GGGCGGGTacagaggggattctcctcat	RNAi	3A
Dm	cDNA	RPLP0	agaagaagaggaggaggccaag	ccgaaacccatatcatcgtc	qPCR	3B
Dm	cDNA	RPS9 PTC+	tgactccgttcaccaatc	tcaagtctgcgccatgtatt	qPCR	3B
Dm	cDNA	Total exogenous RPS9	gtcggctcaagaggaagaac	cctgctagcttacgtcacca	qPCR	3B
Dm	cDNA	Total endogenous RPS9	gtcggctcaagaggaagaac	tcaagtctgcgccatgtatt	qPCR	3B
Dm	cDNA	GAPDH1	cttcttcaggcgcaccccatt	accgaactctgttgtcgtacc	qPCR	3B
Dm	cDNA	RPS9 cDNA ORF	CACCatggtgaacggccgcatac	ttagtcctcctcctcttcagc	subcloning	3B

Chapter 4: Epilogue

4.1 Chromatin structure reflects gene architecture

4.1.1 Splicing is coupled to transcription

While the fundamental aspects of RNA splicing were initially described *in vitro*, the field has since set its sights on the *in vivo* process, where spliceosome assembly and catalysis occur concurrently with transcription and other RNA processing events. Further, a growing body of work suggests that splicing and transcription are coupled, such that one process can affect the other in important ways (Neugebauer, 2002). Much of this research has focused on transcription elongation dynamics and its influence on splicing. For example, elongation speed of RNA polymerase is now well accepted to modulate the inclusion of alternatively spliced exons (Kornblihtt, de la Mata, Fededa, Munoz, & Nogues, 2004). Similarly, the newly discovered association between chromatin marks and gene architecture, specifically H3K36me3 and exons, suggests that the cross-talk between transcription elongation and splicing may include chromatin structure (Kolasinska-Zwierz et al., 2009).

Much attention has been given to the possibility that chromatin marks direct splicing decisions by modulating RNA polymerase speed or splicing factor recruitment. For detailed discussion of these hypotheses, I direct the reader to recent and comprehensive reviews (Luco, Allo, Schor, Kornblihtt, & Misteli, 2011; Schwartz & Ast, 2010). It should be noted, however, that these hypotheses ignore a more fundamental question: how are intronic and exonic chromatin marks localized in the first place?

4.1.2 Is transcription coupled to splicing?

Intronic and exonic chromatin marks are presumably associated with the elongation phase of RNA polymerase and require transcription for their deposition. This alone, however, cannot explain the specific localization of chromatin marks to introns and exons. To date, spliceosome assembly is the only known cellular mechanism capable of demarcating the boundaries between introns and exons; its substrate is precursor RNA, whereby splice sites are recognized and exons and introns are defined. Thus, localization of chromatin marks to introns and exons are likely to require some aspect of spliceosome assembly on precursor mRNA. This notion, however, begs the question: how can pre-mRNA (the substrate of splicing), specify chromatin marks associated with DNA? One possibility is that RNA splicing and transcription are coupled in a previously unappreciated manner: RNA processing events may feedback onto transcription elongation. Several lines of evidence suggest that this may indeed occur: 1) transcriptional elongation is enhanced by snRNP-associated TAT-SF1 (Fong & Zhou, 2001) and SR proteins (Zhong et al., 2009) in humans and 2) splicing-dependent polymerase pausing in yeast (Alexander, Innocente, Barrass, & Beggs, 2010).

4.1.3 Dissecting the mechanisms underlying intronic and exonic chromatin marking

My colleagues and I first tested the hypothesis that RNA splicing specifies chromatin marks to introns and exons (Chapter 2). We altered splice site usage through regulated alternative splicing, rather than splice site mutations, because this 'scalpel' approach changed splicing patterns without altering DNA sequence. While splice site mutations

would certainly perturb RNA splicing, they could also perturb DNA determinants (e.g. nucleosome positioning sequences), and so, effects caused by such a 'hammer' approach cannot distinguish between determinants encoded in the DNA or RNA. Nonetheless, we (and others) have now demonstrated changes in alternative splicing alone are not sufficient to alter chromatin marks at *CD45* (Figure 2.5; Shukla et al., 2011). It is currently unclear whether *CD45* and *YPEL5* (Chapter 2) represent outliers or the norm, but chromatin marks are not always static: H3K36me3 marking does increase with polymerase pausing and exon inclusion at *CD44* (De Almeida et al., 2011). It is therefore possible that different mechanisms of alternative splicing have different effects on chromatin marking. One intriguing possibility is that exon definition, which is unaltered in some cases of alternative splicing (e.g. *CD45*), may specify chromatin marks independently of splice site usage; conversely, increased exon definition would be expected to alter chromatin marks. Alternatively, DNA elements may also play a role; however, no correspondence between exon-like sequences (a.k.a. ECRs) and exonic chromatin marks have been found (Figure 2.12).

The 'hammer' approach, despite its caveats, has since provided further insight into the mechanisms specifying the exonic mark H3K36me3. Small-molecule inhibition of the spliceosome (Spliceostatin A and Meayamycin) alters both the distribution of RNA polymerase and the exonic chromatin mark H3K36me3 (S. Kim, Kim, Fong, Erickson, & Bentley, 2011; De Almeida et al., 2011). In many genes, the distribution of H3K36me3 increases toward the 3' end of the gene, whereas polymerase levels are generally decreased (S. Kim et al., 2011). Similarly, splice site mutations in the beta-globin gene

change the distribution of H3K36me3, extending the mark beyond the normal termination site (S. Kim et al., 2011). These results are striking given that changes in alternative splicing may have no effect on H3K36me3 at some genes. Nonetheless, the extension of H3K36me3 beyond transcription termination sites is unexpected and may reflect the manner in which the H3K36me3 methyltransferase (or demethylase) is recruited to chromatin.

Perhaps not surprisingly, the localization of the human H3K36me3 methyltransferase, SetD2, appears to differ substantially from its *S. cerevisiae* homologue, Set2p. Both homologues bind phosphorylated Ser2 of the RNA polymerase CTD (M. Li et al., 2005; Kizer et al., 2005); however, this does not appear to dictate the pattern of H3K36me3 in humans. Unlike H3K36me3 at *S. cerevisiae* genes (B. Li et al., 2007), H3K36me3 does not correlate with Ser2 phosphorylation at human genes (De Almeida et al., 2011). Instead, H3K36me3 distributions correlate with spliceosome components (i.e. U5 snRNP), even when altered by splice site mutations (S. Kim et al., 2011). Thus, human genes may rely on spliceosome assembly rather than RNA polymerase Ser2 phosphorylation alone for SetD2 localization to chromatin. In further support of this idea, mass spectrometry analysis of exon definition complexes reveals the presence of SetD2 fragments, albeit in sub-stoichiometric amounts (Schneider et al., 2010). Taken together, these data best support a model in which chromatin modifying enzymes are physically associated with the spliceosome and mark exons during spliceosome assembly (Figure 4.1). However, a significant caveat remains: the use of blunt instruments (i.e. Spliceostatin and splice site

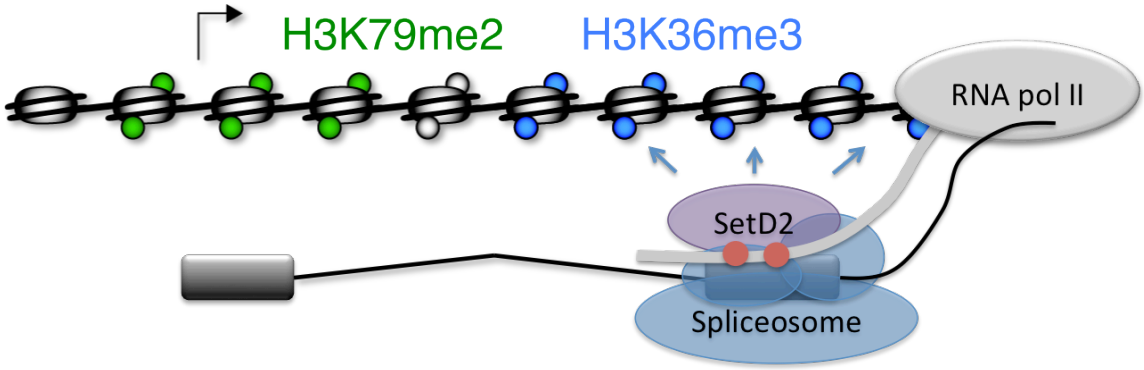
mutations) may have altered transcription elongation or nucleosome positioning sequences in a manner sufficient to change chromatin marking.

4.1.4 Unanswered questions

In Chapter 2, we showed that a large portion of the human epigenome can be explained simply by gene architecture. Out of 41 chromatin marks analyzed, 10 were specifically localized to 5' intronic regions. While subsequent studies (detailed above) have provided evidence that exonic H3K36me3 is determined in part by splicing, the mechanisms underlying intronic marks remain relatively obscure. First, these marks have not been analyzed in the context of splicing inhibition or splice site mutations. Such experiments could determine whether the localization of intronic marks relies on similar or independent mechanisms (Figure 4.1). Second, intronic and exonic regions occasionally overlap near the first 3' splice site or switch back from exonic to intronic in genes with particularly large introns, such as STIM1 (Figure 2.9). Does H3K36me3 at distant exons reflect a chromatin conformation that brings distant exons closer together in a manner analogous to "transcription factories?" Alternatively, does the switch from exonic marks to intronic marks reflect a change from exon definition to intron definition, or vice versa? And lastly, is co-transcriptional splicing sufficient to localize spliceosome components to the chromatin surrounding exons? Are histone methyltrasferases (e.g. SetD2) transferred from the spliceosome to RNA polymerase during spliceosome assembly? Many of these questions will require a greater understanding of the rules governing spliceosome recruitment and assembly at exonic chromatin. Genome-wide

ChIP of splicing factors in genomes with diverse gene architectures will certainly provide a useful starting point.

Figure 4.1 Model: Spliceosome recruitment specifies chromatin marks



4.2 Regulated unproductive splicing: a recurring theme among eukaryotic RNA binding proteins

4.2.1 Regulated unproductive splicing is common in S. cerevisiae

Regulated alternative splicing has been intensely studied as a major mechanism for generating protein diversity (Nilsen & Graveley, 2010). More recently, unproductive splicing has been proposed to be a similarly important mechanism regulating gene expression (Lewis, Green, & Brenner, 2003). The relative importance of these two processes is unclear in animals. In *S. cerevisiae*, only two cases of alternative splicing are known to produce multiple protein isoforms (Juneau et al., 2009; Grund et al., 2008), while unproductive splicing (in the form of "regulated splicing") is commonplace. Unproductive splicing allows *S. cerevisiae* to adapt to different physiological challenges that include environmental stress, development, and homeostasis. All 13 meiotic genes (Juneau et al., 2007) and most of the 103 RPGs (Pleiss, Whitworth, Bergkessel, & Guthrie, 2007a) contain introns that have been shown to be inefficiently spliced when expression would be physiologically inappropriate (asexual cell division and amino acid

starvation, respectively). *YRA1* and *DBP2* each contain a single intron that is inefficiently spliced to maintain homeostasis in response to excess protein production (Barta & Iggo, 1995; Preker, Kim, & Guthrie, 2002). And incredibly, some RPGs, including *RPL30* (Dabeva & Warner, 1993), *RPS14B* (Fewell & Woolford, 1999), *RPS9A* and *RPS9B* (Chapter 3), contain a single intron whose splicing is regulated to meet both the challenges of environmental stress and homeostasis.

It is currently unclear how common unproductive splicing is in other eukaryotes. Recent sequence based assessments of the human transcriptome have revealed staggering numbers of alternative splicing events, such that nearly all genes produce multiple mRNA isoforms (E. T. Wang, Sandberg, et al., 2008). Approximately half these events disrupt the coding reading frame (E. T. Wang, Sandberg, et al., 2008). While the potential for regulated unproductive splicing in humans appears tremendous, it is unclear what fraction of these are functional. Most are not conserved (Pan et al., 2006) and may simply reflect noise (Melamud & Moult, 2009). However, there is evidence that unproductive splicing regulates certain gene classes in animals, including core components of the spliceosome (Saltzman et al., 2008), the entire SR protein gene family (Lareau, Inada, et al., 2007) and several ribosomal proteins (Q M Mitrovich & Anderson, 2000; Cuccurese et al., 2005). SR protein genes contain ultraconserved elements in animals, unlike many unproductively spliced introns (Lareau, Inada, et al., 2007). These examples highlight the potential of unproductive splicing for homeostatic regulation by negative feedback. It is also important to note two of these examples involve proteins that are directly involved in the process of splicing. In Chapter 3, I

showed that RPS9 orthologs in yeast and animals contain conserved elements within introns similar to those found in SR proteins. While not a splicing factor, L30 binds its own nascent pre-mRNA and blocks recruitment of the U2 snRNP (Macías, Bragulat, Tardiff, & Vilardell, 2008). Thus, a recurring theme of unproductive splicing involves direct autoregulation by proteins that bind pre-RNA alone or as part of the splicing machinery.

4.2.2 Does S9 directly regulate alternative splicing?

4.2.2.1 RPS9 fits the theme

Given the theme above, in which RNA binding proteins, including ribosomal proteins, are common and conserved targets of unproductive splicing, I speculated that S9 directly binds to its own pre-mRNA to alter splicing in very different organisms (Figure 3.5); however, this has yet to be tested. Intriguingly, RPS9 alternative splicing in different organisms takes on many different forms (Figure 3.4). Therefore, determining if RPS9 binds its own transcript only solves part of the mystery. How does excess S9 promote intron retention in *S. cerevisiae*, alternative 5' splice site usage in *D. melanogaster*, and exon inclusion in mammals?

Formally, there are many possible pathways for autoregulation, both direct and indirect. In *S. cerevisiae*, S9 has been shown to function in both ribosome biogenesis (Bernstein, Gallagher, Mitchell, Granneman, & Baserga, 2004) and translational fidelity (Vincent & Liebman, 1992) as a component of the small subunit processome (Bernstein et al., 2004) and the mature ribosome. Both of these functions are thought to be mediated by RNA

binding to the precursor and mature 18S rRNA. Thus, RNA binding activity of S9 could indirectly regulate splicing through the translation of splicing factors.

I favor a model in which direct binding of S9 to its own pre-mRNA might stabilize RNA structures that influence splicing. I prefer this model given the precedent set by S4, the bacterial ortholog of S9, which represses its own mRNA translation as part of the α operon of *E. coli* (D. Dean & Nomura, 1980). In Chapter 3, I provide evidence consistent with a similarly direct role for S9 in *RPS9* alternative splicing: specifically, increased in *RPS9* expression results in alternative splicing of introns associated with conserved RNA structures (Figure 3.5). Because S4 translational repression of the α operon mRNA has been extensively studied in *E. coli*, it provides a useful case study to consider when discussing the potential for S9 regulation of alternative splicing. Importantly, S4 autoregulation is unlike the autoregulation of other bacterial ribosomal proteins in two key aspects: the RNA structure of the regulatory binding site 1) does not mimic its cognate rRNA binding site and 2) does not compete between S4 and the ribosome binding. Here, I discuss these aspects and their potential relevance to RPS9 regulation.

4.2.2.2 A study in S4

Unlike other autoregulatory ribosomal proteins, S4 recognizes two distinct and complex RNA structures. The α operon contains a regulatory S4 binding site that cannot be accurately defined by conventional RNA folding algorithms. Instead, comprehensive structural studies, including nucleotide protection assays specific for single- or double-stranded RNA (Deckman & Draper, 1987) and compensatory base pair mutations to confirm predicted helices (C. K. Tang & Draper, 1989) were undertaken to reveal that S4

binds an unusual “double pseudoknot” structure. Intriguingly, this structure can be found in the archeal α operon from *H. halobium* (Sano, Taguchi, Furumoto, Uda, & Itoh, 1999), but not in *B. subtilis*, which utilizes an entirely different RNA structure (Grundy & Henkin, 1991). Surprisingly, none of these structure bear sequence or structural similarity to the 16S rRNA binding site: a complex five-way helical junction (Bellur & Woodson, 2009; Rabl, Leibundgut, Ataide, Haag, & Ban, 2011). Eukaryotic S9 also binds this five-way helical junction along in addition to contacts with two eukaryotic expansions of the eukaryotic 18S rRNA (Rabl et al., 2011). Like the S4 psuedoknot, *in silico* RNA structure predictions of the conserved elements in *S. cerevisiae* and animal introns have failed to identify obvious similarities.

Also unique to S4 translational repression, S4 has only modest affinity to the α operon compared to ribosome binding. However, even a small excess of S4 can repress translation through a combination of factors, including 1) an inherently weak ribosomal binding site, 2) stabilization of an “inactive” RNA conformation by S4 and 3) entrapment, rather than displacement, of the initiating ribosome. First, expression of the α mRNA is relatively poor *in vivo* (C. K. Tang & Draper, 1990), which is likely a consequence of a 30 fold less active rate in translation initiation measured *in vitro* (Schlax, Xavier, Gluick, & Draper, 2001). Strangely, initiation benefits from a GUC initiation codon instead of the nearly universal AUG, which may be due to an atypical arrangement of initiation codon and Shine-Delgarno sequence in order to form the psuedoknot structure (C. K. Tang & Draper, 1990). Second, the regulatory binding site can fold into two distinct conformations (Gluick, Gerstner, & Draper, 1997) that are bound by the ribosome with

near equal affinity (Schlax et al., 2001). At equilibrium, the regulatory S4 binding site switches back and forth between these two conformations and only in the presence of S4 is an inactive complex stabilized (Schlax et al., 2001). Lastly, *in vitro* experiments with labeled S4, α operon psudoknot, and 30 S subunits revealed that a ternary S4·mRNA·30 S subunit "entraptment" complex forms to prevent initiation, rather than a direct competition between S4 and the ribosome. The significance of this later result is that S4 can inhibit translation despite a relatively modest binding affinity to the pseudoknot structure.

4.2.2.3 Future experiments

Applying this knowledge of S4 translational repression to RPS9 regulation is not entirely straightforward, but I will make a few predictions. First, S9 may not require a mimic of the 16S rRNA contacts of the five-way helical junction or the eukaryote-specific expansions. Further, the double pseudoknot structure found of the α operon switches between two conformations, which may explain why pseudoknot prediction programs predict stem loops rather than the pseudoknot structure determined by Draper and colleagues (Dawson, Fujiwara, & Kawai, 2007). It is then entirely possible that the conserved stem loops predicted by *in silico* analysis of RPS9 introns represent just one of multiple RNA structure conformations. Like S4, S9 might bind and stabilize an "inactive" conformation. Thus, it would appear that nucleotide protection assays followed by compensatory base-pair mutational analyses, in the presence and absence of S9, may be needed to properly identify regulatory S9 binding sites.

Second, it is entirely unclear how a highly conserved ribosomal protein might promote very different forms of alternative splicing in different organisms. Because S9 is highly conserved, the precise mechanisms by which S9 promotes intron retention, alternative 5' splice site usage, and exon inclusion may reflect differences in local RNA structure. Further, the entrapment mechanism used by S4 suggests that S9 may not need to directly compete with splicing factors for binding to splice sites. Instead, S9 may stabilize RNA conformations that promote particular splicing patterns.

In yeast, intron retention is equivalent to splicing inhibition, and so, RNA structures that obscure splice sites may be sufficient to produce this form of alternative splicing. Intriguingly, (unpublished) microarray data form mutant strains defective in specific degradation pathways implicate different mechanisms for *RPS9A* and *RPS9B* pre-mRNA turnover: RPS9A pre-mRNA may be largely degraded in the nucleus (most stabilized by Δrrp6 mutation), while RPS9B may be largely degraded in the cytoplasm (most stabilized by Δupf1 mutation). Thus, S9 may stabilize different RNA structures that inhibit splicing differently enough to trigger different RNA surveillance mechanisms. Perhaps spliceosome components that become "entrapped" during assembly would trigger discard and nuclear degradation, while obscured splice sites would allow unspliced pre-mRNA to export and degradation.

The alternative 5' splice site usage seen in *Drosophila* might reflect competition between weak and strong splice sites, which could then be altered by RNA structures that obscure a strong 5' splice site. Exon inclusion in mammals, on the other hand, is

more difficult to explain by simple splice site competition. In this case, exon definition or the progression to intron definition may need to be altered. Here, an entrapment mechanism could possibly interfere with the assembly of exon or intron definition complexes. However, the role of accessory splicing factors, such as SR and hnRNP proteins should also be considered in animals. These binding sites can be occluded by simple changes in RNA structure (Warf & Berglund, 2010). Alternatively, S9 may promote the association of accessory splicing factors to modulate splicing. Indeed, alternatively splicing and degradation of human rpL3 requires excess L3 in association with hnRNP H1 and other proteins (Cuccurese et al., 2005; A. Russo et al., 2010; A. Russo et al., 2011). Thus, the potential role of S9 in alternative splicing represents a rich area of study that may reveal unappreciated regulatory roles for RNA structure in splicing.

Appendix A: Towards a comprehensive understanding of S. cerevisiae *intron function*

A.1 Progress toward a complete intron deletion collection

Initially, I intended to create a single intron deletion (Δi) mutant for every known or putative intron in the *S. cerevisiae* genome. However, this proved to be overly ambitious and I quickly focused my efforts on studying the effects of these mutations faster than I could generate them. Nonetheless, other researchers had similar goals --- some of whom I have worked with in this regard --- ultimately leading to the generation of a large number of Δi strains (Juneau et al., 2006; Parenteau et al., 2008; Parenteau et al., 2011). My goal here is to catalogue these efforts and to compare their findings. To date, the majority of yeast introns have been deleted, and in many cases, multiple isolates from different groups are available (Table A.1).

A.2 The effect of RPG intron deletion on gene expression

The role of introns in RPG expression has long been speculated about, in large part because RPGs are clearly overrepresented among the intron-containing genes of *S. cerevisiae*. While some studies have shown that particular RPG introns are required to repress expression, others have suggested that introns allow RPGs to be highly expressed. Based on microarray analysis of 14 Δi strains I concluded that RPGs require introns to repress gene expression, rather than to achieve high levels of expression (discussed at length in Chapter 3).

Recently, a systematic analysis of RPG Δi strains has provided evidence that, depending on the gene, introns can enhance or repress gene expression (Parenteau et al., 2011). According to this study, 84% of RPG Δi strains showed altered mRNA expression (+/- 20% of wild-type) of the Δi allele and, of these, 64% also altered the mRNA expression of its paralogous gene copy. Superficially, the results reported by Parenteau et al. are reminiscent of the compensatory expression changes that I describe in Chapter 3 for *RPS9A* and *RPS9B*; however, most (80%) of these changes are not compensatory, rather, they report that the expression of both gene copies is either increased or decreased in response to a single intron deletion. Thus, unlike the S9 expression changes that I report, these data cannot be explained by a simple feedback model. Moreover, many of these qPCR measurements reported by Parenteau et al. are not reproduced by my microarray measurements (where direct comparison can be made).

The *RPS9A* and *RPS9B* mRNA expression data mentioned above serves as a useful example of the type of discrepancy common between these two datasets. Parenteau et al. report that in this strain both *RPS9A* and *RPS9B* genes are greatly reduced in expression relative to wild type (5% and 32% respectively). In contrast, my data show a slight relative increase in *RPS9B* expression, by both microarray and qPCR assays (Figure 3.2). Similar discrepancies appear to be common, where small increases (< 2-fold) measured by my microarrays are reported as modest to large (> 2-fold) decreases by their qPCR (Figure A.1, notice that points near 0 of the y-axis, drift toward negative values of the x-axis). Interestingly, the largest relative changes are reproduced in both studies, which may indicate that large effects are the most robust to differences

between expression analysis platforms or Δi strains. Still, it is not possible to identify the source of what appears to be systematic error in one or both of these studies from these data alone.

In order to gain further insight into the biological significance of the expression changes reported by Parenteau et al., I sought to assess the effect that these expression changes might have on the total number of transcripts encoding a given RP. For example, if *RPS9A* and *RPS9B* gene expression is reduced to 5% and 32% wild-type mRNA levels, then the total number of transcripts encoding S9 must be less than wild-type… but by how many transcripts? Interpreting the biological significance of relative mRNA expression changes for the duplicated RPGs of *S. cerevisiae* can be problematic as these genes sometimes contribute very different amounts of mRNA. In the above example, the total amount of mRNA encoding S9 could be as little as 5% or as much as 95% of wild-type, depending on whether *RPS9B* contributes little or most of the total fraction of S9 mRNA in wild-type cells. To overcome these ambiguities, I have determined the relative number of transcripts produced by each gene of an RPG pair using two published RNA-seq datasets (Nagalakshmi et al., 2008; C.-Y. Wu, Rolfe, Gifford, & Fink, 2010). By simply multiplying this value by the fold-change as determined by qPCR, the relative number of transcripts produced by each gene for each Δi strain were determined.

Based on RNA-seq data for a wild-type *S. cerevisiae* strain, *RPS9A* generates only 6 - 16% of the total mRNA encoding S9 (Nagalakshmi et al., 2008; C.-Y. Wu et al., 2010). Thus,

relatively large changes in *RPS9A* expression are required to compensate for relatively small changes in *RPS9B* expression. Indeed, I observed such differences between *RPS9A* and *RPS9B* relative expression changes (Figure 3.2), but more importantly, the total number of transcripts encoding S9 never exceed > 130% of wild-type levels due to compensation by the other gene copy (Figure A.2, left two columns). In contrast, the Parenteau et al. data reveal a large decrease in the total number of transcripts encoding S9 (as little as 30% of wild-type levels in the *rps9bΔi* mutant; Figure A.2, right most column).

While the above example highlights the type of discrepancy observed between these two datasets, I have also shown that large relative changes were the most robust between datasets (Figure A.1). Interestingly, RPGs with the largest relative changes in mRNA expression also contribute the least number of transcripts encoding an RP (Figure A.3). Thus, the largest changes observed in the Parenteau et al. dataset are the most likely to be accurate and these genes are candidates for intron-dependent repression. Like *RPS9A* and *RPS14B* intron-dependent repression (discussed in Chapter 3), *RPL22B* and *RPL18A* expression changes showed large relative expression increases when their introns were deleted, which suggests that these genes may be inefficiently spliced in wild-type cells. Consistenly, RNA-seq reads for both of these genes reveal detectable levels of pre-mRNA in wild-type cells (Figures A.4 and A.5). Further, *RPL22B* and *RPL18A* introns are associated with conserved elements, which in theory, could mediate splicing regulation (Figures A.4 and A.5). And much like the compensatory expression changes observed for *RPS9A*, *RPL22B* expression was greatly decreased when the *RPL22A* intron

was deleted (Figure A.3), suggesting that this gene is not merely inefficiently spliced, but regulated at the level of splicing. Taken together, these data suggest that additional examples of autoregulation indeed exist, but await further testing.

A.3 The effect of intron deletion on fitness

Few studies have assessed the requirement for introns on growth and fitness. For *S. cerevisiae*, at least, a consensus for growth under standard laboratory conditions is emerging. In line with early studies on an *ACT1Δi* strain (Ng, Domdey, Larson, Rossi, & Abelson, 1985), systematic studies of non-RPG and RPG Δi strains (Parenteau et al., 2008; Parenteau et al., 2011; Chapter 3) have found only subtle effects on growth, with a few notable exceptions, such as *YRA1* and *GLC7* (Preker et al., 2002; Juneau et al., 2006). Because obvious growth phenotypes are typically lacking, more sensitive tests that assess fitness, specifically, competition assays with co-cultured strains have more recently been employed. Parenteau et al. reported that a large number of RPG intron deletions can either decrease or increase fitness when co-cultured with wild-type strains (26 and 39 Δi strains, respectively), despite a lack of growth defects in most other conditions. The authors suggest that these surprising findings may reflect non-redundant functions that differ between the copies of RPG pairs (Komili et al., 2007). However, a simpler explanation, which has not been tested, is that decreased RP dosage is responsible for the fitness defects.

The question of whether the two RPG copies have unique or redundant functions has long been debated: an earlier view was that the copies of RPG pairs may generate

different amounts of mRNA, but encode proteins with redundant function (Warner, 1989). This view formed early, based initially on S17 (previously RP51) gene dosage studies. Through gene deletions and northern blots, it was shown that *RPS17A* contributes 60-70% of the total transcripts encoding S17 and that deletion of this gene copy results in a slow growth phenotype, whereas *RPS17B* deletion was less deleterious (Abovich & Rosbash, 1984). Demonstrating functional redundancy by complementation, growth defects caused by *RPS17A* deletion were suppressed by increased *RPS17B* expression (Abovich & Rosbash, 1984). Moreover, deletion of both gene copies resulted in synthetic lethality, which suggests that both gene copies contribute to the same process (Abovich & Rosbash, 1984).

Functional redundancy, as determined by synthetic lethality, has been since been assessed systematically for the duplicated genes in *S. cerevisiae* using gene deletion collections (E. J. Dean, Davis, Davis, & Petrov, 2008). In this study, almost all of the RPG pair deletions were synthetic lethal in combination, which is a much higher percentage than non-RPG pairs (83% and 23%, respectively). This global redundancy among RPGs is likely mediated through gene dosage. Indeed, gene deletion collections have also been employed to demonstrate that *S. cerevisiae* fitness is highly sensitive to the RPG mRNA production. By co-culturing gene deletion collection strains marked by different "genetic barcodes," fitness can be measured by the amount of barcode that hybridizes to a microarray after competition (Pierce, Davis, Nislow, & Giaever, 2007). In one such experiment, quantitative assessments of heterozygous gene deletions revealed that (57%) of RPGs were haploinsufficient for growth, suggesting that as little as a 50%

reduction of transcripts from just one of the two RPG pair copies can impair fitness (Deutschbauer et al., 2005). Taken together, these systematic studies are consistent with two of the observations reported earlier for *RPS17A* and *RPS17B* gene deletions. Specifically that 1) both genes of an RPG pair contribute to a redundant function and 2) *S. cerevisiae* fitness is highly sensitive to RP dosage. Nonetheles, an important conclusion from the *RPS17A* and *RPS17B* gene dosage studies has not been systematically tested: specifically, it has not been shown that there is a quantitative relationship between strain fitness and the total number of transcripts encoding a RPs.

I sought to test the hypothesis that fitness defects are proportional to the total number of RP transcripts decreased by mutations, whether these are caused by gene deletion or intron deletion. If this relationship were known, then strain fitness could be predicted by gene expression analysis alone. Because the two genes of an RPG pair often contribute different amounts of mRNA for a given RP, I predicted that decreases in gene deletion stain fitness would be proportional to the number of transcripts contributed by that gene copy. In order to determine the number of transcripts decreased by each gene deletion, I once again calculated the relative fraction of transcripts produced by each gene copy of an RPG pair as determined by RNA-seq. For each homozygous gene deletions (e.g. *rps17aΔ*), I calculated the number of transcripts encoding a RP (e.g. S17) after gene deletion by simply subtracting that gene's relative fraction of transcripts from the total (C.-Y. Wu et al., 2010). Similarly, for heterozygous gene deletions, ½ of the relative fraction of transcripts were subtracted from the total. Using these calculations, I found a direct correlation (Pearson, 0.6) between fitness and the expected RP transcript

levels (Figure A.6). While a sizable fraction of the data can be explained by this relationship (R^2 = 0.38), there are several biological phenomena that might have added to the unexplained variance in the data. First, many homozygous RPG deletions are known to frequently acquire suppressive mutations that increase the amount of RP mRNA (Hughes et al., 2000). Second, several RPGs can compensate for gene deletions through autoregulation (Warner, 1999). And finally, the relationship between fitness and RP expression may be non-linear.

Nonetheless, having identified a relationship between fitness and RP expression, I sought to determine if this relationship held for Δi strains as well as gene deletions. Parenteau et al. performed co-cultured competition assays for each Δi strain (*ADE2*) and a wild-type (*ade2Δ*) strain, which were counted by red/white colony color. In theory, this should assess fitness similarly to the barcode method mentioned above (although a total of 50 generations elapsed instead of 20). To assess the relationship between the amount of RP mRNA and fitness in this dataset, I first estimated the total amount of RP mRNA based on their reported expression changes and RNA-seq data as described above. Surprisingly, no correlation (Pearson, 0) was found. Limiting the analysis to only Δi strains that decreased the amount of RP mRNA also showed no correlation (data not shown). In fact, Δi strains that greatly reduce total RP mRNA levels are as likely to increase fitness as they are to decrease fitness (Figure A.7).

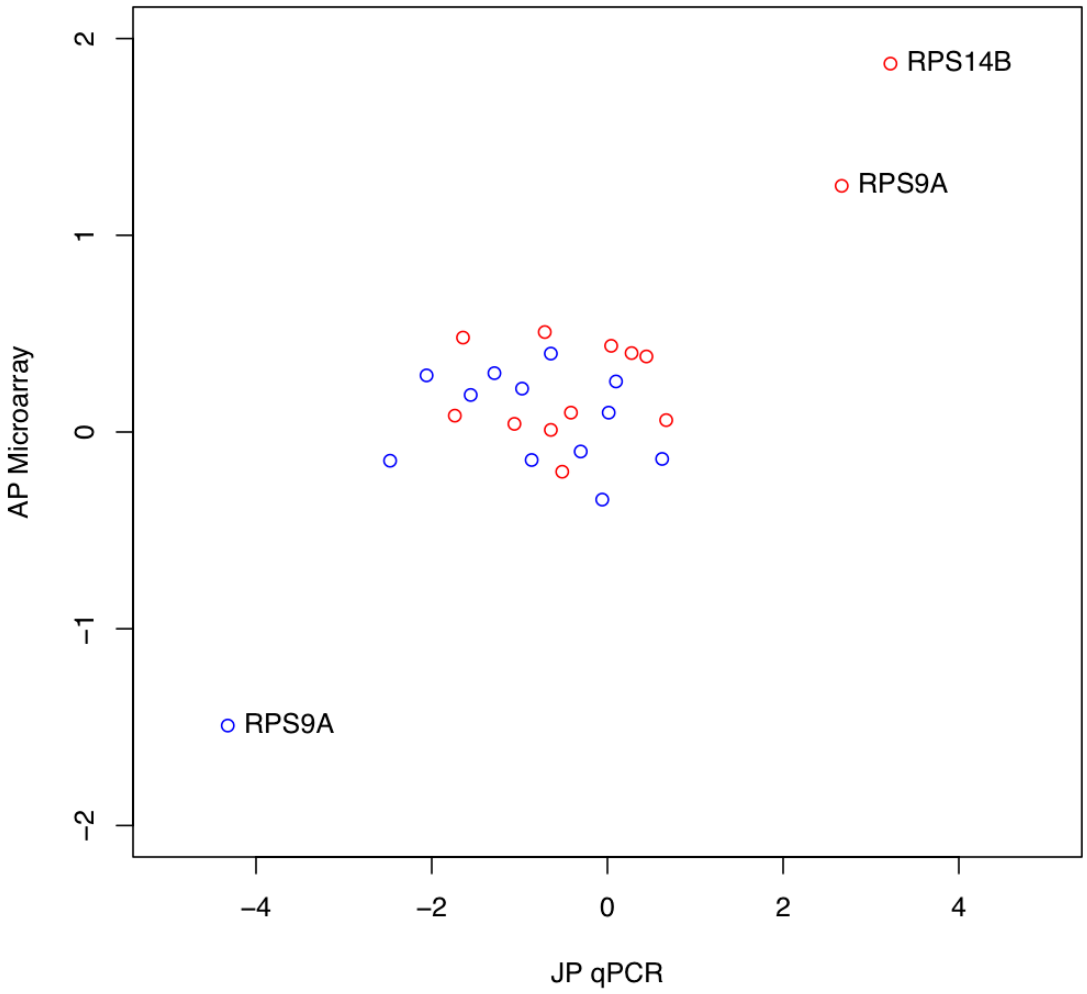
Parenteau et al. report that RPG Δi strains often do not phenocopy gene deletion strains. This may not be surprising given that they report that intron deletions have

various effects on mRNA expression unlike gene deletions. Nonetheless, they report qPCR quantification of expression changes, which in theory, should allow direct comparisons between RPG Δi strains and gene deletions based on the amount of RP mRNA caused by either type of mutation. The lack of correlation between fitness and RP expression in Δi strains suggests that either the expression data and/or the fitness data reported by Parenteau et al. is inaccurate.

In summary, the microarray expression dataset described in Chapter 3 and the Parenteau et al. qPCR expression dataset differ systematically, primarily due to decreases in RPG expression that they report, but are not observed in my dataset. Importantly, these decreased mRNA levels are surprising because they are not correlated with fitness as would be expected from gene deletion studies. Unfortunately, I am unable to identify the source of the discrepancies and so I cannot determine which data points should be disregarded with high confidence. Nonetheless, some of the largest expression changes, such as those for *RPL22B*, may be important for the regulation RPGs and the biology of *S. cerevisiae*.

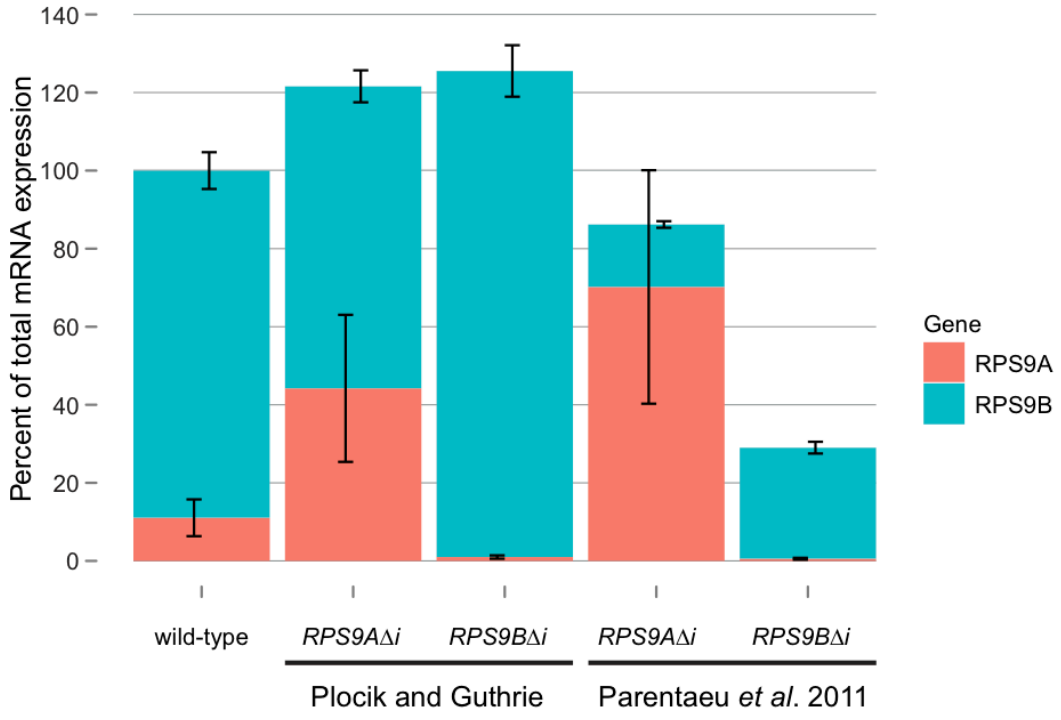
4.3 Figures and Tables

Figure A.1 Microarray data does not support mRNA decreases reported for Δi alleles



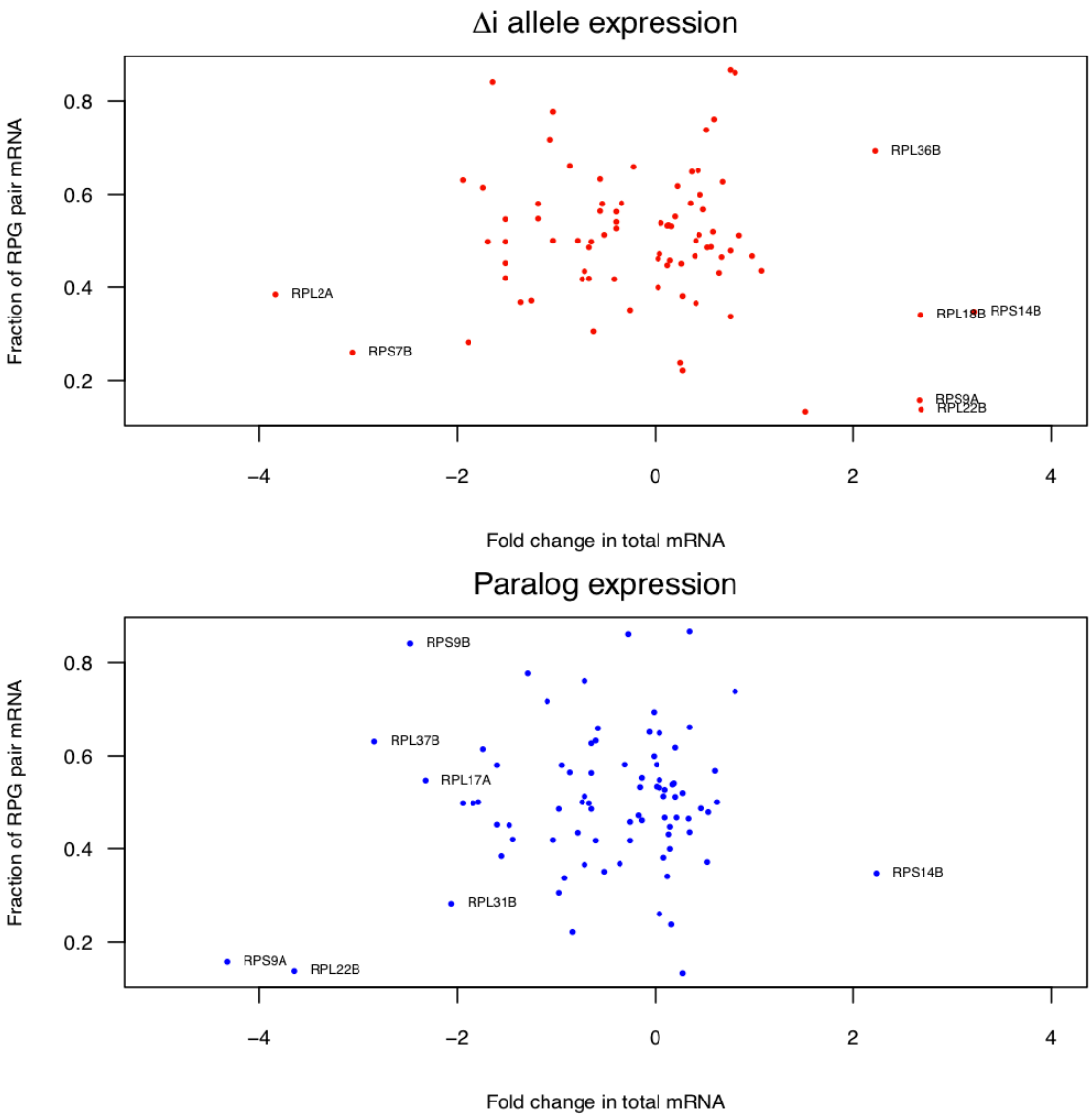
Bivariate scatterplot comparing data from Figure 3.2 (AP microarray) and Parenteau et al. 2011 data (JP qPCR). The fold change for the Δi allele (red) and its paralog (blue) are shown as circles. Note the largest fold changes are similar between datasets, while small (mostly positive) microarray fold changes are observed as moderate or large (mostly negative) qPCR changes.

Figure A.2 Total RPS9A/B mRNA expression measurements are strikingly different between two different studies



Stacked barplot comparing *RPS9A* (red) and *RPS9B* (blue) mRNA expression measurements between two independent studies. The mean relative fraction of total *RPS9A/B* mRNA for *RPS9A* and *RPS9B* genes in wild-type cells (first column) were determined based on the relative number of unique sequence reads for each gene from two RNA-seq datasets of wild-type strains (Nagalakshmi et al., 2008; C.-Y. Wu et al., 2010). For each mutant strain, percent mRNA expression (relative to wild-type) was calculated as the relative fraction of total *RPS9A/B* mRNA for that gene (first column) multiplied by its relative fold changes in the mutant strain (Figure 3.2 and Parentaeu *et al*. 2011). Error bars indicate the range of the difference between the two RNA-seq datasets.

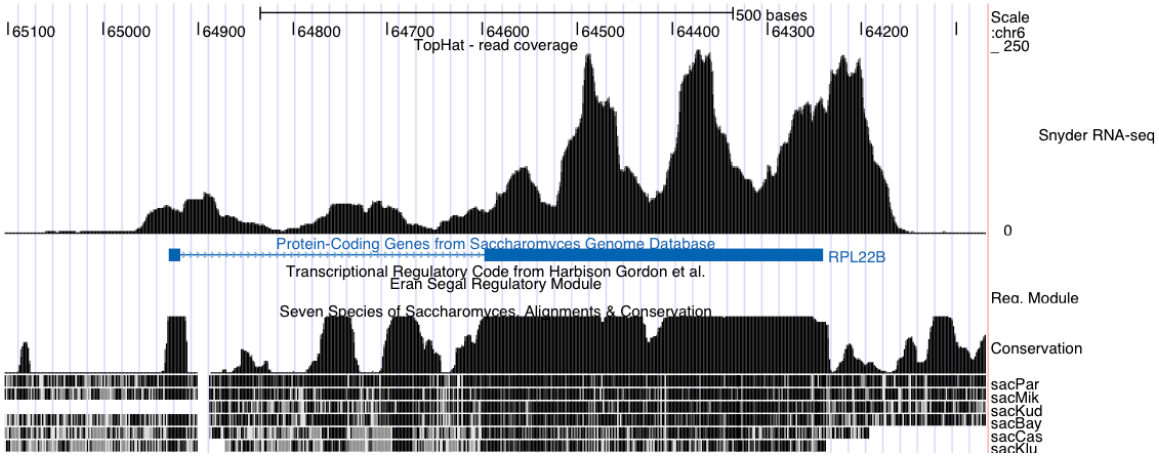
Figure A.3 RPGs most strongly affected by intron deletion are among the least highly expressed



Scatterplot comparing RPG expression level to fold change due to intron deletion. The affect on the Δi allele (red) and the paralogous gene copy (blue) are shown as dots and the affected gene copy is labeled to the right for fold changes > 4 fold. The estimated total ribosomal mRNA was calculated as in Figure A.2 except using only the RNA-seq dataset with the least missing data (C.-Y. Wu et al., 2010).

Figure A.4 Evidence for unspliced RPL22B mRNA associated with conserved intronic regions

A)



B)

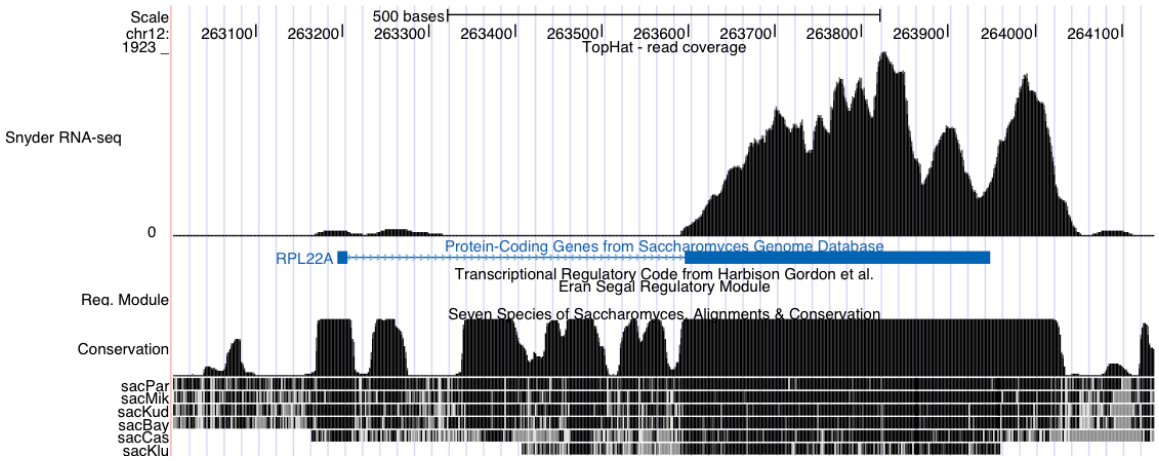


Figure A.5 Evidence for unspliced RPL18B mRNA associated with conserved intronic regions

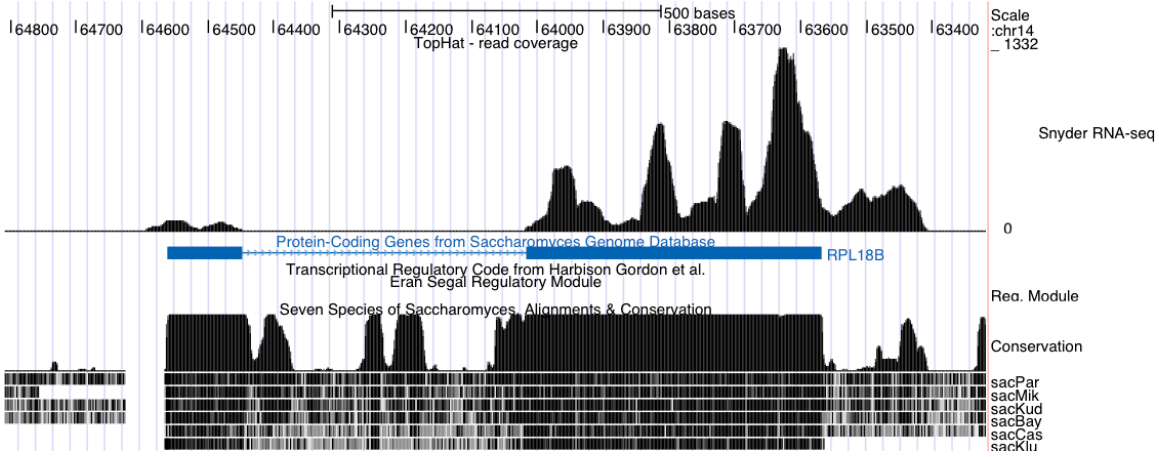
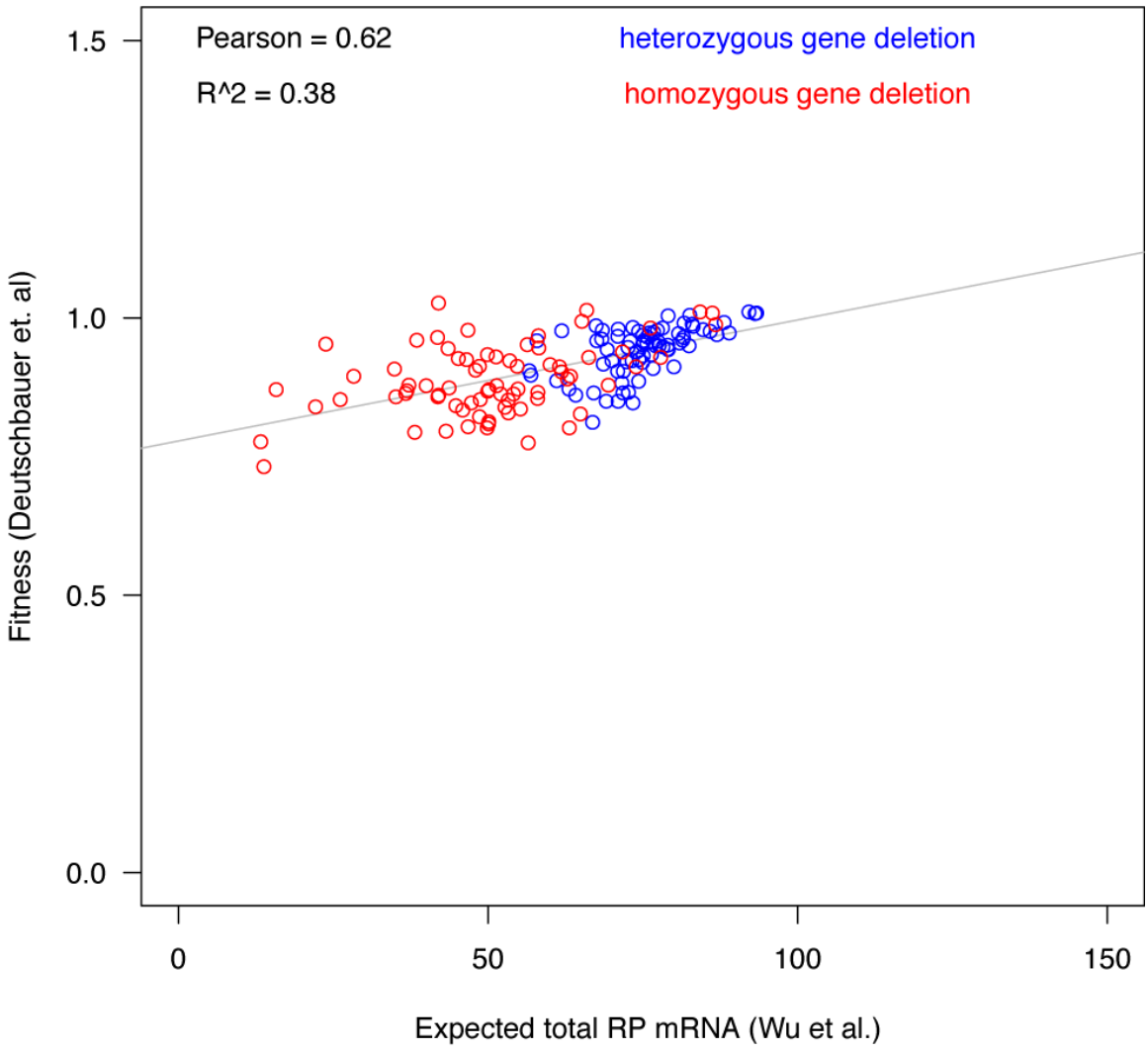
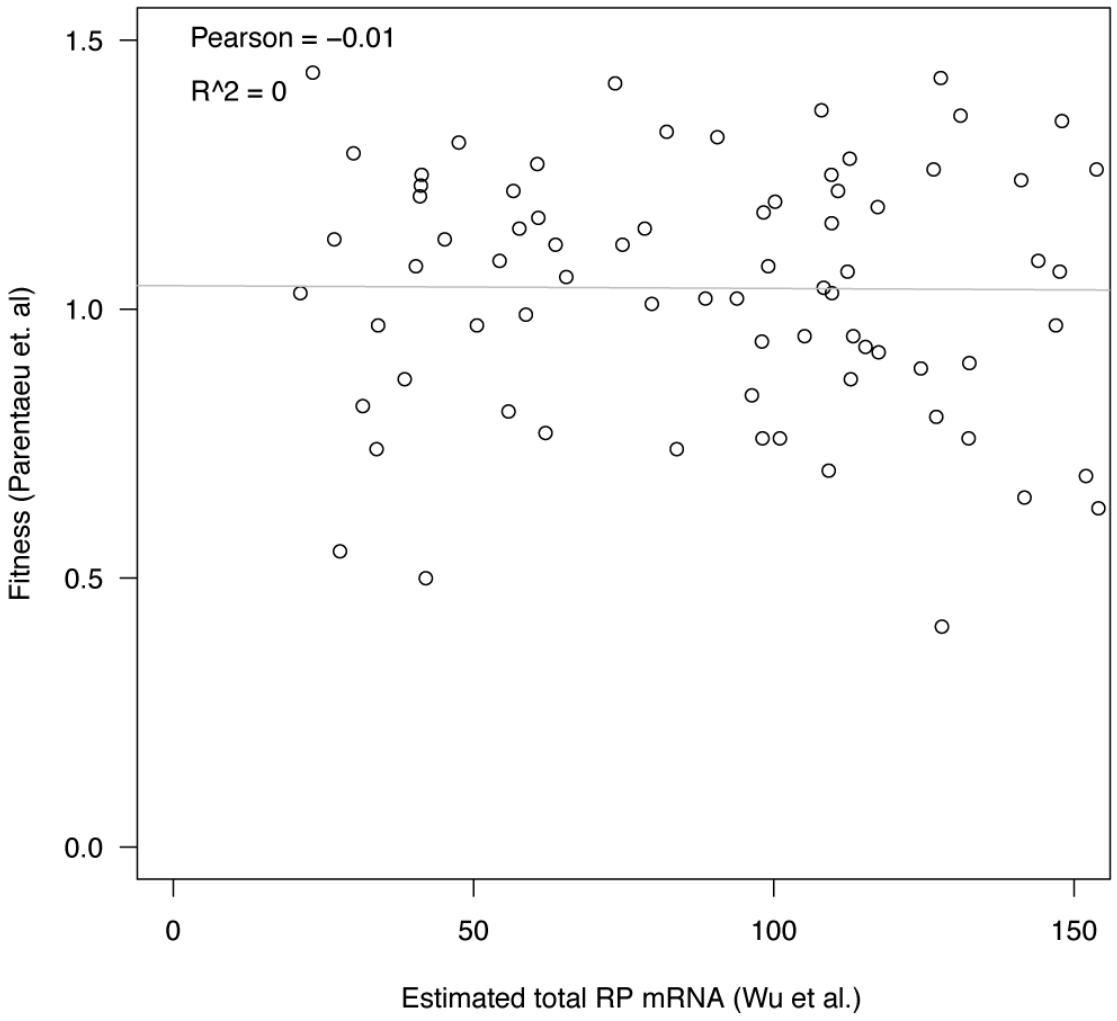


Figure A.6 Fitness correlates with total RPG mRNA levels predicted by gene deletion



Scatterplot illustrating the relationship between fitness (Deutschbauer et al., 2005) and expected total RPG pair expression. The expected total ribosomal mRNA was calculated for heterozygous gene deletions (blue) and homozygous gene deletions (red) assuming complete loss of the deleted gene and no change in other genes. Calculations were performed similar to Figure A.2 except using only the RNA-seq dataset with the least missing data (C.-Y. Wu et al., 2010).

Figure A.7 Fitness does not correlate with total RPG mRNA levels caused by intron deletion



Scatterplot illustrating the relationship between fitness (Parenteau et al., 2011) and estimated total RPG pair mRNA expression. The estimated total ribosomal mRNA was calculated as in Figure A.2 except using only the RNA-seq dataset with the least missing data (C.-Y. Wu et al., 2010).

Table 4.1 S. cerevisiae Δi strains

#	ORF	Gene name	Gene class	Essential	Intron deletion*
1	YBL027W	RPL19B	RPG	NO	1, 3
2	YBL072C	RPS8A	RPG	NO	1, 3
3	YBL087C	RPL23A	RPG	NO	1, 3
4	YBL092W	RPL32	RPG	YES	3
5	YBR048W	RPS11B	RPG	NO	3
6	YBR084C-A	RPL19A	RPG	NO	1, 3
7	YBR181C	RPS6B	RPG	NO	1, 3
8	YBR189W	RPS9B	RPG	NO	1, 3
9	YBR191W	RPL21A	RPG	NO	1, 3
10	YCR031C	RPS14A	RPG	NO	1, 3
11	YDL061C	RPS29B	RPG	NO	3
12	YDL075W	RPL31A	RPG	NO	1, 3
13	YDL082W	RPL1, 3A	RPG	NO	1, 3
14	YDL083C	RPS16B	RPG	NO	1, 3
15	YDL136W	RPL35B	RPG	NO	1, 3
16	YDL191W	RPL35A	RPG	NO	1, 3
17	YDR025W	RPS11A	RPG	NO	1, 3
18	YDR064W	RPS13	RPG	YES	3
19	YDR447C	RPS17B	RPG	NO	1, 3
20	YDR450W	RPS18A	RPG	NO	1, 3
21	YDR471W	RPL27B	RPG	NO	1, 3
22	YDR500C	RPL37B	RPG	NO	1, 3
23	YER056C-A	RPL34A	RPG	NO	1, 3
24	YER074W	RPS24A	RPG	NO	1, 3
25	YER102W	RPS8B	RPG		1, 3
26	YER117W	RPL23B	RPG	NO	1, 3
27	YER131W	RPS26B	RPG	NO	1
28	YFL034C-A	RPL22B	RPG		3
29	YFR031C-A	RPL2A	RPG	NO	1, 3
30	YFR032C-A	RPL29	RPG		1, 3
31	YGL030W	RPL30	RPG	YES	3
32	YGL031C	RPL24A	RPG	NO	3
33	YGL076C_0	RPL7A	RPG	NO	3
34	YGL076C_1	RPL7A	RPG	NO	3
35	YGL103W	RPL28	RPG	YES	3
36	YGL189C	RPS26A	RPG	NO	3
37	YGR027C	RPS25A	RPG	NO	3
38	YGR034W	RPL26B	RPG	NO	3
39	YGR118W	RPS23A	RPG	NO	3
40	YGR148C	RPL24B	RPG	NO	3
41	YGR214W	RPS0A	RPG	NO	3
42	YHL001W	RPL14B	RPG		3
43	YHR010W	RPL27A	RPG	NO	3
44	YHR021C	RPS27B	RPG	NO	3

45	YHR141C	RPL42B	RPG		1, 3
46	YHR203C	RPS4B	RPG	NO	3
47	YIL018W	RPL2B	RPG	NO	3
48	YIL052C	RPL34B	RPG	NO	3
49	YIL069C	RPS24B	RPG	NO	3
50	YIL133C	RPL16A	RPG	NO	3
51	YIL148W	RPL40A	RPG	NO	3
52	YJL136C	RPS21B	RPG	NO	3
53	YJL177W	RPL17B	RPG	NO	3
54	YJL189W	RPL39	RPG	NO	3
55	YJL191W	RPS14B	RPG	NO	3
56	YJR094W-A	RPL43B	RPG	NO	3
57	YJR145C	RPS4A	RPG	NO	3
58	YKL006W	RPL14A	RPG	NO	3
59	YKL156W	RPS27A	RPG	NO	1, 3
60	YKL180W	RPL17A	RPG	YES	1, 3
61	YKR057W	RPS21A	RPG	NO	3
62	YKR094C	RPL40B	RPG	NO	3
63	YLR048W	RPS0B	RPG	NO	3
64	YLR061W	RPL22A	RPG	NO	3
65	YLR185W	RPL37A	RPG	NO	1, 3
66	YLR287C-A	RPS30A	RPG	NO	1, 3
67	YLR333C	RPS25B	RPG	NO	3
68	YLR344W	RPL26A	RPG	NO	3
69	YLR367W_0	RPS22B	RPG	NO	3
70	YLR367W_1	RPS22B	RPG	NO	1, 3
71	YLR388W	RPS29A	RPG	NO	3
72	YLR406C	RPL31B	RPG	NO	3
73	YLR448W	RPL6B	RPG	NO	3
74	YML024W	RPS17A	RPG	NO	3
75	YML026C	RPS18B	RPG	NO	3
76	YML073C	RPL6A	RPG	NO	3
77	YMR142C	RPL13B	RPG	NO	3
78	YMR143W	RPS16A	RPG	NO	3
79	YMR194W	RPL36A	RPG	NO	3
80	YMR230W	RPS10B	RPG	NO	3
81	YMR242C	RPL20A	RPG	NO	3
82	YNL069C	RPL16B	RPG	NO	3
83	YNL096C	RPS7B	RPG	NO	3
84	YNL162W	RPL42A	RPG	YES	3
85	YNL301C	RPL18B	RPG	NO	3
86	YNL302C	RPS19B	RPG	NO	3
87	YOL120C	RPL18A	RPG	YES	3
88	YOL121C	RPS19A	RPG	NO	3
89	YOL127W	RPL25	RPG	YES	3
90	YOR096W	RPS7A	RPG	NO	3
91	YOR182C	RPS30B	RPG	NO	3
92	YOR234C	RPL33B	RPG	NO	3

93	YOR293W	RPS10A	RPG	NO	3
94	YOR312C	RPL20B	RPG	NO	3
95	YPL079W	RPL21B	RPG	NO	3
96	YPL081W	RPS9A	RPG	NO	3
97	YPL090C	RPS6A	RPG	NO	3
98	YPL143W	RPL33A	RPG	YES	3
99	YPL198W_0	RPL7B	RPG	NO	3
100	YPL198W_1	RPL7B	RPG	NO	3
101	YPL249C-A	RPL36B	RPG		3
102	YPR043W	RPL43A	RPG	NO	3
103	YPR132W	RPS23B	RPG	NO	3
104	SNR17A	SNR17A	non-RPG	NO	1
105	SNR17B	SNR17B	non-RPG	NO	1
106	YAL001C	TFC3	non-RPG	YES	
107	YAL003W	EFB1	non-RPG	YES	
108	YAL030W	SNC1	non-RPG	NO	1
109	YBL018C	POP8	non-RPG	YES	2
110	YBL026W	LSM2	non-RPG	YES	2
111	YBL040C	ERD2	non-RPG	YES	
112	YBL050W	SEC17	non-RPG	YES	
113	YBL059C-A	YBL059C-A	non-RPG		1
114	YBL059W	YBL059W	non-RPG	NO	1
115	YBL091C-A	YBL091C-A	non-RPG	NO	1, 2
116	YBR062C	YBR062C	non-RPG	NO	
117	YBR078W	ECM33	non-RPG	NO	1, 2
118	YBR082C	UBC4	non-RPG	NO	1, 2
119	YBR090C	YBR090C	non-RPG	NO	
120	YBR111W-A_0	SUS1	non-RPG	YES	
121	YBR111W-A_1	SUS1	non-RPG	YES	1
122	YBR119W	MUD1	non-RPG	NO	1, 2
123	YBR186W	PCH2	non-RPG	NO	1
124	YBR219C	YBR219C	non-RPG	NO	1
125	YBR230C	YBR230C	non-RPG	NO	1
126	YBR255C-A	YBR255C-A	non-RPG		
127	YCL002C	YCL002C	non-RPG	NO	1
128	YCL005W-A_0	VMA9	non-RPG		1, 2
129	YCL005W-A_1	VMA9	non-RPG		2
130	YCL012C	YCL012C	non-RPG	NO	1
131	YCR028C-A	RIM1	non-RPG		1, 2
132	YCR097W_0	HMRA1	non-RPG	NO	
133	YCR097W_1	HMRA1	non-RPG	NO	
134	YDL012C	YDL012C	non-RPG	NO	1
135	YDL029W	ARP2	non-RPG	YES	2
136	YDL064W	UBC9	non-RPG	YES	2
137	YDL079C	MRK1	non-RPG	NO	1, 2
138	YDL108W	KIN28	non-RPG	YES	1
139	YDL125C	HNT1	non-RPG	NO	1, 2
140	YDL130W	RPP1B	non-RPG	NO	1, 2

141	YDL137W	ARF2	non-RPG	NO	1
142	YDL189W	RBS1	non-RPG	NO	
143	YDL219W	YDL219W	non-RPG	NO	1
144	YDR005C	MAF1	non-RPG	NO	1
145	YDR059C	UBC5	non-RPG	NO	2
146	YDR092W	UBC13	non-RPG	NO	2
147	YDR099W	BMH2	non-RPG	NO	
148	YDR129C	SAC6	non-RPG	NO	1, 2
149	YDR139C	RUB1	non-RPG	NO	1, 2
150	YDR305C	HNT2	non-RPG	NO	1, 2
151	YDR318W	MCM21	non-RPG	NO	1
152	YDR367W	YDR367W	non-RPG	YES	
153	YDR381C-A	YDR381C-A	non-RPG		
154	YDR381W	YRA1	non-RPG	YES	2
155	YDR397C	NCB2	non-RPG	YES	
156	YDR424C_0	DYN2	non-RPG	NO	1, 2
157	YDR424C_1	DYN2	non-RPG	NO	1, 2
158	YDR535C	YDR535C	non-RPG	NO	
159	YEL012W	UBC8	non-RPG	NO	1, 2
160	YER003C	PMI40	non-RPG	YES	2
161	YER007C-A	YER007C-A	non-RPG	NO	1
162	YER014C-A	BUD25	non-RPG	NO	
163	YER044C-A	MEI4	non-RPG	NO	1, 2
164	YER074W-A_0	YOS1	non-RPG		
165	YER074W-A_1	YOS1	non-RPG		
166	YER093C-A	YER093C-A	non-RPG	NO	1
167	YER133W	GLC7	non-RPG	YES	2
168	YER179W	DMC1	non-RPG	NO	1, 2
169	YFL031W	HAC1	non-RPG	NO	
170	YFL034C-B	MOB2	non-RPG	YES	2
171	YFL039C	ACT1	non-RPG	YES	2
172	YFR024C-A	LSB3	non-RPG	NO	1, 2
173	YGL033W	HOP2	non-RPG	NO	2
174	YGL087C	MMS2	non-RPG	NO	2
175	YGL137W	SEC27	non-RPG	YES	
176	YGL178W	MPT5	non-RPG	NO	2
177	YGL183C	MND1	non-RPG	NO	2
178	YGL187C	COX4	non-RPG	NO	2
179	YGL226C-A	OST5	non-RPG	NO	
180	YGL232W	YGL232W	non-RPG	NO	
181	YGL251C	HFM1	non-RPG	NO	2
182	YGR001C_0	YGR001C	non-RPG	NO	
183	YGR001C_1	YGR001C	non-RPG	NO	
184	YGR029W	ERV1	non-RPG	YES	2
185	YGR183C	QCR9	non-RPG	NO	2
186	YGR225W	AMA1	non-RPG	NO	2
187	YHR001W-A	QCR10	non-RPG	NO	2
188	YHR012W	VPS29	non-RPG	NO	

189	YHR016C	YSC84	non-RPG	NO	2
190	YHR039C-A	VMA10	non-RPG	NO	2
191	YHR041C	SRB2	non-RPG	NO	
192	YHR077C	NMD2	non-RPG	NO	2
193	YHR079C-A	SAE3	non-RPG	NO	2
194	YHR097C	YHR097C	non-RPG	NO	
195	YHR101C	BIG1	non-RPG	YES	2
196	YHR123W	EPT1	non-RPG	NO	2
197	YHR199C-A	YHR199C-A	non-RPG		
198	YIL004C	BET1	non-RPG	YES	
199	YIL073C	SPO22	non-RPG	NO	2
200	YIL106W	MOB1	non-RPG	YES	2
201	YIL111W	COX5B	non-RPG	NO	1, 2
202	YIL156W-B	YIL156W-B	non-RPG		
203	YJL001W	PRE3	non-RPG	YES	1, 2
204	YJL024C	APS3	non-RPG	NO	1
205	YJL031C	BET4	non-RPG	YES	
206	YJL041W	NSP1	non-RPG	YES	1, 2
207	YJL130C	URA2	non-RPG	NO	
208	YJL205C	NCE101	non-RPG		
209	YJR021C	REC107	non-RPG	NO	2
210	YKL002W	YKL002W	non-RPG	NO	1
211	YKL006C-A	SFT1	non-RPG	YES	1
212	YKL081W	TEF4	non-RPG	NO	
213	YKL157W	APE2	non-RPG	NO	2
214	YKL186C	MTR2	non-RPG	YES	2
215	YKL190W	CNB1	non-RPG	NO	1
216	YKR004C	ECM9	non-RPG	YES	
217	YKR095W-A	PCC1	non-RPG	NO	
218	YLL050C	COF1	non-RPG	YES	1, 2
219	YLR054C	YLR054C	non-RPG	NO	
220	YLR078C	BOS1	non-RPG	YES	1
221	YLR093C	NYV1	non-RPG	NO	
222	YLR128W	DCN1	non-RPG	NO	2
223	YLR199C	YLR199C	non-RPG	NO	1
224	YLR202C	YLR202C	non-RPG	NO	
225	YLR211C	YLR211C	non-RPG	NO	
226	YLR275W	SMD2	non-RPG	YES	2
227	YLR306W	UBC12	non-RPG	NO	2
228	YLR312W-A	MRPL15	non-RPG	NO	
229	YLR316C_0	TAD3	non-RPG	YES	2
230	YLR316C_1	TAD3	non-RPG	YES	1, 2
231	YLR329W	REC102	non-RPG	NO	2
232	YLR426W	YLR426W	non-RPG	NO	
233	YLR445W	YLR445W	non-RPG	NO	
234	YML017W	PSP2	non-RPG	NO	
235	YML025C	YML025C	non-RPG	YES	2
236	YML034W	SRC1	non-RPG	NO	2

237	YML036W	CGI121	non-RPG	NO	
238	YML056C	IMD4	non-RPG	NO	2
239	YML067C	ERV41	non-RPG	NO	
240	YML085C	TUB1	non-RPG	YES	2
241	YML094W	GIM5	non-RPG	NO	2
242	YML124C	TUB3	non-RPG	NO	
243	YMR033W	ARP9	non-RPG	YES	2
244	YMR079W	SEC14	non-RPG	YES	
245	YMR116C	ASC1	non-RPG	NO	
246	YMR125W	STO1	non-RPG	NO	2
247	YMR133W	REC114	non-RPG	NO	
248	YMR194C-B	YMR194C-B	non-RPG		
249	YMR201C	RAD14	non-RPG	NO	2
250	YMR225C	MRPL44	non-RPG	NO	2
251	YMR292W	GOT1	non-RPG	NO	
252	YNL012W	SPO1	non-RPG	NO	2
253	YNL038W	YNL038W	non-RPG	YES	2
254	YNL044W	YIP3	non-RPG	NO	
255	YNL050C	YNL050C	non-RPG	NO	
256	YNL066W	SUN4	non-RPG	NO	
257	YNL112W	DBP2	non-RPG	YES	2
258	YNL130C	CPT1	non-RPG	NO	2
259	YNL138W-A	YSF3	non-RPG	YES	2
260	YNL147W	LSM7	non-RPG	NO	2
261	YNL246W	VPS75	non-RPG	NO	2
262	YNL265C	IST1	non-RPG	NO	
263	YNL312W	RFA2	non-RPG	YES	2
264	YNL339C	YRF1-6	non-RPG	NO	
265	YNR053C	YNR053C	non-RPG	YES	
266	YOL047C	YOL047C	non-RPG	NO	
267	YOL048C	YOL048C	non-RPG	NO	
268	YOR122C	PFY1	non-RPG	YES	2
269	YOR318C	YOR318C	non-RPG	NO	
270	YPL031C	PHO85	non-RPG	NO	2
271	YPL075W	GCR1	non-RPG	YES	2
272	YPL109C	YPL109C	non-RPG	NO	
273	YPL129W	ANC1	non-RPG	NO	
274	YPL175W	SPT14	non-RPG	YES	2
275	YPL218W	SAR1	non-RPG	YES	
276	YPL230W	YPL230W	non-RPG	NO	
277	YPL241C	CIN2	non-RPG	NO	2
278	YPR028W	YIP2	non-RPG	NO	2
279	YPR063C	YPR063C	non-RPG	NO	
280	YPR098C	YPR098C	non-RPG	NO	
281	YPR170W-B	YPR170W-B	non-RPG		
282	YPR187W	RPO26	non-RPG	YES	

* 1 = This study, 2 = Parenteau et al., 2008, 3 = Parenteau et al., 2011

References

Abovich, N., & Rosbash, M. (1984). Two genes for ribosomal protein 51 of Saccharomyces cerevisiae complement and contribute to the ribosomes. *Molecular and Cellular Biology*, *4*(9), 1871-1879.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., et al. (2000). The genome sequence of Drosophila melanogaster. *Science (New York, N.Y.)*, *287*(5461), 2185-2195.

Alexander, R. D., Innocente, S. A., Barrass, J. D., & Beggs, J. D. (2010). Splicing-dependent RNA polymerase pausing in yeast. *Molecular Cell*, *40*(4), 582-593. doi:10.1016/j.molcel.2010.11.005

Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., & Komorowski, J. (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Research*, *19*(10), 1732-1741. doi:10.1101/gr.092353.109

Ares, M, Jr, Grate, L., & Pauling, M. H. (1999). A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA (New York, N.Y.)*, *5*(9), 1138-1139.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., et al. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, *129*(4), 823-837. doi:10.1016/j.cell.2007.05.009

Barta, I., & Iggo, R. (1995). Autoregulation of expression of the yeast Dbp2p "DEAD-box" protein is mediated by sequences in the conserved DBP2 intron. *The EMBO Journal*, *14*(15), 3800-3808.

Bellur, D. L., & Woodson, S. A. (2009). A minimized rRNA-binding site for ribosomal protein S4 and its implications for 30S assembly. *Nucleic Acids Research*, *37*(6), 1886-1896. doi:10.1093/nar/gkp036

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2011). GenBank. *Nucleic Acids Research*, *39*(Database issue), D32-37. doi:10.1093/nar/gkq1079

Berger, S. L. (2007). The complex language of chromatin regulation during transcription. *Nature*, *447*(7143), 407-12. doi:10.1038/nature05915

Bernstein, K. A., Gallagher, J. E. G., Mitchell, B. M., Granneman, S., & Baserga, S. J. (2004). The small-subunit processome is a ribosome assembly intermediate. *Eukaryotic Cell*, *3*(6), 1619-1626. doi:10.1128/EC.3.6.1619-1626.2004

Bon, E., Casaregola, S., Blandin, G., Llorente, B., Neuvéglise, C., Munsterkotter, M., Guldener, U., et al. (2003). Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Research*, *31*(4), 1121-1135.

Bonnal, S., Martínez, C., Förch, P., Bachi, A., Wilm, M., & Valcárcel, J. (2008). RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. *Molecular Cell*, *32*(1), 81-95. doi:10.1016/j.molcel.2008.08.008

Bradnam, K. R., & Korf, I. (2008). Longer first introns are a general property of eukaryotic gene structure. *PloS One*, *3*(8), e3093. doi:10.1371/journal.pone.0003093

Brinster, R. L., Allen, J. M., Behringer, R. R., Gelinas, R. E., & Palmiter, R. D. (1988). Introns increase transcriptional efficiency in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(3), 836-840.

Brès, V., Yoh, S. M., & Jones, K. A. (2008). The multi-tasking P-TEFb complex. *Current Opinion in Cell Biology*, *20*(3), 334-340. doi:10.1016/j.ceb.2008.04.008

Brès, V., Yoshida, T., Pickle, L., & Jones, K. A. (2009). SKIP interacts with c-Myc and Menin to promote HIV-1 Tat transactivation. *Molecular Cell*, *36*(1), 75-87. doi:10.1016/j.molcel.2009.08.015

Byrne, K. P., & Wolfe, K. H. (2005). The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research*, *15*(10), 1456-1461. doi:10.1101/gr.3672305

Chen, M., & Manley, J. L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Reviews. Molecular Cell Biology*, *10*(11), 741-754. doi:10.1038/nrm2777

Cherry, J. M., Ball, C., Weng, S., Juvik, G., Schmidt, R., Adler, C., Dunn, B., et al. (1997). Genetic and physical maps of Saccharomyces cerevisiae. *Nature*, *387*(6632 Suppl), 67-73.

Cuccurese, M., Russo, G., Russo, A., & Pietropaolo, C. (2005). Alternative splicing and nonsense-mediated mRNA decay regulate mammalian ribosomal gene expression. *Nucleic Acids Research*, *33*(18), 5965-5977. doi:10.1093/nar/gki905

Dabeva, M. D., & Warner, J. R. (1993). Ribosomal protein L32 of Saccharomyces cerevisiae regulates both splicing and translation of its own transcript. *The Journal of Biological Chemistry*, *268*(26), 19669-19674.

Damgaard, C. K., Kahns, S., Lykke-Andersen, S., Nielsen, A. L., Jensen, T. H., & Kjems, J. (2008). A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Molecular Cell*, *29*(2), 271-8. doi:S1097-2765(08)00007-5

Dawson, W. K., Fujiwara, K., & Kawai, G. (2007). Prediction of RNA pseudoknots using heuristic modeling with mapping and sequential folding. *PloS One*, *2*(9), e905. doi:10.1371/journal.pone.0000905

De Almeida, S. F., Grosso, A. R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., et al. (2011). Splicing enhances recruitment of methyltransferase

HYPB/Setd2 and methylation of histone H3 Lys36. *Nature Structural & Molecular Biology*, *18*(9), 977-983. doi:10.1038/nsmb.2123

Deal, R. B., & Henikoff, S. (2011). Histone variants and modifications in plant gene regulation. *Current Opinion in Plant Biology*, *14*(2), 116-122. doi:10.1016/j.pbi.2010.11.005

Dean, D., & Nomura, M. (1980). Feedback regulation of ribosomal protein gene expression in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(6), 3590-3594.

Dean, E. J., Davis, J. C., Davis, R. W., & Petrov, D. A. (2008). Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genetics*, *4*(7), e1000113. doi:10.1371/journal.pgen.1000113

Deckman, I. C., & Draper, D. E. (1987). S4-alpha mRNA translation regulation complex. II. Secondary structures of the RNA regulatory site in the presence and absence of S4. *Journal of Molecular Biology*, *196*(2), 323-332.

Deutschbauer, A. M., Jaramillo, D. F., Proctor, M., Kumm, J., Hillenmeyer, M. E., Davis, R. W., Nislow, C., et al. (2005). Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics*, *169*(4), 1915-1925. doi:10.1534/genetics.104.036871

Dohm, J. C., Lottaz, C., Borodina, T., & Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*, *36*(16), e105. doi:10.1093/nar/gkn425

Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L. T., Wu, X., et al. (2001). The highly reduced genome of an enslaved algal nucleus. *Nature*, *410*(6832), 1091-1096. doi:10.1038/35074092

Edmunds, J. W., Mahadevan, L. C., & Clayton, A. L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *The EMBO Journal*, *27*(2), 406-20. doi:7601967

Feng, Y., Chen, M., & Manley, J. L. (2008). Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. *Nature Structural & Molecular Biology*, *15*(10), 1040-1048. doi:10.1038/nsmb.1485

Fewell, S. W., & Woolford, J. L., Jr. (1999). Ribosomal protein S14 of Saccharomyces cerevisiae regulates its expression by binding to RPS14B pre-mRNA and to 18S rRNA. *Molecular and Cellular Biology*, *19*(1), 826-834.

Fink, G R. (1987). Pseudogenes in yeast? *Cell*, *49*(1), 5-6.

Fong, Y. W., & Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, *414*(6866), 929-933. doi:10.1038/414929a

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, *5*(10), R80. doi:10.1186/gb-2004-5-10-r80

Gluick, T. C., Gerstner, R. B., & Draper, D. E. (1997). Effects of Mg2+, K+, and H+ on an equilibrium between alternative conformations of an RNA pseudoknot. *Journal of Molecular Biology*, *270*(3), 451-463. doi:10.1006/jmbi.1997.1119

Grund, S. E., Fischer, T., Cabal, G. G., Antúnez, O., Pérez-Ortín, J. E., & Hurt, E. (2008). The inner nuclear membrane protein Src1 associates with subtelomeric genes and alters their regulated gene expression. *The Journal of Cell Biology*, *182*(5), 897-910. doi:10.1083/jcb.200803098

Grundy, F. J., & Henkin, T. M. (1991). The rpsD gene, encoding ribosomal protein S4, is autogenously regulated in Bacillus subtilis. *Journal of Bacteriology*, *173*(15), 4595-4602.

Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., & Young, R. A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, *130*(1), 77-88. doi:10.1016/j.cell.2007.05.042

Hansen, K. D., Lareau, L. F., Blanchette, M., Green, R. E., Meng, Q., Rehwinkel, J., Gallusser, F. L., et al. (2009). Genome-wide identification of alternative splice forms down-regulated by nonsense-mediated mRNA decay in Drosophila. *PLoS Genetics*, *5*(6), e1000525. doi:10.1371/journal.pgen.1000525

Hedges, S. B., Dudley, J., & Kumar, S. (2006). TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics (Oxford, England)*, *22*(23), 2971-2972. doi:10.1093/bioinformatics/btl505

Hon, G., Wang, W., & Ren, B. (2009). Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Computational Biology*, *5*(11), e1000566. doi:10.1371/journal.pcbi.1000566

House, A. E., & Lynch, K. W. (2006). An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. *Nature Structural & Molecular Biology*, *13*(10), 937-944. doi:10.1038/nsmb1149

Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., et al. (2000). Widespread aneuploidy revealed by DNA microarray expression profiling. *Nature Genetics*, *25*(3), 333-337. doi:10.1038/77116

Jeffares, D. C., Mourier, T., & Penny, D. (2006). The biology of intron gain and loss. *Trends in Genetics: TIG*, *22*(1), 16-22. doi:10.1016/j.tig.2005.10.006

Jin, C., & Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes & Development*, *21*(12), 1519-1529. doi:10.1101/gad.1547707

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., & Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark “nucleosome-free

regions" of active promoters and other regulatory regions. *Nat Genet*, *41*(8), 941-945. doi:10.1038/ng.409

Juneau, K., Miranda, M., Hillenmeyer, M. E., Nislow, C., & Davis, R. W. (2006). Introns regulate RNA and protein abundance in yeast. *Genetics*, *174*(1), 511-518. doi:10.1534/genetics.106.058560

Juneau, K., Nislow, C., & Davis, R. W. (2009). Alternative splicing of PTC7 in Saccharomyces cerevisiae determines protein localization. *Genetics*, *183*(1), 185-194. doi:10.1534/genetics.109.105155

Juneau, K., Palm, C., Miranda, M., & Davis, R. W. (2007). High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(5), 1522-1527. doi:10.1073/pnas.0610354104

Katinka, M. D., Duprat, S., Cornillot, E., Méténier, G., Thomarat, F., Prensier, G., Barbe, V., et al. (2001). Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. *Nature*, *414*(6862), 450-453. doi:10.1038/35106579

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, *12*(6), 996-1006. doi:10.1101/gr.229102. Article published online before print in May 2002

Kim, J., Guermah, M., McGinty, R. K., Lee, J.-S., Tang, Z., Milne, T. A., Shilatifard, A., et al. (2009). RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell*, *137*(3), 459-471. doi:10.1016/j.cell.2009.02.027

Kim, J., Hake, S. B., & Roeder, R. G. (2005). The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. *Molecular Cell*, *20*(5), 759-770. doi:10.1016/j.molcel.2005.11.012

Kim, S., Kim, H., Fong, N., Erickson, B., & Bentley, D. L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13564-13569. doi:10.1073/pnas.1109475108

Kim, T., & Buratowski, S. (2009). Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell*, *137*(2), 259-272. doi:10.1016/j.cell.2009.02.045

Kizer, K. O., Phatnani, H. P., Shibata, Y., Hall, H., Greenleaf, A. L., & Strahl, B. D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Molecular and Cellular Biology*, *25*(8), 3305-3316. doi:10.1128/MCB.25.8.3305-3316.2005

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X. S., & Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genetics*, *41*(3), 376-381. doi:10.1038/ng.322

Komili, S., Farny, N. G., Roth, F. P., & Silver, P. A. (2007). Functional specificity among ribosomal proteins regulates gene expression. *Cell*, *131*(3), 557-571. doi:10.1016/j.cell.2007.08.037

Kornblihtt, A. R., de la Mata, M., Fededa, J. P., Munoz, M. J., & Nogues, G. (2004). Multiple links between transcription and splicing. *RNA (New York, N.Y.)*, *10*(10), 1489-1498. doi:10.1261/rna.7100104

Kouskouti, A., & Talianidis, I. (2005). Histone modifications defining active genes persist after transcriptional and mitotic inactivation. *The EMBO Journal*, *24*(2), 347-357. doi:10.1038/sj.emboj.7600516

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, *128*(4), 693-705. doi:10.1016/j.cell.2007.02.005

Lareau, L. F., Brooks, A. N., Soergel, D. A. W., Meng, Q., & Brenner, S. E. (2007). The coupling of alternative splicing and nonsense-mediated mRNA decay. *Advances in Experimental Medicine and Biology*, *623*, 190-211.

Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., & Brenner, S. E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, *446*(7138), 926-929. doi:10.1038/nature05676

Latham, J. A., & Dent, S. Y. R. (2007). Cross-regulation of histone modifications. *Nat Struct Mol Biol*, *14*(11), 1017-1024. doi:10.1038/nsmb1307

Le Hir, H., Nott, A., & Moore, M. J. (2003). How introns influence and enhance eukaryotic gene expression. *Trends in Biochemical Sciences*, *28*(4), 215-220.

Lee, B. M., & Mahadevan, L. C. (2009). Stability of histone modifications across mammalian genomes: Implications for “epigenetic” marking. *Journal of Cellular Biochemistry*, *108*(1), 22-34. doi:10.1002/jcb.22250

Lee, R. C. H., Gill, E. E., Roy, S. W., & Fast, N. M. (2010). Constrained intron structures in a microsporidian. *Molecular Biology and Evolution*, *27*(9), 1979-1982. doi:10.1093/molbev/msq087

Lewis, B. P., Green, R. E., & Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(1), 189-192. doi:10.1073/pnas.0136770100

Li, B., Carey, M., & Workman, J. L. (2007). The role of chromatin during transcription. *Cell*, *128*(4), 707-19. doi:10.1016/j.cell.2007.01.015

Li, M., Phatnani, H. P., Guan, Z., Sage, H., Greenleaf, A. L., & Zhou, P. (2005). Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(49), 17636-17641. doi:10.1073/pnas.0506350102

Loyola, A., & Almouzni, G. (2007). Marking histone H3 variants: how, when and why? *Trends in Biochemical Sciences*, *32*(9), 425-33. doi:10.1016/j.tibs.2007.08.004

Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., & Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell*, *144*(1), 16-26. doi:10.1016/j.cell.2010.11.056

Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., & Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science (New York, N.Y.)*, *327*(5968), 996-1000. doi:10.1126/science.1184208

Macías, S., Bragulat, M., Tardiff, D. F., & Vilardell, J. (2008). L30 binds the nascent RPL30 transcript to repress U2 snRNP recruitment. *Molecular Cell*, *30*(6), 732-742. doi:10.1016/j.molcel.2008.05.002

Malygin, A. A., Parakhnevitch, N. M., Ivanov, A. V., Eperon, I. C., & Karpova, G. G. (2007). Human ribosomal protein S13 regulates expression of its own gene at the splicing step by a feedback mechanism. *Nucleic Acids Research*, *35*(19), 6414-6423. doi:10.1093/nar/gkm701

Mattox, W., Ryner, L., & Baker, B. S. (1992). Autoregulation and multifunctionality among trans-acting factors that regulate alternative pre-mRNA processing. *The Journal of Biological Chemistry*, *267*(27), 19023-19026.

McGhee, J. D., & Felsenfeld, G. (1983). Another potential artifact in the study of nucleosome phasing by chromatin digestion with micrococcal nuclease. *Cell*, *32*(4), 1205-1215.

McGinty, R. K., Kim, J., Chatterjee, C., Roeder, R. G., & Muir, T. W. (2008). Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature*, *453*(7196), 812-816. doi:10.1038/nature06906

Melamud, E., & Moult, J. (2009). Stochastic noise in splicing machinery. *Nucleic Acids Research*, *37*(14), 4873-4886. doi:10.1093/nar/gkp471

Minsky, N., Shema, E., Field, Y., Schuster, M., Segal, E., & Oren, M. (2008). Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nature Cell Biology*, *10*(4), 483-488. doi:10.1038/ncb1712

Mitrovich, Q M, & Anderson, P. (2000). Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in C. elegans. *Genes & Development*, *14*(17), 2173-2184.

Mitrovich, Quinn M, Tuch, B. B., Guthrie, C., & Johnson, A. D. (2007). Computational and experimental approaches double the number of known introns in the pathogenic yeast Candida albicans. *Genome Research*, *17*(4), 492-502. doi:10.1101/gr.6111907

Mohan, M., Herz, H.-M., Takahashi, Y.-H., Lin, C., Lai, K. C., Zhang, Y., Washburn, M. P., et al. (2010). Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). *Genes & Development*, *24*(6), 574-589. doi:10.1101/gad.1898410

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science (New York, N.Y.)*, *320*(5881), 1344-1349. doi:10.1126/science.1158441

Nahkuri, S., Taft, R. J., & Mattick, J. S. (2009). Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle (Georgetown, Tex.)*, *8*(20), 3420-3424.

Neugebauer, K. M. (2002). On the importance of being co-transcriptional. *Journal of Cell Science*, *115*(Pt 20), 3865-3871.

Ng, R., Domdey, H., Larson, G., Rossi, J. J., & Abelson, J. (1985). A test for intron function in the yeast actin gene. *Nature*, *314*(6007), 183-184.

Ni, J. Z., Grate, L., Donohue, J. P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., et al. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes & Development*, *21*(6), 708-718. doi:10.1101/gad.1525507

Nilsen, T. W., & Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature*, *463*(7280), 457-463. doi:10.1038/nature08909

Nomura, M., Gourse, R., & Baughman, G. (1984). Regulation of the synthesis of ribosomes and ribosomal components. *Annual Review of Biochemistry*, *53*, 75-117. doi:10.1146/annurev.bi.53.070184.000451

Oberdoerffer, S., Moita, L. F., Neems, D., Freitas, R. P., Hacohen, N., & Rao, A. (2008). Regulation of CD45 Alternative Splicing by Heterogeneous Ribonucleoprotein, hnRNPLL. *Science*, *321*(5889), 686-691. doi:10.1126/science.1157610

Pan, Q., Saltzman, A. L., Kim, Y. K., Misquitta, C., Shai, O., Maquat, L. E., Frey, B. J., et al. (2006). Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes & Development*, *20*(2), 153-158. doi:10.1101/gad.1382806

Parenteau, J., Durand, M., Morin, G., Gagnon, J., Lucier, J.-F., Wellinger, R. J., Chabot, B., et al. (2011). Introns within ribosomal protein genes regulate the production

and function of yeast ribosomes. *Cell*, *147*(2), 320-331. doi:10.1016/j.cell.2011.08.044

Parenteau, J., Durand, M., Véronneau, S., Lacombe, A.-A., Morin, G., Guérin, V., Cecez, B., et al. (2008). Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Molecular Biology of the Cell*, *19*(5), 1932-1941. doi:10.1091/mbc.E07-12-1254

Perales, R., & Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Molecular Cell*, *36*(2), 178-191. doi:10.1016/j.molcel.2009.09.018

Pierce, S. E., Davis, R. W., Nislow, C., & Giaever, G. (2007). Genome-wide analysis of barcoded Saccharomyces cerevisiae gene-deletion mutants in pooled cultures. *Nature Protocols*, *2*(11), 2958-2974. doi:10.1038/nprot.2007.427

Pirngruber, J., Shchebet, A., Schreiber, L., Shema, E., Minsky, N., Chapman, R. D., Eick, D., et al. (2009). CDK9 directs H2B monoubiquitination and controls replication-dependent histone mRNA 3'-end processing. *EMBO Reports*, *10*(8), 894-900. doi:10.1038/embor.2009.108

Pleiss, J. A., Whitworth, G. B., Bergkessel, M., & Guthrie, C. (2007a). Rapid, transcript-specific changes in splicing in response to environmental stress. *Molecular Cell*, *27*(6), 928-937. doi:10.1016/j.molcel.2007.07.018

Pleiss, J. A., Whitworth, G. B., Bergkessel, M., & Guthrie, C. (2007b). Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. *PLoS Biology*, *5*(4), e90. doi:10.1371/journal.pbio.0050090

Pnueli, L., & Arava, Y. (2007). Genome-wide polysomal analysis of a yeast strain with mutated ribosomal protein S9. *BMC Genomics*, *8*, 285. doi:10.1186/1471-2164-8-285

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

R Development Core Team. (2011). *Development Core Team, R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A., & Ban, N. (2011). Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science (New York, N.Y.)*, *331*(6018), 730-736. doi:10.1126/science.1198308

Raisner, R. M., Hartley, P. D., Meneghini, M. D., Bao, M. Z., Liu, C. L., Schreiber, S. L., Rando, O. J., et al. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell*, *123*(2), 233-248. doi:10.1016/j.cell.2005.10.002

Rogers, S. L., & Rogers, G. C. (2008). Culture of Drosophila S2 cells and their use for RNAi-mediated loss-of-function studies and immunofluorescence microscopy. *Nature Protocols*, *3*(4), 606-611. doi:10.1038/nprot.2008.18

Rossbach, O., Hung, L.-H., Schreiner, S., Grishina, I., Heiner, M., Hui, J., & Bindereif, A. (2009). Auto- and Cross-Regulation of the hnRNP L Proteins by Alternative Splicing. *Mol. Cell. Biol.*, *29*(6), 1442-1451. doi:10.1128/MCB.01689-08

Roy, S. W., & Gilbert, W. (2005). Complex early genes. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(6), 1986-1991. doi:10.1073/pnas.0408355101

Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology (Clifton, N.J.)*, *132*, 365-386.

Russo, A., Catillo, M., Esposito, D., Briata, P., Pietropaolo, C., & Russo, G. (2011). Autoregulatory circuit of human rpL3 expression requires hnRNP H1, NPM and KHSRP. *Nucleic Acids Research*, *39*(17), 7576-7585. doi:10.1093/nar/gkr461

Russo, A., Siciliano, G., Catillo, M., Giangrande, C., Amoresano, A., Pucci, P., Pietropaolo, C., et al. (2010). hnRNP H1 and intronic G runs in the splicing control of the human rpL3 gene. *Biochimica Et Biophysica Acta*, *1799*(5-6), 419-428. doi:10.1016/j.bbagrm.2010.01.008

Saltzman, A. L., Kim, Y. K., Pan, Q., Fagnani, M. M., Maquat, L. E., & Blencowe, B. J. (2008). Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. *Molecular and Cellular Biology*, *28*(13), 4320-4330. doi:10.1128/MCB.00361-08

Sano, K., Taguchi, A., Furumoto, H., Uda, T., & Itoh, T. (1999). Cloning, sequencing, and characterization of ribosomal protein and RNA polymerase genes from the region analogous to the alpha-operon of escherichia coli in halophilic archaea, halobacterium halobium. *Biochemical and Biophysical Research Communications*, *264*(1), 24-28. doi:10.1006/bbrc.1999.1480

Schlax, P. J., Xavier, K. A., Gluick, T. C., & Draper, D. E. (2001). Translational repression of the Escherichia coli alpha operon mRNA: importance of an mRNA conformational switch and a ternary entrapment complex. *The Journal of Biological Chemistry*, *276*(42), 38494-38501. doi:10.1074/jbc.M106934200

Schneider, M., Will, C. L., Anokhina, M., Tazi, J., Urlaub, H., & Lührmann, R. (2010). Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Molecular Cell*, *38*(2), 223-235. doi:10.1016/j.molcel.2010.02.027

Schones, D. E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., et al. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell*, *132*(5), 887-898. doi:10.1016/j.cell.2008.02.022

Schwartz, S., & Ast, G. (2010). Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. *The EMBO Journal*, *29*(10), 1629-1636. doi:10.1038/emboj.2010.71

Schwartz, S., Meshorer, E., & Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nature Structural & Molecular Biology*, *16*(9), 990-995. doi:10.1038/nsmb.1659

Sharma, S., Kohlstaedt, L. A., Damianov, A., Rio, D. C., & Black, D. L. (2008). Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nature Structural & Molecular Biology*, *15*(2), 183-191. doi:10.1038/nsmb.1375

Shema, E., Tirosh, I., Aylon, Y., Huang, J., Ye, C., Moskovits, N., Raver-Shapira, N., et al. (2008). The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes & Development*, *22*(19), 2664-2676. doi:10.1101/gad.1703008

Shi, J., Hu, Z., Pabon, K., & Scotto, K. W. (2008). Caffeine Regulates Alternative Splicing in a Subset of Cancer-Associated Genes: a Role for SC35. *Mol. Cell. Biol.*, *28*(2), 883-895. doi:10.1128/MCB.01345-07

Shubin, N., Tabin, C., & Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature*, *457*(7231), 818-823. doi:10.1038/nature07891

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., et al. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, *479*(7371), 74-79. doi:10.1038/nature10442

Spellman, R., Llorian, M., & Smith, C. W. J. (2007). Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Molecular Cell*, *27*(3), 420-434. doi:10.1016/j.molcel.2007.06.016

Spies, N., Nielsen, C. B., Padgett, R. A., & Burge, C. B. (2009). Biased chromatin signatures around polyadenylation sites and exons. *Molecular Cell*, *36*(2), 245-254. doi:10.1016/j.molcel.2009.10.008

Spingola, M., Grate, L., Haussler, D., & Ares, M. (1999). Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *RNA (New York, N.Y.)*, *5*(2), 221-234.

Stajich, J. E., Dietrich, F. S., & Roy, S. W. (2007). Comparative genomic analysis of fungal genomes reveals intron-rich ancestors. *Genome Biology*, *8*(10), R223. doi:10.1186/gb-2007-8-10-r223

Storici, F., Durham, C. L., Gordenin, D. A., & Resnick, M. A. (2003). Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(25), 14994-14999. doi:10.1073/pnas.2036296100

Tang, C. K., & Draper, D. E. (1989). Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. *Cell*, *57*(4), 531-536.

Tang, C. K., & Draper, D. E. (1990). Evidence for allosteric coupling between the ribosome and repressor binding sites of a translationally regulated mRNA. *Biochemistry*, *29*(18), 4434-4439.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., & Guigó, R. (2009). Nucleosome positioning as a determinant of exon recognition. *Nature Structural & Molecular Biology*, *16*(9), 996-1001. doi:10.1038/nsmb.1658

Topp, J. D., Jackson, J., Melton, A. A., & Lynch, K. W. (2008). A cell-based screen for splicing regulators identifies hnRNP LL as a distinct signal-induced repressor of CD45 variable exon 4. *RNA (New York, N.Y.)*, *14*(10), 2038-2049. doi:10.1261/rna.1212008

Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W. M. P., van Schaik, F. M. A., Varier, R. A., Baltissen, M. P. A., et al. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*, *131*(1), 58-69. doi:10.1016/j.cell.2007.08.016

Vincent, A., & Liebman, S. W. (1992). The yeast omnipotent suppressor SUP46 encodes a ribosomal protein which is a functional and structural homolog of the Escherichia coli S4 ram protein. *Genetics*, *132*(2), 375-386.

Vojnic, E., Simon, B., Strahl, B. D., Sattler, M., & Cramer, P. (2006). Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription. *The Journal of Biological Chemistry*, *281*(1), 13-15. doi:10.1074/jbc.C500423200

Von Hippel, P. H., Kowalczykowski, S. C., Lonberg, N., Newport, J. W., Paul, L. S., Stormo, G. D., & Gold, L. (1982). Autoregulation of gene expression. Quantitative evaluation of the expression and function of the bacteriophage T4 gene 32 (single-stranded DNA binding) protein system. *Journal of Molecular Biology*, *162*(4), 795-818.

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell*, *136*(4), 701-718. doi:10.1016/j.cell.2009.02.009

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*(7221), 470-476. doi:10.1038/nature07509

Wang, Zhibin, Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., Cui, K., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics*, *40*(7), 897-903. doi:ng.154

Warf, M. B., & Berglund, J. A. (2010). Role of RNA structure in regulating pre-mRNA splicing. *Trends in Biochemical Sciences*, *35*(3), 169-178. doi:10.1016/j.tibs.2009.10.004

Warner, J. R. (1989). Synthesis of ribosomes in Saccharomyces cerevisiae. *Microbiological Reviews*, *53*(2), 256-271.

Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends in Biochemical Sciences*, *24*(11), 437-440.

Washietl, S., Hofacker, I. L., & Stadler, P. F. (2005). Fast and reliable prediction of noncoding RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(7), 2454-2459. doi:10.1073/pnas.0409169102

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer New York.

Wolfe, K H, & Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, *387*(6634), 708-713. doi:10.1038/42711

Wu, C.-Y., Rolfe, P. A., Gifford, D. K., & Fink, G. R. (2010). Control of transcription by cell size. *PLoS Biology*, *8*(11), e1000523. doi:10.1371/journal.pbio.1000523

Wu, Z., Jia, X., de la Cruz, L., Su, X.-C., Marzolf, B., Troisch, P., Zak, D., et al. (2008). Memory T cell RNA rearrangement programmed by heterogeneous nuclear ribonucleoprotein hnRNPLL. *Immunity*, *29*(6), 863-875. doi:10.1016/j.immuni.2008.11.004

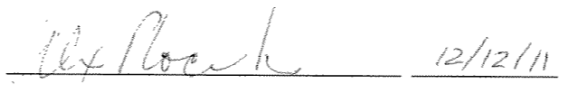
Yoh, S. M., Lucas, J. S., & Jones, K. A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes & Development*, *22*(24), 3422-34. doi:10.1101/gad.1720008

Zhang, Z., Hesselberth, J. R., & Fields, S. (2007). Genome-wide identification of spliced introns using a tiling microarray. *Genome Research*, *17*(4), 503-509. doi:10.1101/gr.6049107

Zhong, X.-Y., Wang, P., Han, J., Rosenfeld, M. G., & Fu, X.-D. (2009). SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Molecular Cell*, *35*(1), 1-10. doi:10.1016/j.molcel.2009.06.016

Zhu, B., Zheng, Y., Pham, A.-D., Mandal, S. S., Erdjument-Bromage, H., Tempst, P., & Reinberg, D. (2005). Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. *Molecular Cell*, *20*(4), 601-611. doi:10.1016/j.molcel.2005.09.025

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



Author Signature Date